

Università degli Studi di Torino
Doctoral School of Sciences and Innovative Technologies
Ph.D. Program in Pharmaceutical and Biomolecular Sciences
Cycle XXXVII



**UNIVERSITÀ
DI TORINO**

**GREEN(ER) Transition in Analytical
Measurements for GREEN(NES) FOOD**

PhD Candidate

Andrea Caratti

Supervisor

Chiara Emilia Irma Cordero



UNIONE EUROPEA

Fondo Sociale Europeo

UNIVERSITÀ DEGLI STUDI DI TORINO

DIPARTIMENTO DI SCIENZA E TECNOLOGIA DEL FARMACO

DOTTORATO DI RICERCA IN SCIENZE FARMACEUTICHE E
BIOMOLECOLARI

CICLO: XXXVII

TITOLO DELLA TESI: GREEN(ER) Transition in Analytical Measurements for GREEN(NES)
FOOD

TESI PRESENTATA DA: Andrea Caratti

SUPERVISORE: Prof.ssa Chiara Emilia Cordero

COORDINATORE DEL DOTTORATO: Prof.ssa Roberta Cavalli

ANNI ACCADEMICI: 2021/2022 – 2022/2023 – 2023/2024

SETTORE SCIENTIFICO-DISCIPLINARE DI AFFERENZA: CHEM-07/B (ex CHIM/10)

Progetto svolto grazie al sostegno finanziario del **Programma Operativo Nazionale
Ricerca e Innovazione 2014-2020 (CC12014IT16M2OP005)**, risorse FSE REACT-EU,
Azione IV.5 “Dottorati su tematiche Green”

Università degli Studi di Torino



UNIVERSITÀ
DI TORINO

Ph.D. Program in Pharmaceutical and Biomolecular Sciences

Department of Drug Science and Technology

Cycle XXXVII

Thesis title - GREEN(ER) Transition in Analytical Measurements for GREEN(NES) FOOD

Ph.D. Candidate - Dott. Andrea Caratti

Tutor - Prof. Chiara Emilia Irma Cordero

Ph.D. Coordinator - Prof. Roberta Cavalli

Academic years - 2021/2022 – 2022/2023 – 2023/2024

Scientific field of application - Food Chemistry – SSD CHEM/07-B

Thesis reviewers – Prof. Gianni Sagratini, Prof. Eugenio Aprea

Jury members – Prof. Chiara Emilia Irma Cordero, Prof. Gianni Sagratini, Prof. Eugenio Aprea, Dott. Daniela Peroni.

Contents

Abstract	11
Publications during enrollment	13
Oral and poster contributions	15
Oral presentations:	15
Poster presentations	15
Thesis including published works declaration.....	16
Synopsis of thesis investigation.....	17
Introductory Section	20
Chapter 1 Green Analytical Chemistry	21
1.1 Historical overview and concept of Green Analytical Chemistry.....	22
1.2 GAC principles.....	23
2.3 GAC metrics and tools.....	28
2.4 White Analytical Chemistry.....	32
References.....	35
Chapter 2 Gas Chromatography-Mass Spectrometry	39
2.1 . History, evolution, and impact of capillary Gas Chromatography-Mass Spectrometry in modern analytical chemistry	40
2.2 Mass Spectrometric detectors.....	42
2.2.1. High-Resolution Mass Spectrometry	43
2.3 Applications of GC-MS.....	45
2.3.1 Food Applications of GC-MS.....	47
2.4 Use of alternative carrier gasses in GC-MS	51
2.4.1 Hydrogen in GC-MS	54
2.5 Method translation	56
2.6 Fast GC.....	58
2.7 Portable GC systems.....	59
References.....	65
Chapter 3 Comprehensive Two-Dimensional Gas Chromatography-Mass Spectrometry: Principles and Tools	71
3.1 Introduction.....	72
3.2 Experiments set-up in different application fields.....	74
3.2.1 Sample Preparation.....	74
Gas samples and volatile analytes.....	75
Non-volatile analytes	75
Hyphenation with liquid chromatography.....	76

3.2.2 Column Geometries and Orthogonality	76
3.2.3 Modulators for GC×GC	78
3.3 Detection requirements in GC×GC	79
3.3.1 Single-channel detectors	79
3.3.2 Multi-channel detectors	80
3.4 Raw data interpretation	81
3.4.1 Data visualization and representation	81
3.4.2 Data processing: From raw-data to results	83
Pre-processing: Phase-shift and baseline correction	83
Two-dimensional peak detection and analyte identification	83
Chromatogram analysis	85
3.5 Limitations and platform selection	86
Laboratory effectiveness and efficiency	87
Instruments and parameter selection	87
3.6 Conclusions and Perspectives	88
References	91
Chapter 4 Artificial Intelligence in <i>food-omic</i> applications	97
4.1 Introduction	98
4.1.1 Nature's complexity and analytical system dimensions	101
4.1.2 Food – omics domains and the role of comprehensive two-dimensional chromatography	103
4.2 Machine Learning techniques: general overview	105
4.2.1 Current applications of AI in correlating analytical data and food properties	106
4.3 AI and comprehensive two-dimensional chromatography data processing	107
4.3.1 Image pattern recognition and computer vision applied to C2DC data: approaches and tools	108
Computer vision by datapoint features	109
Computer vision by region features	111
Computer vision by peak-region features	114
4.4 Artificial Intelligence Smelling based on sensomics: concept and tools	119
4.5 Concluding remarks	120
Supplementary information	122
Machine Learning techniques: tools description	122
References	128
Chapter 5 Comprehensive Two-Dimensional Gas Chromatography as a Key-Technology in Food-omics	145

5.1 Introduction	146
5.2 Analytical chemistry toolbox to advance knowledge at the boundary between chemistry and biology	147
5.2.1 Analytical dimensions.....	148
5.3 Food-Omics fields and opportunities to advance existing knowledge on food composition and properties.....	152
5.3.1 Approaches of investigation and terminology.....	154
5.4 Applications of GC×GC in the food-omic domains	155
5.4.1 Food-metabolomics and volatilomics.....	155
5.4.2 Nutrimetabolomics	158
5.4.3 Sensomics and Artificial Intelligence smelling.....	160
5.4.4 Food safety	163
5.5 The greenness potential of GC×GC when applied to food analysis	165
5.6 Conclusions.....	166
References.....	167
Experimental Section I Green(er) Transition of Established Analytical Methods	180
Chapter 6 EPA Method 8260D Making Greener the Volatile Organic Compounds Detection in Environmental Samples by Gas Chromatography-Mass Spectrometry.	181
6.1 Introduction to EPA Method 8260D.....	182
6.1.1 Procedural outline.....	182
Sample Preparation:.....	182
Calibration and Instrument Tuning:.....	183
Sample Analysis:.....	183
6.1.2 Quality Assurance and Controls	183
Instrument Tuning and Performance Check:.....	183
Calibration Curve and Continuing Calibration Verification (CCV):.....	183
Quality Control Samples:.....	183
Detection Limit Determination:.....	184
6.1.3 Validation and Data Analysis	184
Validation Criteria:.....	184
Data Analysis and Interpretation:	184
Reporting and Documentation:.....	185
6.2 P&T for VOCs in Aqueous Samples (EPA Method 5030)	185
6.2.1 Procedural Outline of EPA Method 5030	186
Sample Collection and Preservation:	186
Sample Preparation:.....	186
P&T Process:.....	186

Thermal Desorption:.....	186
6.2.2 Target VOCs	186
6.2.3 P&T recommended conditions.....	189
Purging Chamber	189
Trap Specifications and Conditioning.....	189
Desorber	189
Trap Packing Materials and Alternatives	189
Heated Purging Chamber	190
6.3 Capillary GC columns requirements	190
6.4 Use of hydrogen as alternative carrier gas	191
6.5 Italian legislation framework: Decree 156/06.....	192
6.6 Nitrogen as purge gas.....	193
6.6.1 Evaluation of requested parameters.....	193
6.7 Hydrogen as carrier gas in GC-MS	200
6.7.1 Evaluation of requested parameters	201
References.....	207
Chapter 7 Validation of a High-Throughput Method for the Accurate Quantification of Secondary Products of Lipid Oxidation in High-Quality Hazelnuts (<i>Corylus avellana</i> L.): A Robust Tool for Quality Assessment	209
7.1 Introduction	211
7.2 Materials and methods.....	213
7.2.1 Chemicals	213
7.2.2 Reference solutions and calibration mixtures	213
7.2.3 Hazelnut samples	213
7.2.4 Multiple headspace solid-phase microextraction: devices and conditions.....	214
7.2.5 GC-MS: instrument set-up and conditions.....	214
7.2.6 Multiple Headspace Extraction - principle of operation and reference equations	215
7.2.7 Responses normalization and choice of internal standards	216
7.2.8 Validation parameters.....	217
7.2.9 Data acquisition	217
7.3 Results and Discussion	217
7.3.1 Target analytes selection and quantification strategy	217
7.3.2 Method performance parameters.....	219
Identity confirmation.....	219
Response function and linearity within the working range	220
Precision: repeatability (r) and intermediate precision (R).....	221

Trueness/Accuracy	222
Limit of detection and limit of quantitation	224
Uncertainty associated with the results	225
7.3.3 Real-world samples application results and remarks	225
7.4 Conclusions.....	226
Electronic Supplementary Material	228
Supplementary Figure SF4.....	229
References	241
Chapter 8 Enhanced Selectivity VOCs Detection Using Novel GC-QEPAS - EVOQUE Project.....	247
8.1 Objectives and background of the EVOQUE project.....	248
8.1.1 The importance of VOCs	248
8.2 The EVOQUE project applications.....	249
8.3 Current state of the art vs. the EVOQUE system	251
8.4 EVOQUE system components.....	252
8.4.1 Portable Fast-GC.....	252
8.4.2 QEPAS detector	252
8.4.3 Metasurfaces.....	254
8.5 Aims of the EVOQUE project	255
8.6 Gen zero GC-QEPAS development: Testing column geometries for targeted VOCs separation	256
8.6.1 Linearity	258
8.6.2 LOD and LOQ.....	259
8.6.3 Alkanes discrimination	259
8.6.4 Resolution	260
8.6.5 Efficiency	262
8.6.6 Separation measure	263
8.7 Preliminary conclusions and future work	264
References	268
Experimental Section II Maximizing the Information Power by Multidimensional Techniques and AI Aided Strategies	270
Chapter 9 Artificial Intelligence Decision-Making Tools Based on Comprehensive Two-Dimensional Gas Chromatography Data: The Challenge of Quantitative Volatilomics in Food Quality Assessment	271
9.1 Introduction	272
9.2 Materials and methods.....	273
9.2.1 Reference standards and solvents.....	273
9.2.2 Reference solutions and calibration standards	273

9.2.3 Hazelnut samples	274
9.2.4 Headspace solid-phase microextraction: sample preparation, devices and conditions.....	274
9.2.5 GC×GC-MS/FID: instrument setup and conditions	275
9.2.6 Data acquisition and processing	275
9.3 Results and discussion.....	275
9.3.1 Raw hazelnut volatile fingerprint	276
9.3.2 Multiple markers quantitative analysis: accuracy and precision.....	279
9.3.3 AI decision-making tools: augmented smelling and robust classification.....	284
9.4 Conclusions	285
Electronic Supplementary Material	287
References	292
Chapter 10 Augmented Visualization by Computer Vision and Chromatographic Fingerprinting on Comprehensive Two-dimensional Gas Chromatographic Patterns: Unraveling Diagnostic Signatures in Food Volatilome	299
10.1 Introduction	301
10.2 Materials and methods	302
10.2.1 Reference standards and solvents.....	302
10.2.2 Reference solutions and calibration mixtures.....	302
10.2.3 Hazelnut samples.....	303
10.2.4 Headspace solid-phase microextraction: devices and conditions.....	303
10.2.5 GC×GC-TOFMS: instrument set-up and conditions.....	304
10.2.6 Combined Untargeted and Targeted (UT) fingerprinting by smart templates....	304
10.2.7 Comparative visualization by visual features	309
10.2.8 Data acquisition and 2D data processing	310
10.3 Results and discussion	310
10.3.1 Computer Vision by datapoint features: chromatograms transform and registration	312
10.3.2 Augmented Visualization by UT fingerprinting on composite class images: principle and workflow.....	314
10.3.3 Molecular patterns in butter processing.....	315
10.4 Conclusions	319
Acknowledgments	320
Supplementary Information.....	321
References	324

Chapter 11 Translating 2D-Chromatographic Fingerprinting to Quantitative Volatilomics: Unrevealing Compositional Changes in Maize Silage Volatilome for Robust Markers Discovery	329
11.1 Introduction	330
11.2 Materials and methods	331
11.2.1 Reference chemicals and solvents	331
11.2.2 Reference solutions and calibration mixtures.....	331
11.2.3 Fermented maize silage samples.....	332
11.2.4 Chemical and microbial analyses on silage samples.....	332
11.2.4 VOCs sampling and analysis by HS-SPME-GC×GC-TOF MS and TM with the reference method	333
Headspace solid-phase microextraction: devices and conditions	333
TM GC×GC-TOF MS: instrument set-up and conditions	334
11.2.5 VOCs sampling and analysis by HS-SPME-GC×GC-qMS/FID and FM in translated conditions.....	334
Multiple headspace solid phase microextraction: devices and conditions.....	334
FM GC×GC-qMS/FID: instrument set-up and conditions	335
11.2.6 External standard calibration by MHS-SPME-GC×GC-qMS/FID and FID predicted RRFs	336
11.2.7 Comprehensive chromatographic fingerprinting by image pattern recognition ..	338
11.2.8 Data acquisition, 2D data processing and statistical analysis	338
11.3 Results and discussion	339
11.3.1 Main fermentative and microbial characteristics of the silages	339
11.3.2 Volatilome compositional diversity and the impact of functional variables	339
11.3.3 Method translation from TM to FM GC×GC: information capacity and targeted analytes coverage	342
11.3.4 Quantitative volatilomics by FM GC×GC-qMS/FID: accuracy validation and predicted FID relative response factors	344
11.3.5 Quantitative volatilomics by FM GC×GC-qMS/FID: results	346
Supplementary Information.....	350
References	361
Chapter 12 Beyond Current Quality Indices: Quantitative Volatilomics Unrevealed Cultivar Traits, Harvesting Practices Impact, and Aroma Blueprint of Extra-Virgin Olive Oils	367
12.1 Introduction	368
12.2 Materials and methods.....	369
12.2.1 Extra virgin olive oil samples	369

12.2.3 Reference standards and solvents	370
12.2.4 Multiple headspace solid phase microextraction: devices and conditions	374
12.2.5 Comprehensive two-dimensional gas chromatography: instrument set-up and conditions.....	374
12.2.6 Analytes identification criteria.....	375
12.2.7 Analytical data acquisition, processing and mining	375
Combined Untargeted and Targeted (UT) fingerprinting by smart templates.....	375
12.2.8 Data acquisition and statistical analysis	376
12.3 Results and discussion.....	376
12.3.1 Picual olive oils: Organic vs. conventional cultivation diagnostic volatiles signature	379
12.3.2 Hojiblanca olive oils: Organic vs. conventional cultivation diagnostic volatiles signature.....	382
12.3.3 Accurate quantification of aroma active compounds: Artificial Intelligence smelling	382
12.3.4 Aroma Blueprinting of organic extra-virgin olive oils.....	384
12.4 Conclusions	386
Acknowledgements.....	387
Supplementary Information.....	388
References.....	395
Conclusive Remarks	401
Acknowledgements	405

Abstract

This doctoral thesis focuses on the development of innovative and sustainable analytical methodologies for food quality and safety assessment, with a strong emphasis on the transition to greener practices. The research integrates Green Analytical Chemistry (GAC) principles, advanced chromatographic techniques, and artificial intelligence (AI)-based systems to address the growing demand for efficient, cost-effective, and environmentally friendly analytical solutions in food quality control.

Several key areas of analytical chemistry were explored, primarily focusing on the development of fast, high-throughput methods for food analysis. An increase of efficiency and speed of analytical workflows were achieved by optimizing chromatographic techniques such as fast gas chromatography (fast-GC), while maintaining accuracy and reliability. These advancements are particularly relevant for food quality control, where rapid analysis is essential for meeting industrial demands. The analysis of large numbers of samples in a shorter time frame, supporting both sustainability and operational efficiency in food analysis was achieved thanks to the implementation of high-throughput methods; the development of portable GC systems was explored and preliminarily tested enhancing the accessibility and sustainability of food analysis by on-site systems. The ultimate goal, as a part of the EVOQUE project, is the development of a portable, compact GC-QEPAS (Quartz-Enhanced PhotoAcoustic Spectroscopy) system for volatile organic compounds (VOCs) detection with confirmatory attitudes as mass spectrometry (MS) but lower operational costs and background requirements.

Comprehensive two-dimensional gas chromatography (GC×GC) coupled with MS detection, a technique that enhances separation capacity and provides detailed insights into complex food fractions, was also considered. This high-information capacity platform, when combined with AI tools like machine learning and computer vision, allows for the efficient analysis of large, complex datasets, improving food quality assessments and enabling more confident traceability, and multi-level quality characterization. Four applications were developed exploiting GC×GC-MS full potential; they include hazelnut quality leveling, butter processing characterization, maize silage VOCs quantification, and extra virgin olive oil (EVOO) aroma blueprinting. By identifying key volatile compounds (i.e., markers) and correlating them with quality or authenticity traits, the research offers valuable tools for shelf life quality predictions and stability, process monitoring during the industrial production chain, spoilage detection, and the identification of products' origins. The integration of AI-driven chemometric models further enhances the reliability and precision of these analyses, supporting decision-making in food production and quality control.

The development of greener methods for detecting contaminant VOCs was another key aspect of the research project. The use of renewable gases in GC workflows, such as hydrogen and nitrogen, alongside the optimization of purge-and-trap techniques, demonstrated how sustainability can be incorporated into VOC analysis without compromising regulatory compliance or performance standards.

Overall, this thesis provides a comprehensive framework for modernizing food analysis by combining advanced chromatographic techniques, AI, and sustainable practices. The findings contribute to the advancement of green analytical chemistry, offering innovative solutions for improving food quality control, supporting sustainability, and advancing the green transition in analytical methodologies.

Publications during enrollment

1. **A. Caratti**, S. Squara, F. Stilo, S. Battaglini, E. Liberto, I. Cincera, G. Genova, N. Spigolon, C. Bicchi, C. Cordero, Integrated Strategy for Informative Profiling and Accurate Quantification of Key-Volatiles in Dried Fruits and Nuts: An Industrial Quality Control Perspective, *Foods*. 11 (2022) 3111. DOI: 10.3390/foods11193111.
2. S. Squara, **A. Caratti**, F.O. Gavilan, P. Bolzoni, N. Spigolon, G. Genova, G. Castello, M.G.B. González, L. Cuadros-Rodriguez, C. Bicchi, C. Cordero, Validation of a high-throughput method for the accurate quantification of secondary products of lipid oxidation in high-quality hazelnuts (*Corylus avellana* L.): A robust tool for quality assessment, *J. Food Compos. Anal.* 114 (2022) 104766. DOI: 10.1016/j.jfca.2022.104766.
3. S. Squara, F. Manig, T. Henle, M. Hellwig, **A. Caratti**, C. Bicchi, S.E. Reichenbach, Q. Tao, M. Collino, C. Cordero, Extending the breadth of saliva metabolome fingerprinting by smart template strategies and effective pattern realignment on comprehensive two-dimensional gas chromatographic data, *Anal. Bioanal. Chem.* 415 (2023) 2493–2509. DOI: 10.1007/s00216-023-04516-x.
4. **A. Caratti**, S. Squara, C. Bicchi, Q. Tao, D. Geschwender, S.E. Reichenbach, F. Ferrero, G. Borreani, C. Cordero, Augmented visualization by computer vision and chromatographic fingerprinting on comprehensive two-dimensional gas chromatographic patterns: Unraveling diagnostic signatures in food volatilome, *J. Chromatogr. A*. 1699 (2023) 464010. DOI: 10.1016/j.chroma.2023.464010.
5. S. Squara, **A. Caratti**, A. Fina, E. Liberto, N. Spigolon, G. Genova, G. Castello, I. Cincera, C. Bicchi, C. Cordero, Artificial Intelligence decision-making tools based on comprehensive two-dimensional gas chromatography data: the challenge of quantitative volatilomics in food quality assessment, *J. Chromatogr. A*. 1700 (2023) 464041. DOI: 10.1016/j.chroma.2023.464041.
6. S. Squara, S. Moraglio, **A. Caratti**, A. Fina, E. Liberto, C. Bicchi, C.H. Weinert, S.T. Soukup, L. Tavella, C. Cordero, Unrevealing the Halyomorpha halys Damage Fingerprint on Hazelnut Metabolome by Multiomic Platforms and AI-Aided Strategies, *J. Agric. Food Chem.* 72 (2024) 24109–24129. DOI: 10.1021/acs.jafc.4c06888.
7. S. Squara, S. Moraglio, **A. Caratti**, A. Fina, E. Liberto, C. Bicchi, C.H. Weinert, S.T. Soukup, L. Tavella, C. Cordero, Unrevealing the Halyomorpha halys Damage Fingerprint on Hazelnut Metabolome by Multiomic Platforms and AI-Aided Strategies, *J. Agric. Food Chem.* 72 (2024) 24109–24129. DOI: 10.1021/acs.jafc.4c06888.
8. **A. Caratti**, A. Fina, F. Trapani, C. Bicchi, E. Liberto, C. Cordero, F. Magagna, Artificial Intelligence Sensing: Effective Flavor Blueprinting of Tea Infusions for a Quality Control Perspective, *Molecules*. 29 (2024) 565. DOI: 10.3390/molecules29030565.
9. **A. Caratti**, S. Squara, C. Bicchi, E. Liberto, M. Vincenti, S.E. Reichenbach, Q. Tao, D. Geschwender, E. Alladio, C. Cordero, Boosting comprehensive two-dimensional chromatography with artificial intelligence: Application to food-omics, *TrAC Trends Anal. Chem.* 174 (2024) 117669. DOI: 10.1016/j.trac.2024.117669.

10. N.S. Brillhante, H.R. Bizzo, **A. Caratti**, S. Squara, C. Cordero, Artificial intelligence tools and concepts give access to authenticity and quality information in Brazilian olive oil volatilome, *J. Food Compos. Anal.* 136 (2024) 106826. DOI: 10.1016/j.jfca.2024.106826.
11. S. Squara, **A. Caratti**, A. Fina, E. Liberto, N. Koljančić, I. Špánik, G. Genova, G. Castello, C. Bicchi, A. de Villiers, C. Cordero, Artificial intelligence decision making tools in food metabolomics: Data fusion unravels synergies within the hazelnut (*Corylus avellana* L.) metabolome and improves quality prediction, *Food Res. Int.* 194 (2024) 114873. DOI: 10.1016/j.foodres.2024.114873.
12. **A. Caratti**, F. Ferrero, E. Tabacco, F. Gerbaldo, E. Liberto, G. Borreani, C. Cordero, Translating 2D-Chromatographic Fingerprinting to Quantitative Volatilomics: Unrevealing Compositional Changes in Maize Silage Volatilome for Robust Marker Discovery, *J. Agric. Food Chem.* 72 (2024) 23616–23630. DOI: 10.1021/acs.jafc.4c05877.
13. G. Rolando, S.T. Moraglio, **A. Caratti**, C. Cordero, G. Borreani, L. Tavella, Quantitative and qualitative damage caused by *Halyomorpha halys* (Hemiptera: Pentatomidae) on soybean crop at different growth stages, *Crop Prot.* 187 (2025) 106987. DOI: 10.1016/j.cropro.2024.106987.
14. **A. Caratti**, A. Fina, F. Trapani, E. Liberto, B. Jiménez-Herrera, L. Arce, R.M. Callejón, C. Cordero, Beyond current quality indices: Quantitative volatilomics unveiled cultivar traits, harvesting practices impact, and aroma blueprint of extra-virgin olive oils, *J. Food Compos. Anal.* 137 (2025) 106975. DOI: 10.1016/j.jfca.2024.106975.

Oral and poster contributions

Oral presentations:

1. **Caratti A.**, Squara S., Bicchi C., Reichenbach S.E., Tao Q., Ferrero F., Borreani G., Cordero C. "Computer Vision to analyze chemical signatures: a novel workflow for rationalizing raw data exploration in GC×GC". XIII Congresso Nazionale di Chimica degli Alimenti, Marsala (29th-31st May 2023).
2. **Caratti A.**, Squara S., Fina A., Reichenbach S. E., Tao Q., Liberto E., Bicchi C., Borreani G., Ferrero F., Cordero C. "A New Workflow for Rationalizing Raw Data Exploration in GC×GC: Computer Vision to Unravel Diagnostic Signatures". Autumn School in Food Chemistry 2nd edition Italian School in Food Chemistry for PhD student. Pavia, Italia (September 20th-22nd 2023).
3. **Caratti A.**, Fina A., Trapani F., Bicchi C., Reichenbach S.E., Tao Q., Geschwender D., Cordero C. "Innovations and challenges in quantitative volatilomics: the role of FID/MS chromatogram fusion in enhancing pattern recognition". 11th International Symposium on Recent Advances In Food Analysis (RAFA), Prague, Czech Republic (5th-8th November 2024).

Poster presentations

1. **Caratti A.**, Bolzoni P., Squara S., Giaccone D., Borreani G., Dolci P., Nucera D., Bruno F., Bicchi C., Sgorbini B., Cordero C. "Challenges and Opportunities for Micro-Extraction Techniques to Accurately Monitor Cheese Volatilome Evolution During Ripening and Promptly Detect Spoilage" 2nd European Sample Preparation e-Conference and the 1st Green and Sustainable Analytical Chemistry e-Conference (14th-16th March 2022).
2. **Caratti A.**, Squara S., Liberto E., Bicchi C., Reichenbach S.E., Callejon Fernandez R.M., Rodriguez L.C., Cordero C. "Study of the high-quality extra-virgin olive oils volatilome: potentiality of "comprehensive" two-dimensional gas chromatography for the discrimination of olive cultivation methodologies". 7th MS Food Day, Firenze, Italy (5th-7th October 2022).
3. **Caratti A.**, Squara S., Reichenbach S.E., Tao Q., Bicchi C., Borreani G., Ferrero F., Cordero C. "Computer vision enables effective detection of compositional differences in complex samples: validation of a workflow based on chromatographic fingerprinting and pattern recognition" 14th Multidimensional Chromatography workshop, Liège, Belgium (30th -1st February 2023).
4. **Caratti A.**, Bizzo H., Squara S., Bicchi C., Cordero C. "Artificial intelligence smelling by GC×GC-MS/FID as a tool for food aroma blueprinting: the unique aroma of Brazilian olive oils" Brilhante N., 14th Multidimensional Chromatography workshop, Liège, Belgium (30th -1st February 2023).
5. **Caratti A.**, Brilhante N., Squara S., Bicchi C., Bizzo H., Cordero C. "Augmented smelling to reveal Brazilian Extra Virgin Olive Oil aroma blueprint: accurate quantification of key-aroma compounds by comprehensive two-dimensional gas

chromatography and parallel detection by MS and FID” XIII Congresso Nazionale di Chimica degli Alimenti, Marsala, Italy (29th-31st May 2023).

Thesis including published works declaration

I hereby declare that this thesis contains no material that has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published, except where due reference is made in the text of the thesis.

This thesis includes six original papers published in peer-reviewed journals. A book chapter has been submitted for publication but to date not yet published. Various papers on GC and GC×GC applications on different aspects of food quality were published during the course of this Ph.D. program, providing background knowledge for this thesis and helping with the preparation of the introductory Chapters.

The ideas, development and writing up of all the papers in the thesis were the principal responsibility of myself, the student, working within the Dipartimento di Scienza e Tecnologia del Farmaco, University of Turin under the supervision of Professor Chiara Emilia Irma Cordero and industry supervisor Dr. Armando Miliazza (SRA Instruments, SpA).

The inclusion of co-authors reflects the fact that the work came from active collaboration between researchers and acknowledges input into team-based research.

Synopsis of thesis investigation

This doctoral thesis is part of a broader European project aimed at developing innovative and sustainable methodologies for analytical chemistry. The research aligns with the principles of Green Analytical Chemistry (GAC), addressing the scientific community's growing commitment to reducing the environmental impact of analytical methods. By combining green principles, advanced chromatographic techniques, and artificial intelligence (AI), this work aims to balance high analytical performance with sustainability, overcoming challenges such as the energy-intensive nature of some methods required in food analysis.

In the **Introductory section**, the **Chapter 1** introduces the principles of GAC and the methodologies for assessing the sustainability of analytical processes. The focus is on understanding how green principles can be integrated into analytical workflows to reduce their environmental impact while maintaining high performance. This chapter also highlights the broader concept of White Analytical Chemistry (WAC), which incorporates safety and sustainability considerations into the evaluation of analytical methods. By establishing a theoretical foundation, this chapter provides the context for the development of greener and more innovative approaches in food chemistry, setting the stage for the subsequent research presented in the thesis.

The **second chapter** focuses on gas chromatography-mass spectrometry (GC-MS) and its fundamental role in analytical chemistry. Given the increasing scarcity and cost of helium, the chapter explores alternative carrier gases such as hydrogen and nitrogen, evaluating their feasibility and performance in various contexts. The development of portable GC systems is also discussed, highlighting their potential for on-site analyses, which expand the applicability of GC-MS while aligning with green principles by reducing logistical and resource demands.

A focus on the principles and advancements GC×GC, a technique that addresses the limitations of traditional one-dimensional chromatography is reported in **Chapter 3**. GC×GC enhances peak capacity and selectivity through the use of a modulator, enabling the detailed separation of complex mixtures. This chapter discusses the technological innovations in thermal and flow modulation that have improved the performance and applicability of GC×GC, highlighting its versatility and potential for use in a wide range of analytical scenarios. Furthermore, GC×GC enables the simultaneous monitoring of hundreds of analytes in a single run. This capability generates large and complex data matrices, offering unparalleled insight into sample composition and presenting new opportunities for data-driven research.

Chapter 4 examines the potential that the use of AI in analytical workflows can offer. Advanced chromatographic techniques, such as GC×GC, generate huge and complex datasets, which traditional methods struggle to process efficiently. This chapter highlights the role of AI-driven tools, such as machine learning and computer vision, in managing and interpreting these datasets. By uncovering patterns and insights, AI enhances the reliability and depth of food-omics research, enabling more precise quality assessments and origin tracing.

Considering the potential of the GC×GC and the support AI can bring to data processing, **Chapter 5** focuses on the integration of this technique in food-omics, emphasizing its role in food safety, quality control, and origin tracing. GC×GC is presented as a powerful tool for analyzing complex food samples, offering high-resolution separation of volatile and non-volatile compounds. The chapter highlights its applications in food metabolomics, nutrimentalomics, sensomics, and volatilomics, showing how it enhances the reliability and depth of food-omics research. By providing examples of its capabilities, the chapter demonstrates how GC×GC contributes to more precise food quality assessments and the identification of both known and unknown components, making it an essential tool in modern food analysis.

The **Experimental Section** is organized in two parts. The first one focuses on the translation or adaptation of conventional or official methods to greener practices (Chapters 6, 7, and 8). In the second part, smart and efficient data processing work-flows, based on AI principles and tools, are explored to maximize the information potential of single analytical measurements performed with high-density information platforms as GC×GC-MS(/FID).

In the framework of the green transition of analytical methods, **Chapter 6** explores the adaptation of EPA Method 8260D for the detection of volatile organic compounds (VOCs) in environmental samples, with a focus on improving sustainability in GC-MS workflows. The chapter emphasizes the use of purge-and-trap (P&T) techniques for VOC enrichment and evaluates greener alternatives, such as hydrogen as a carrier gas and nitrogen as purge gas. By presenting these sustainable approaches, the chapter highlights the importance of modernizing VOC detection methods, promoting eco-friendly practices while maintaining analytical precision and regulatory compliance.

Chapter 7 focuses on the development and validation of a fast-GC method for the analysis and accurate quantification of four linear saturated aldehydes in hazelnuts. The study highlights the method's ability to achieve high throughput while maintaining the required performance standards, such as repeatability, accuracy, and sensitivity. By optimizing multiple headspace solid-phase microextraction (MHS-SPME) coupled with fast GC-MS, the chapter demonstrates how the method can efficiently yet accurately quantify aldehydes like hexanal, heptanal, octanal, and nonanal, which are key markers of oxidation in hazelnuts. The validation process confirms the method's reliability, making it suitable for high-volume industrial applications where rapid analysis and precise results are crucial for quality control.

Chapter 8 discusses the contribution to the European project "EVOQUE," which aims to develop a compact, cost-effective analyzer for detecting VOCs using GC systems integrated with a Quartz-Enhanced PhotoAcoustic Spectroscopy (QEPAS) detector. Building on the laboratory-based methods, of which an example is provided in the previous chapter, this chapter represents the future direction of VOC detection by working towards the development of a portable system capable of performing at-line analyses directly in production lines. The focus is on optimizing the GC-QEPAS system to operate with nitrogen as a carrier gas, ensuring compatibility with the QEPAS detector. Preliminary results demonstrate the system's sensitivity, linearity, and chromatographic efficiency, marking a

significant step toward creating a sustainable and affordable VOC detection solution for industrial applications.

Chapter 9 applies quantitative volatilomics using GC×GC-MS/FID to analyze the volatilome of hazelnuts, focusing on the characterization of certain features such as origin, post-harvest practices, and storage conditions. By using MHS-SPME, the study identifies key volatile compounds and correlates them with quality parameters. The integration of artificial intelligence (AI) tools enables the development of chemometric models for spoilage detection, rancidity assessment, and origin tracing. These findings offer valuable insights for improving product quality and consistency in the food industry.

The integration of Computer Vision (CV), an AI approach, with GC×GC to analyze complex chemical data in food quality assessments is explored in **Chapter 10**. Through CV, the study demonstrates how more straightforward yet efficient methods can be developed to analyze chemical signatures in food samples. By using pattern recognition algorithms, the chapter introduces a novel workflow that combines untargeted and targeted fingerprinting, enabling a deeper understanding of the biological phenomena related to specific chemical patterns. This approach enhances the analysis of volatile compounds in food, offering a more efficient and accurate means of food quality control compared to traditional methods.

Chapter 11 explores the translation of GC×GC-MS from thermal modulation to differential-flow modulation, a development that allows for the accurate quantification of volatile compounds of maize silage using predicted relative response factors (RRF). This approach enhances the efficiency of sample analysis while maintaining the analytical performance required for accurate food quality assessments. The ability to analyze complex biological samples more quickly and efficiently opens up new opportunities for monitoring silage quality and stability over time, with potential applications extending to broader food and agricultural quality assessments, contributing to more sustainable and effective quality control practices.

The last chapter (**Chapter 12**) investigates the application of GC×GC coupled with MS and flame ionization detection (FID) for the analysis of extra virgin olive oil (EVOO) quality. By focusing on volatile compounds, the study identifies markers that distinguish between different olive cultivars and cultivation methodologies. The research highlights how untargeted fingerprinting and quantitative volatilomics can define robust markers for EVOO quality, including cultivar-specific markers and indicators of cultivation methods. The findings contribute to a better understanding of how volatile profiles evolve during olive oil production, offering valuable insights for quality control and authenticity verification, essential for maintaining the integrity of EVOO in the marketplace.

In summary, this thesis bridges foundational green chemistry principles with cutting-edge analytical techniques and artificial intelligence. By addressing sustainability challenges and advancing methodological innovations, it provides a comprehensive framework for modernizing food analysis while pragmatically adhering to the principles of GAC.

Introductory Section

Chapter 1

Green Analytical Chemistry

This chapter explores the principles of Green Analytical Chemistry (GAC) and the tools available for assessing the sustainability and environmental impact of analytical methods. It delves into key metrics such as the Analytical GREENess (AGREE) metric, the Green Analytical Procedure Index (GAPI), and the National Environmental Methods Index (NEMI), emphasizing their roles in evaluating greenness across various dimensions of analytical processes. The concept of White Analytical Chemistry (WAC) is also discussed, integrating considerations of safety and sustainability alongside greenness. By examining these frameworks and their applications, the chapter highlights the progress and challenges in adopting environmentally conscious practices within analytical sciences.

1.1 Historical overview and concept of Green Analytical Chemistry

The issue of the "environmental impact" of chemical substances first entered public discourse and gained full recognition as a significant problem in the early 1960s. Prior to this period, environmental regulations governing the production, use, and disposal of chemicals were virtually nonexistent [1]. It was not uncommon for chemical substances to be released directly into the atmosphere, water bodies, and lands. Furthermore, there was a deep lack of scientific understanding regarding the hazards posed by these chemicals and their potential effects on both environmental and human health. It was within this context that the initial negative consequences of chemical misuse began to emerge, prompting widespread public concern. This, in turn, brought to the development of legislative frameworks and regulatory measures aimed at controlling the manufacture, management, and disposal of chemical substances. The advancement of environmental regulations led to a greater focus on treating waste before its release or addressing it after release to minimize risks to human health and the environment. However, the stricter limitations on the use of certain chemicals also created strong incentives for industries and chemists (faced with significant pressure not only to decrease the release of toxic substances into the environment but also to reduce the use of hazardous chemicals) to seek out replacements, substitutes, or alternative solutions [2].

The implementation of analytical procedures can pose significant environmental risks and potential health hazards to operators, making it critical to consider the broader implications of these methods. Consequently, from both an environmental and an economic perspective, it is essential to focus on the inherent risks associated with the types of samples, the quantities of reagents and solvents used, the energy consumption linked to increasingly sophisticated instruments, and, importantly, the waste and emissions generated throughout the analytical process [3].

A key development in the conceptualization of Green Analytical Chemistry (GAC) occurred in 1993 when the potential of combining photo-assisted degradation with flow injection analysis was explored for the spectrometric determination of phenolic compounds, thus demonstrating how analytical methods could be adapted to reduce environmental impact and paving the way for the broader adoption of green chemistry principles in analytical practices [4]. Another significant conceptual advance was the introduction of the term "clean waste" as an alternative to "waste". This term suggested a shift in thinking, emphasizing the reduction of environmental impact through additional chemical processing that could produce waste less harmful. The idea of "clean waste" marked the early development of the clean analytical chemistry concept, which focuses on minimizing the negative environmental footprint of analytical procedures. Later, in 1995, further validation of these concepts came with the publication "Towards Environmentally Conscientious Analytical Chemistry through Miniaturization, Containment, and Reagent Replacement", which is widely regarded as the first formal articulation of the principles that would later define GAC [5]. The importance of reducing the environmental burden of analytical procedures through strategies such as

miniaturizing analytical systems, containing hazardous substances, and replacing toxic reagents with safer alternatives was highlighted.

Since these foundational developments, the adoption of green principles in analytical chemistry has accelerated rapidly, with improvements made not only in analytical procedures but also in the instrumentation used, reflected in the ongoing evolution of analytical methodologies, which continue to become more efficient, less resource-intensive, and less harmful to the environment. Moreover, the basic ideas for an integrated environmentally friendly approach in analytical chemistry have been established, emphasizing the need to consider the entire lifecycle of analytical methods from sample collection and preparation to analysis and disposal of by-products. These ideas advocate for a comprehensive approach that considers the environmental, economic, and social impacts of analytical practices. This approach not only aims to reduce the environmental footprint of chemical analyses but also aims to improve the safety and well-being of those involved in these processes[6].

In conclusion, the development and implementation of GAC highlights a growing recognition of the need to balance scientific innovation with environmental stewardship, ensuring that analytical techniques do not compromise the health of our planet or its inhabitants.

1.2 GAC principles

The 12 principles of GAC, developed by J. Namieśnik's research team, have provided a fundamental framework for guiding the practice of analytical chemistry toward greater environmental responsibility. These principles, when integrated with the 12 principles of Green Chemistry proposed by Paul Anastas [1], are collectively summarized in the publication "The 12 principles of green analytical chemistry and the SIGNIFICANCE mnemonic of green analytical practices" [7]. This approach serves as an excellent guide for transitioning from traditional analytical chemistry, which focuses primarily on enhancing the analytical features of methods, to a more comprehensive perspective that deeply considers the potential side effects on both human operators and the environment. In the following section, the twelve principles of Green Analytical Chemistry will be introduced and briefly explained, highlighting their significance in promoting environmentally conscious analytical practices.

1. **Direct analytical techniques should be applied to avoid sample treatment.** The emphasis on using direct analytical methods aims to bypass extensive sample preparation, which often involves harmful solvents and chemicals. By analyzing samples in their natural state, this approach not only minimizes chemical waste but also speeds up the analytical process. Direct techniques can include methods like spectroscopy or electrochemical sensors, which provide rapid and accurate results with minimal environmental impact. Direct techniques may not always be feasible due to the nature of the sample. However, it is preferable to employ analytical methods that require minimal sample manipulation prior to analysis.
2. **Minimal sample size and minimal number of samples are goals.** In analytical chemistry, less is more. Reducing the amount of sample needed for analysis not only

conserves valuable resources but principally reduces the generation of waste. This principle is especially important in fields like pharmacology or environmental science, where samples may be limited or precious. Advances in technology have enabled the development of highly sensitive instruments that can achieve precise measurements from minute sample quantities.

3. **In situ measurements should be performed.** Conducting analyses directly at the site where samples are collected eliminates the need for transportation and storage, which can introduce contamination and degrade sample integrity. This approach is particularly advantageous in environmental monitoring, where real-time data is crucial for assessing immediate conditions. Additionally, the use of portable and field-deployable instruments, which are becoming increasingly sophisticated, allows for comprehensive analysis on-site without the need for transporting samples to a laboratory.
4. **Integration of analytical processes and operations saves energy and reduces the use of reagents.** By integrating multiple steps of the analytical process, such as sampling, separation, and detection, into a single, streamlined operation, resource consumption, and analysis time can be significantly reduced. This approach promotes the development of multi-functional instruments capable of performing various tasks simultaneously, resulting in more efficient workflows and a lower environmental impact.
5. **Automated and miniaturized methods should be selected.** The shift towards automation and miniaturization is driven by the need for improved analytical processes. Automated systems enhance reproducibility by reducing human error and can operate more precisely and efficiently, leading to shorter analysis times. Additionally, miniaturized devices require smaller amounts of sample and reagents, which can further reduce processing time.
6. **Derivatization should be avoided.** Derivatization, often used to improve detection in analytical chemistry, typically involves additional chemicals that increase waste and complexity. By avoiding or minimizing derivatization, we simplify the analytical process and reduce the use of potentially harmful reagents. This principle encourages the development of techniques and instruments that can achieve high sensitivity and selectivity without the need for chemical modification of the analyte.
7. **Generation of a large volume of analytical waste should be avoided and proper management of analytical waste should be provided.** The generation of large volumes of organic waste is a significant environmental concern in traditional analytical chemistry. GAC promotes strategies such as solvent recycling, waste minimization protocols, and the use of greener solvents (e.g., water or ethanol) to reduce the environmental footprint of analytical processes. Proper waste management practices are essential to ensure that any waste generated is handled responsibly.
8. **Multi-analyte or multi-parameter methods are preferred versus methods using one analyte at a time.** Methods capable of analyzing multiple analytes or parameters in a single run are preferred because they reduce the need for multiple analyses, thereby saving time, energy, and resources. This approach is particularly useful in

complex matrices, such as environmental samples, biological fluids, or food samples, where multiple constituents need to be monitored simultaneously. Techniques like mass spectrometry, multiplex assays, and multidimensional approaches are examples of multi-analyte methods that align with this principle.

9. **The use of energy should be minimized.** Reducing the energy consumption of analytical methods is critical for their sustainability. Energy-efficient instruments and processes not only lower operational costs but also contribute to reducing the environmental impact of the laboratory. This principle encourages the adoption of methods that operate at lower temperatures, require shorter analysis times, or utilize energy-efficient technologies.
10. **Reagents obtained from renewable sources should be preferred.** The shift towards using reagents derived from renewable sources supports the broader goal of sustainability in chemical processes. Renewable reagents, such as those sourced from plants or other natural materials, reduce dependency on non-renewable resources and often have a lower environmental impact. This principle promotes the development and use of greener chemistry practices in analytical laboratories.
11. **Toxic reagents should be eliminated or replaced.** Toxic reagents pose significant risks to both human health and the environment. This principle calls for the elimination of such substances or their replacement with safer, less hazardous alternatives. The use of green solvents, non-toxic indicators, and safer reaction conditions are examples of how this principle can be applied. By reducing the toxicity of reagents, we protect the health of the laboratory team and the environment.
12. **The safety of the operator should be increased.** The safety of those conducting analytical procedures is crucial. This principle emphasizes designing methods and selecting reagents that minimize risk to the operator, including reducing exposure to hazardous chemicals and ensuring safe working conditions. Safety can also be enhanced through proper training, the use of personal protective equipment, and the implementation of safety protocols in the laboratory.

Despite the broad acceptance of GAC's environmental and economic benefits, the aspect of user safety has often been overlooked, being frequently confused with environmental and cost considerations. It is evident that strategies such as automation and miniaturization play a crucial role in enhancing the safety of laboratory personnel. By reducing the need for direct handling of samples, reagents, and solvents, these strategies minimize exposure to potentially hazardous substances, thereby improving safety. Simultaneously, automation and miniaturization can significantly reduce costs and decrease the environmental footprint associated with traditional analytical methods. A third critical factor in reducing the impact of analytical methods on both operator safety and the environment is the advanced treatment of analytical data through the application of chemometric techniques. Chemometrics, which involves the comprehensive exploration of analytical signals, aligns closely with the eighth principle of GAC. This principle advocates for the use of multianalyte or multiparameter methods rather than focusing on single-analyte analyses. By leveraging chemometric approaches, researchers can extract multiple pieces of information from a single analysis

without the need for extensive sample preparation or the use of different instruments. This ability to maximize information without additional resource consumption and minimize the environmental and safety risks associated with sample handling and preparation makes chemometrics a powerful tool in advancing the goals of Green Analytical Chemistry [8].

The majority of the principles and concerns are centered around the sample preparation and treatment step. This is because the sample preparation stage is one of the most critical phases in analytical chemistry, posing significant risks to both the operator and the environment. These risks primarily derive from the use of hazardous reagents and solvents, which can be toxic, flammable, or otherwise dangerous. Sample treatment includes a range of processes such as sample digestion for inorganic analysis, analyte extraction for organic analysis, and specific treatments like preconcentration or clean-up to remove interferences. This step is often regarded as the bottleneck of analytical procedures due to its resource-intensive nature and its impact on both the accuracy and reliability of analytical results, which typically consumes substantial amounts of resources, including energy, solvents, and time, and implies significant financial costs. Moreover, it is frequently the primary source of analytical errors, contributing to issues such as contamination, analyte loss, or degradation, all of which can compromise the integrity of the analytical data [9].

The selection of an appropriate sample treatment method is influenced by various factors, including the concentration of the analyte within the sample, the nature of analyte-sample matrix interactions, and the specific requirements of the analytical instrumentation used to measure target analytes at the desired concentration levels. Historically, green considerations have been largely overlooked in the development of analytical methods, with past approaches primarily focused on achieving optimal analytical performance metrics, such as sensitivity, accuracy, and precision. This narrow focus often led to the use of methods that were resource-intensive, hazardous, and environmentally damaging. As previously mentioned, the past decades have witnessed a significant paradigm shift, which led to the recognition that the use of direct methods, which eliminate the need for extensive sample treatment, represents the greenest strategy available. Direct methods minimize the consumption of reagents and solvents, reduce energy use, and lower the risk of exposure to hazardous chemicals. However, the complete elimination of sample treatment is not feasible in many analytical scenarios. The need for sample preparation arises from several factors, including the heterogeneity of the sample, which requires homogenization or extraction to obtain a representative analytical portion, the presence of matrix components that interfere with the accurate determination of target analytes, necessitating clean-up procedures to remove these interferences, issues related to the sensitivity of the analytical method, and difficulties in calibrating direct methods, which may require the use of reference materials or additional calibration strategies to ensure accuracy [10].

Advancements in this area include the development of more efficient solvent-free extraction techniques, the use of benign reagents derived from renewable resources, and the implementation of advanced technologies such as solid-phase microextraction, supercritical fluid extraction, and microwave-assisted extraction. These methods offer the potential to

reduce the use of solvents, lower energy consumption, and minimize waste production, all while delivering high-quality analytical performance. The adaptations are susceptible to the nature of the sample under study: for instance, specifically for gaseous samples, the main challenges are primarily related to the inherent difficulties associated with handling gas samples, which must be carefully managed to prevent leaks and contamination. Additionally, the storage of gaseous samples can be complex due to their often low stability and the necessity to maintain their integrity over time. Moreover, many target analytes in gaseous samples are typically present at very low concentrations, requiring sensitive analytical methods to detect and quantify them accurately [11]. When it comes to determining both organic and inorganic analytes in the vapor phase of air samples, direct methodologies are often employed. These methods are considered the greenest analytical procedures because they minimize the use of additional reagents and reduce the need for extensive sample preparation steps [12]. Nevertheless, in many cases, a sampling step is necessary to concentrate the target analytes. Two main sampling strategies are used for this purpose: active sampling and passive sampling. The first one involves the active concentration and stabilization of target analytes within an extractant phase, which can be either liquid or solid. Solid-phase extraction (SPE) is widely utilized in active sampling for the detection of various organic pollutants. This method employs different types of sorbents, including carbon-based materials (such as carboxen and carbograph), organic polymers (such as Tenax, Chromosorb, and XAD), and inorganic materials (such as Florisil and silicate molecular sieves) [13]. Analytes retained in sampling tubes are typically thermally desorbed and directly analyzed using gas chromatography (GC). This approach avoids the use of organic solvents and achieves moderate desorption times, usually ranging from 10 to 30 minutes. Passive sampling, on the other hand, allows the analyte to move freely from the air into a collecting medium, driven by the difference in chemical potentials until equilibrium is achieved. This strategy enables long-term sampling studies without the need for pumps, providing a time-weighted average concentration of the target analytes. Various passive sampling devices are employed for this purpose, with the larger use being represented by semipermeable membrane devices, *i.e.*, solid-phase microextraction (SPME) fibers and arrows, but also some polyurethane foams, diffusive badges, Radiello diffusive samplers, *etc.* [14][15].

When it comes down to liquid samples, this has traditionally been accomplished using liquid-liquid extraction (LLE) and solid-phase extraction (SPE) methods. While these techniques have been effective in isolating and concentrating analytes, they come with significant drawbacks. These conventional methods are often associated with high costs in terms of time, energy, and the use of solvents and reagents. Moreover, they generate substantial amounts of waste, contributing to considerable environmental impact. In response to these concerns, the development of environmentally friendly extraction methods has advanced rapidly over the past few decades, with a strong emphasis on automation and miniaturization of analytical techniques. Flow systems reduce the direct contact between samples, reagents, and operators, thereby enhancing occupational safety, increasing sample throughput, and improving the precision and reproducibility of analytical results. At the same time, these systems reduce the consumption of reagents, lower the risk of contamination, and

minimize the generation of waste and emissions. Various flow-based sample treatment procedures have been developed, including flow injection analysis (FIA), sequential injection analysis (SIA), multicommutation techniques, lab-on-a-valve systems, and lab-on-a-chip devices. These approaches have demonstrated excellent analytical performance while incorporating environmental considerations [16]. Finally, for the extraction of organic compounds from solid samples, classical methods have typically employed solvent extraction techniques enhanced by shaking or heating, such as Soxhlet extraction. However, more advanced extraction techniques have emerged as valuable alternatives, including ultrasound-assisted extraction (UAE), microwave-assisted extraction (MAE), pressurized solvent extraction (PSE), and supercritical fluid extraction (SFE). These high-performance extraction systems have gained widespread acceptance in routine analytical laboratories worldwide. Although these methods often require the acquisition of sophisticated and expensive equipment, they provide highly efficient extraction of target analytes within significantly shorter analysis times (reducing the duration from hours to minutes), consume minimal solvent volumes (typically less than 100 mL per sample), and substantially decrease occupational hazards associated with handling solvents and reagents [17]. Following the principle that was previously mentioned – *i.e.*, minimal treatment is the best treatment, headspace (HS) extraction emerges as an outstanding green technique for the direct and automated analysis of volatile organic compounds (VOCs) from solid samples. In HS extraction, no solvent is required to extract the target compounds. Instead, the sample is heated to release VOCs, which are then directly introduced into a gas chromatography (GC) system for analysis. This method is particularly advantageous due to its simplicity, automation potential, and minimal environmental footprint. The combination of headspace extraction with solid-phase microextraction (HS-SPME) can further enhance the sensitivity of the method, making it suitable for the detection of semi-volatile organic compounds. Moreover, vacuum-assisted HS-SPME has been proposed to improve extraction kinetics and sensitivity by utilizing short extraction times and mild temperatures, thus reducing energy consumption and preserving the integrity of the sample [18][19].

The green methodologies can significantly enhance the environmental profile of powerful instrumental techniques, particularly those requiring precise sample treatment. Notable examples include chromatographic techniques and methods involving inductively coupled plasma (ICP) spectrometry, which demand high-efficiency digestion protocols to mitigate matrix effects [20]. For instance, ICP measurements can be made more environmentally friendly by incorporating laser ablation systems for the direct analysis of solid samples. However, while this approach reduces the need for chemical reagents, it often compromises the quality of the analytical data, limiting it to a semi-quantitative level due to challenges in achieving accurate calibration. In many cases, the use of reference materials becomes necessary to address these calibration difficulties.

2.3 GAC metrics and tools

Green chemistry has emerged as a central driving force in both laboratory and industrial settings, fueling the pursuit of sustainable development across various sectors. The

foundation of this movement is the twelve principles of green chemistry, which serve as a comprehensive guide for those aiming to adopt eco-friendly practices in chemical processes. These principles provide a structured framework for reducing environmental impact, minimizing waste, and enhancing the overall sustainability of chemical operations. They have become a cornerstone for chemists working in various fields such as organic synthesis, chemical engineering, and analytical chemistry, guiding them toward more responsible and sustainable practices.

Since the formalization of these principles in 1998, they have been systematically integrated into chemical processes, shaping the way chemists approach sustainability. However, the challenge of accurately assessing the "greenness" of chemical processes remains a significant obstacle. It is well-recognized that processes that cannot be measured cannot be effectively managed or improved. Therefore, the development and application of metrics to evaluate the environmental impact of chemical processes are crucial. These metrics serve as tools to compare the greenness of existing and newly developed solutions, enabling chemists to choose the most sustainable alternatives available [21].

In most industrial processes, the primary objective is to produce a specific product, with any inputs that do not convert into the product being classified as waste. The principles of green chemistry provide a methodology for measuring the impact of chemical reactions, associated products, and generated waste. However, when it comes to analytical chemistry, the situation differs significantly from traditional synthetic chemistry or large-scale industrial processes. In analytical chemistry, the outcome is not a physical product but rather analytical results, which are obtained through various processes involving solvents, reagents, and other inputs. Although these substances are used in much smaller quantities than in industrial processes, they still contribute to environmental pollution, albeit on a more dispersed scale [22].

To assist researchers in assessing the greenness of their analytical methods, several metrics have been developed. These metrics offer a systematic approach to evaluating the environmental impact of analytical procedures, allowing chemists to make informed decisions about the sustainability of their methods [10][23].

This paragraph provides an overview of the most widely used GAC metrics, offering insights into their application and significance. By understanding these tools, researchers can better evaluate and enhance the greenness of their analytical methods, contributing to the broader goal of sustainable chemistry. This discussion not only underscores the importance of adopting green practices in analytical chemistry but also highlights the need for continuous innovation and improvement in the field to meet the evolving demands of sustainability. Based on this overview, the following sections will detail the principal green metric tools, exploring their specific applications and how they contribute to the evaluation and improvement of greener analytical methods.

- **National Environmental Methods Index (NEMI).** NEMI, introduced in 2002, serves as a searchable database for environmental techniques, protocols, and methods. It

helps scientists locate and compare analytical methods for environmental monitoring. The NEMI tool is visually represented through a pictogram divided into four segments (Persistent Bioaccumulative Toxic (PBT), Hazardous, Corrosive, and Waste) each color-coded, from green to red, based on the assessment. While NEMI is simple to use, its reliance on manual procedures without quantification metrics limits its precision and applicability in evaluating the comprehensive greenness of analytical methods [24].

- **Modified NEMI.** The Modified NEMI, developed in 2009 by Raynie and Driver, enhances the original NEMI by adding a fifth segment to the pictogram, which now includes Health, Safety, Quality of Waste, Environmental Impact, and Energy Usage. The assessment criteria are color-coded (green, yellow, or red) based on the method's compliance with these parameters. Although it provides a broader evaluation, the Modified NEMI remains limited by the lack of software for quantification, which restricts its utility in offering precise and measurable assessments of method greenness [25].
- **Analytical Eco-Scale.** The Analytical Eco-Scale, introduced in 2006, is a semi-quantitative tool that assesses the eco-friendliness of analytical methods based on penalty points assigned for the use of hazardous chemicals, energy consumption, occupational hazards, and waste generation. A score above 75 is considered excellent, while scores below 50 indicate non-eco-friendly methods. Despite its effectiveness in providing a measurable assessment of greenness, the absence of automated calculation software makes it a labor-intensive and potentially error-prone tool [26][27].
- **Green Analytical Procedure Index (GAPI).** GAPI, developed in 2018, offers a visual representation of the greenness of analytical methods from sample collection to final determination. GAPI uses a color-coded pentagonal pictogram to qualitatively assess various aspects of the method, including sample sourcing, method type, sample preparation, reagent use, and instrumentation. While GAPI provides a comprehensive view of method greenness, it lacks quantifiable metrics and can be challenging to apply due to the absence of software support [28].
- **Complex GAPI.** Complex GAPI is an updated version of GAPI that includes a hexagonal pictogram to provide a more detailed assessment. This tool combines GAPI with software support, allowing for a more refined evaluation of method greenness. However, similar to GAPI, it primarily provides a qualitative assessment and lacks a clear numerical value, which can complicate the interpretation of results [29].
- **Analytical GREEnness (AGREE).** The AGREE tool, developed by the Gdańsk University of Technology, is a modern software tool that quantitatively evaluates the greenness of analytical methods based on the twelve principles of green chemistry. AGREE uses a circular pictogram divided into twelve segments, each representing a different green chemistry principle. The tool offers both qualitative and quantitative assessments, with scores ranging from 0 to 1, making it one of the more advanced and comprehensive tools available for evaluating analytical method greenness [30].

- **AGREEprep.** AGREEprep is a specialized version of the AGREE tool, developed in 2022, and designed specifically to assess the greenness of sample preparation procedures. AGREEprep evaluates ten criteria, including solvent use, sample size, waste generation, and energy consumption, assigning scores from 0 to 1. This tool provides a focused assessment of the sample preparation stage, making it particularly useful for researchers aiming to enhance the sustainability of this critical aspect of analytical chemistry[31], [32].

Representative outputs obtained with the metrics described above are schematically shown in **Figure 1.1**.

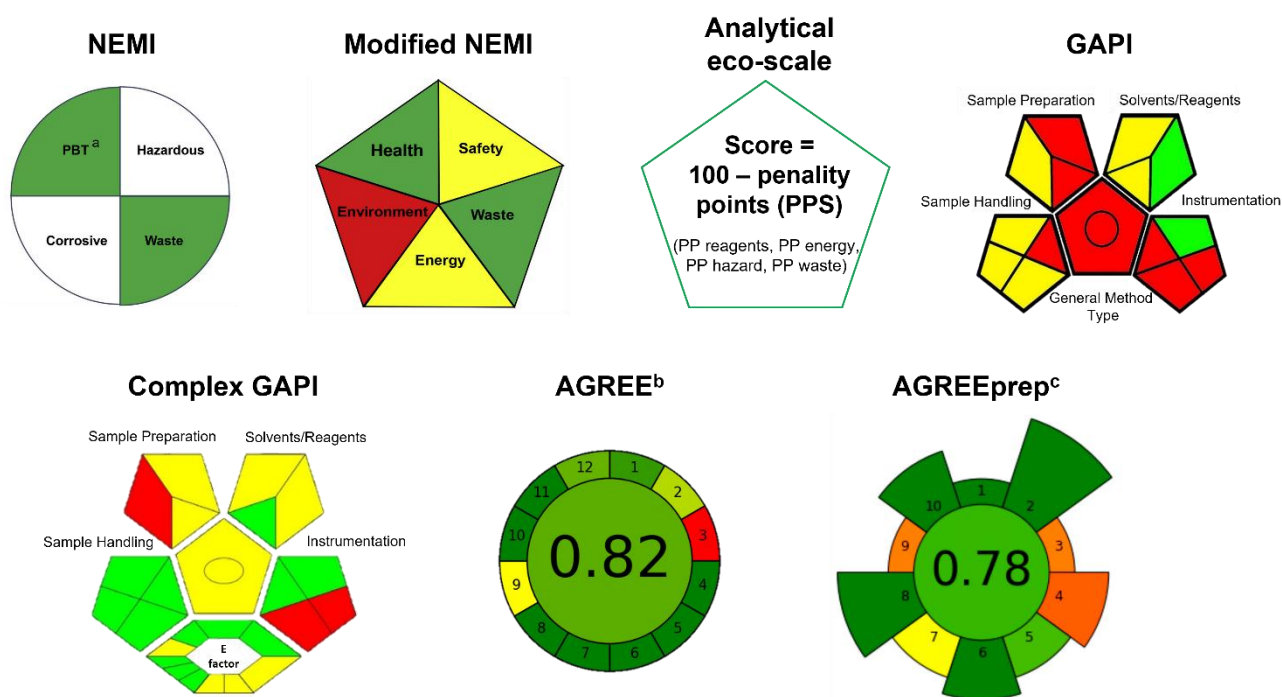


Figure 1.1: Representative output of the main metrics developed to critically evaluate analytical procedure in the chemical field.

^a Persistent, bioaccumulative, toxic

^b Criterion considered by the metric: 1. Direct analytical techniques should be applied to avoid sample treatment, 2. Minimal sample size and minimal number of samples are goals, 3. Measurements should be performed in situ, 4. Integration of analytical processes and operations saves energy and reduces the use of reagents, 5. Automated and miniaturized methods should be selected, 6. Derivatization should be avoided, 7. Generation of a large volume of analytical waste should be avoided, and proper management of analytical waste should be provided, 8. Multi-analyte or multi-parameter methods are preferred versus methods using one analyte at a time, 9. The use of energy should be minimized, 10. Reagents obtained from renewable sources should be preferred, 11. Toxic reagents should be eliminated or replaced, 12. Operator's safety should be increased [30].

^c Criterion considered by the metric: 1. Sample preparation placement, 2. Hazardous material, 3. Sustainability and renewability of materials, 4. Waste, 5. Size economy of the sample, 6. Sample throughput, 7. Integration and automation, 8. Energy consumption, 9. Post-sample preparation configuration for analysis, 10. Operator's safety.

2.4 *White Analytical Chemistry*

The field of green analytical chemistry (GAC) has been instrumental in advancing green practices within analytical laboratories. Over the past two decades, GAC has provided a framework that emphasizes the reduction of environmental impact through principles such as minimizing the use of hazardous reagents, conserving energy, and reducing waste generation. These principles have impacted significant progress in making analytical processes more sustainable. However, as the field has matured, it has become increasingly clear that a singular focus on environmental aspects can sometimes be in contrast with other critical dimensions of analytical methods, such as their functionality, accuracy, and overall practicality.

One of the key challenges that has emerged from the implementation of GAC principles is the need to balance the greenness of a method with its effectiveness and practical utility. Analytical methods are not just evaluated on their environmental impact; they must also meet stringent criteria related to accuracy, precision, sensitivity, and reproducibility. Furthermore, these methods must be economically viable, time-efficient, and straightforward enough to be implemented reliably across various contexts. These additional demands highlight the complexity of developing methods that are not only green but also robust and versatile enough to be useful in real-world applications [33].

To address these challenges, the concept of White Analytical Chemistry (WAC) has been introduced as an evolution and extension of GAC. WAC aims to merge the environmental objectives of GAC with the functional and practical requirements of analytical methods, creating a more balanced and comprehensive approach. The WAC framework is built on the integration of three key dimensions: analytical efficiency (represented by the color red), environmental friendliness (represented by green), and practical and economic efficiency (represented by blue). This RGB model is designed to evaluate methods in a way that ensures they are not only sustainable but also fit for purpose in their intended applications [34].

The novelty of WAC lies in its comprehensive approach, which goes beyond the environmental considerations of GAC to include a broader set of criteria that are essential for the development of effective analytical methods. The twelve principles of WAC, analogous to the twelve principles of GAC, are structured to address these additional dimensions. For instance, while GAC emphasizes the reduction of toxic reagents and waste, WAC also considers the importance of minimizing costs, maximizing time efficiency, and ensuring operational simplicity. By doing so, WAC provides a more complete structure for evaluating analytical methods, one that recognizes the interconnectedness of environmental, practical, and economic factors [35].

A critical component of WAC is the concept of "whiteness," which is used as a quantitative measure of how well an analytical method balances the three key dimensions of the RGB model. The "whiteness" of a method reflects its overall sustainability, taking into account not just its environmental impact but also its effectiveness and practicality. This approach acknowledges that a method can be green without being fully functional or

economically feasible, and vice versa. The goal of WAC is to ensure that all three aspects are optimized, resulting in a method that is both sustainable and practically useful.

The RGB 12 algorithm, a tool specifically designed for WAC, facilitates the evaluation of methods by allowing researchers to assess them against the twelve WAC principles. This algorithm provides a structured and transparent way to evaluate the "whiteness" of a method, making it easier to identify areas for improvement and to compare different methods in terms of their overall sustainability. The flexibility of the RGB 12 algorithm also allows for adjustments based on specific requirements, making it a versatile tool for a wide range of analytical applications [36].

The introduction of WAC represents a significant advancement in the quest for sustainable analytical practices. It builds on the foundations laid by GAC but expands the focus to include the full spectrum of factors that influence the development and application of analytical methods. This approach is closely aligned with the broader goals of Sustainable Development (SD), which seeks to achieve a balance between economic growth, social well-being, and environmental protection. By incorporating practical and economic considerations into the evaluation of analytical methods, WAC ensures that sustainability is not achieved at the expense of functionality or accessibility. Furthermore, WAC addresses some of the limitations inherent in a purely green-focused approach. While GAC has been successful in promoting environmentally friendly practices, it does not always account for the realities of method implementation, such as the need for high sensitivity, accuracy, and cost-effectiveness. WAC, by contrast, offers a more nuanced approach that recognizes the importance of these factors and integrates them into the overall assessment of a method's sustainability [35].

WAC is assessed using a tool that evaluates the different key parameters. A model of this tool is presented in **Figure 1.2**.

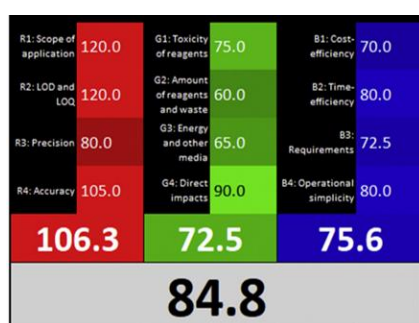


Figure 1.2. WAC model template.

In summary, the emergence of White Analytical Chemistry marks a crucial moment in the evolution of sustainable chemistry. It provides a comprehensive framework that balances environmental, practical, and economic considerations, ensuring that analytical methods are not only green but also effective and accessible. This approach is poised to play a crucial role in the ongoing effort to merge the demands of modern analytical science with the imperatives of environmental stewardship and sustainable development. By embracing the principles of WAC, researchers, and practitioners can contribute to the creation of analytical methods that

are truly fit for the future methods that meet the highest standards of sustainability without compromising on performance or practicality.

References

- [1] J. C. Anastas, Paul T. and Warner, *Green Chemistry theory and practice*, vol. 19, no. 5. 2016.
- [2] P. Tundo *et al.*, "Synthetic pathways and processes in green chemistry. Introductory overview," *Pure Appl. Chem.*, vol. 72, no. 7, pp. 1207–1208, 2000.
- [3] M. De La Guardia and S. Garrigues, "An ethical commitment and an economic opportunity," *RSC Green Chem.*, no. 13, pp. 1–12, 2011.
- [4] M. de la Guardia and S. Armenta, "Origins of green analytical chemistry," *Compr. Anal. Chem.*, vol. 57, no. 11, pp. 1–23, 2011.
- [5] M. de la Guardia and J. Ruzicka, "Towards Environmentally Conscientious Analytical Chemistry Through Miniaturization, Containment and Reagent Replacement," *Analyst*, vol. 120, no. February, p. 1995, 1995.
- [6] P. T. Anastas, "Green Chemistry and the role of analytical methodology development," *Crit. Rev. Anal. Chem.*, vol. 29, no. 3, pp. 167–175, 1999.
- [7] A. Gałuszka, Z. Migaszewski, and J. Namieśnik, "The 12 principles of green analytical chemistry and the SIGNIFICANCE mnemonic of green analytical practices," *TrAC Trends Anal. Chem.*, vol. 50, pp. 78–84, Oct. 2013.
- [8] S. Garrigues, S. Armenta, and M. de la Guardia, "Challenges in Green Analytical Chemistry," *Encycl. Inorg. Bioinorg. Chem.*, pp. 1–9, 2016.
- [9] J. Płotka-Wasyłka, M. Fabjanowicz, K. Kalinowska, and J. Namieśnik, "History and Milestones of Green Analytical Chemistry," pp. 1–17, 2019.
- [10] G. Emanuela, Ed., *Green Approaches for Chemical Analysis*, vol. 19, no. 5. 2016.
- [11] M. M. Moein, R. Said, F. Bassyouni, and M. Abdel-Rehim, "Solid phase microextraction and related techniques for drugs in biological samples," *J. Anal. Methods Chem.*, vol. 2014, 2014.
- [12] F. Karagulian, M. Gerboles, M. Barbieri, A. Kotsev, F. Lagler, and A. Borowiak, *Review of sensors for air quality monitoring*. 2019.
- [13] E. Woolfenden, "Sorbent-based sampling methods for volatile and semi-volatile organic compounds in air. Part 2. Sorbent selection and other aspects of optimizing air monitoring methods," *J. Chromatogr. A*, vol. 1217, no. 16, pp. 2685–2694, 2010.
- [14] F. A. Esteve-turrillas, A. Pastor, and M. De Guardia, *Passive Sampling of Atmospheric Organic Contaminants*, vol. 1, no. 1. Elsevier, 2012.
- [15] S. Armenta, M. De La Guardia, and F. A. Esteve-Turrillas, *Environmental applications (air)*. Elsevier Inc., 2019.
- [16] A. M. Bevanda, S. Talić, and A. Ivanković, "Flow Injection Analysis Toward Green Analytical Chemistry," pp. 299–323, 2019.

- [17] S. Armenta, F. A. Esteve-Turrillas, S. Garrigues, and M. de la Guardia, *Green Analytical Chemistry: The Role of Green Extraction Techniques*, vol. 76, no. 11. Elsevier Ltd, 2017.
- [18] E. Psillakis, "Vacuum-assisted headspace solid-phase microextraction: A tutorial review," *Anal. Chim. Acta*, vol. 986, pp. 12–24, 2017.
- [19] G. Bechis, M. A. Minteguiaga, B. Sgorbini, A. Marengo, P. Rubiolo, and C. Cagliero, "Make the Quality Control of Essential Oils Greener: Fast Enantioselective GC-MS Analysis of Sweet and Bitter Orange as a Case Study," *Molecules*, vol. 28, no. 17, 2023.
- [20] C. A. Bizzi, M. F. Pedrotti, J. S. Silva, J. S. Barin, J. A. Nóbrega, and E. M. M. Flores, "Microwave-assisted digestion methods: Towards greener approaches for plasma-based analytical techniques," *J. Anal. At. Spectrom.*, vol. 32, no. 8, pp. 1448–1466, 2017.
- [21] M. Sajid and J. Płotka-Wasyłka, "Green analytical chemistry metrics: A review," *Talanta*, vol. 238, no. October 2021, 2022.
- [22] L. P. Kowtharapu, N. K. Katari, S. K. Muchakayala, and V. M. Mariseti, "Green metric tools for analytical methods assessment critical review, case studies and crucify," *TrAC - Trends Anal. Chem.*, vol. 166, no. June, p. 117196, 2023.
- [23] M. Shi, X. Zheng, N. Zhang, Y. Guo, M. Liu, and L. Yin, "Overview of sixteen green analytical chemistry metrics for evaluation of the greenness of analytical methods," *TrAC - Trends Anal. Chem.*, vol. 166, no. August, p. 117211, 2023.
- [24] M. Tobiszewski, M. Marć, A. Gałuszka, and J. Namieśnik, "Green chemistry metrics with special reference to green analytical chemistry," *Molecules*, vol. 20, no. 6, pp. 10928–10946, 2015.
- [25] K. F. M. Amin, "Greenness-sustainability metrics for assessment smart-chemometric spectrophotometric strategy for evaluation of the combination of six gastric proton-pump inhibitors with two selected impurities," *MethodsX*, vol. 12, no. March, p. 102670, 2024.
- [26] K. Van Aken, L. Strekowski, and L. Patiny, "EcoScale, a semi-quantitative tool to select an organic preparation based on economical and ecological parameters," *Beilstein J. Org. Chem.*, vol. 2, pp. 1–7, 2006.
- [27] A. Gałuszka, Z. M. Migaszewski, P. Konieczka, and J. Namieśnik, "Analytical Eco-Scale for assessing the greenness of analytical procedures," *TrAC - Trends Anal. Chem.*, vol. 37, pp. 61–72, 2012.
- [28] J. Płotka-Wasyłka, "A new tool for the evaluation of the analytical procedure: Green Analytical Procedure Index," *Talanta*, vol. 181, no. January, pp. 204–209, 2018.
- [29] R. Th. El-Eryan, S. S. Toubar, A. A. Ashour, and M. S. Elshahed, "Application of analytical Eco-Scale and Complex-GAPI tools for green assessment of a new simple nanoparticle modified carbon paste electrode method for voltammetric determination of mosapride citrate in pharmaceutical dosage form and human plasma," *Microchem. J.*, vol. 178, no. December 2021, p. 107347, 2022.
- [30] F. Pena-Pereira, W. Wojnowski, and M. Tobiszewski, "AGREE - Analytical GREEnness

Metric Approach and Software,” *Anal. Chem.*, vol. 92, no. 14, pp. 10076–10082, 2020.

- [31] W. Wojnowski, M. Tobiszewski, F. Pena-Pereira, and E. Psillakis, “AGREEprep – Analytical greenness metric for sample preparation,” *TrAC - Trends Anal. Chem.*, vol. 149, p. 116553, 2022.
- [32] F. Pena-Pereira, M. Tobiszewski, W. Wojnowski, and E. Psillakis, “A Tutorial on AGREEprep an Analytical Greenness Metric for Sample Preparation,” *Adv. Sample Prep.*, vol. 3, no. May, p. 100025, 2022.
- [33] C. M. Hussain, C. G. Hussain, and R. Keçili, “White analytical chemistry approaches for analytical and bioanalytical techniques: Applications and challenges,” *TrAC - Trends Anal. Chem.*, vol. 159, p. 116905, 2023.
- [34] G. Bechis *et al.*, “Assessing the environmental and overall performance of gas chromatographic analyses. Development of a comprehensive evaluation framework and application to routine chiral analyses of fragrances as a case study,” *Sustain. Chem. Pharm.*, vol. 35, no. June, p. 101217, 2023.
- [35] P. M. Nowak, R. Wietecha-Posłuszny, and J. Pawliszyn, “White Analytical Chemistry: An approach to reconcile the principles of Green Analytical Chemistry and functionality,” *TrAC - Trends Anal. Chem.*, vol. 138, 2021.
- [36] P. M. Nowak and P. Kościelniak, “What color is your method? adaptation of the rgb additive color model to analytical method evaluation,” *Anal. Chem.*, vol. 91, no. 16, pp. 10343–10352, 2019.

Chapter 2

Gas Chromatography-Mass Spectrometry

This chapter focuses on gas chromatography-mass spectrometry (GC-MS) and its wide range of applications in analytical chemistry. It will explore the increasing use of alternative carrier gases to helium, prompted by its scarcity and rising costs. Special attention will be given to hydrogen and nitrogen as promising alternatives, evaluating their advantages and challenges in different contexts. Additionally, the chapter will cover the development of portable GC systems, which are gaining popularity due to their ability to perform on-site analyses with efficiency and flexibility, expanding the scope of GC-MS in various fields.

2.1 History, evolution, and impact of capillary Gas Chromatography-Mass Spectrometry in modern analytical chemistry

Capillary Gas Chromatography coupled with Mass Spectrometry (GC-MS) is widely regarded as one of the most powerful and versatile analytical techniques in the field of chemistry, particularly for the analysis of complex mixtures of organic, thermal stable, volatile and semi-volatile compounds. The evolution of GC-MS from its early conceptual stages to its current status as a cornerstone of analytical chemistry is a story of continuous innovation and adaptation, driven by the need to address increasingly complex analytical challenges across various scientific disciplines.

The origins of GC-MS can be traced back to the mid-20th century when both GC and MS were evolving as independent techniques. The combination of these two techniques was a natural progression, as both excelled in their respective areas: GC in separating volatile and semi-volatile compounds based on their physical and chemical properties, and MS in identifying and quantifying these compounds based on their mass-to-charge ratios. The initial attempts to couple GC with MS met significant challenges, particularly due to the technological limitations of the time. Early gas chromatographs used packed columns that required high flow rates, which were incompatible with the vacuum conditions necessary for mass spectrometers to operate effectively. However, these obstacles were gradually overcome with the introduction of capillary columns in gas chromatography. Capillary columns, which are much narrower than packed columns, allowed for lower flow rates, thus reducing the pressure incompatibility. Capillary columns also offered enhanced resolution, allowing for better separation of complex mixtures, which in turn facilitated more accurate and detailed mass spectral analysis [1], [2].

The real breakthrough for GC-MS came with the development of reliable interfaces that could effectively link the gas chromatograph to the mass spectrometer. These interfaces allowed the smooth transfer of separate analytes from the GC to the MS without compromising the integrity of the analysis. The success of this integration was evident in the work of pioneers like Roland Gohlke and Fred McLafferty, who demonstrated the first practical application of GC-MS for the analysis of organic compounds. Their work laid the foundation for the widespread adoption of GC-MS in various fields of research and industry [3].

GC-MS operates by first separating the components of a mixture in the gas chromatograph. As the mixture is heated, the volatile compounds vaporize and travel through the GC column, where they are separated based on their relative vapor pressures due to the different interactions with the column's stationary phase. Each component eluates the column at a different time, known as the retention time, and enters the mass spectrometer. In the MS, the separated components are ionized, and the resulting ions are detected and analyzed based on their mass-to-charge ratios. The mass spectrometer produces a mass

spectrum for each component, which serves as a unique "fingerprint" that can be used to identify the compound. This combination of separation and identification makes GC-MS an incredibly powerful tool for analyzing complex mixtures.

One of the major strengths of GC-MS is its dual capability for both qualitative and quantitative analysis. Initially, GC-MS was primarily used for the qualitative identification of unknown compounds. The ability to obtain a mass spectrum for each separated component allowed researchers to identify compounds with a high degree of confidence, even in complex mixtures. This qualitative capability was particularly valuable in fields like natural products research, where the identification of unknown compounds is often a key objective.

Quantitative analysis, on the other hand, was traditionally performed using other types of detectors, such as Flame Ionization Detectors (FID) or Thermal Conductivity Detectors (TCD), which were considered more suitable for precise quantification. However, the development of more advanced mass spectrometers and data processing systems has expanded the role of GC-MS to include quantitative analysis as well. Modern GC-MS systems are now capable of providing both qualitative and quantitative data in a single analysis, making them highly efficient and versatile tools for a wide range of applications.

In the experimental section of this thesis, the quantification of target molecules in the volatile profile of raw hazelnuts will be discussed in detail [4]. This example will focus on the use of the GC-MS technique with the detector set to *Scan* acquisition mode. This approach will collect the Total Ion Current (TIC), i.e. for each ion fragment a detector response is recorded and for each acquisition point (Scan) the total response is collected and used to reconstruct the chromatogram. By scan acquisition, a comprehensive assessment of the compounds present in the sample is possible, enabling simultaneous monitoring of multiple analytes in a single run. On the other side, the Selected Ion Monitoring (SIM) acquisition mode offers alternative possibilities. By SIM acquisition, just selected ions are monitored and their response recorded by the MS system. Many of the commercial GC-MS systems allow, nowadays, a simultaneous yet consecutive acquisition in multiple or mixed mode by "SIM-scan". When comparing SIM and scan mode, each has its own advantages and limitations. SIM mode is particularly beneficial for improving sensitivity and selectivity as it focuses on a limited number of target ions, which improves detection limits for certain compounds. However, the narrow focus means that other compounds of interest that have not been preselected for monitoring will not be detected. On the other hand, the Scan mode provides broader, more comprehensive detection, allowing the entire mass range to be captured. This can be invaluable when reprocessing previously acquired data, as any compounds that were not originally detected can still be subsequently identified as all potential targets are captured in the scan data. This flexibility makes the scan mode very advantageous for exploratory analyses, even if it comes with a slightly lower sensitivity compared to the SIM mode.

The technological advancements that have driven the evolution of GC-MS are numerous. Improvements in ionization techniques, such as Electron Ionization (EI) and Chemical Ionization (CI), have enhanced the sensitivity and specificity of mass spectrometric detection. The development of capillary columns with better stability, lower bleed, and higher efficiency has improved the quality of chromatographic separations, allowing for more accurate MS analysis, lower detection limits and high resolution separations. Additionally, the integration of advanced data processing systems has revolutionized the way GC-MS data is

analyzed and interpreted. The introduction of mini-computers in the 1960s and later the development of more powerful micro-computers enabled the rapid processing and interpretation of large volumes of GC-MS data, which was previously a major bottleneck in the analytical process.

The impact of these technological advancements is evident in the GC-MS instruments available today. These instruments are more compact, user-friendly, and capable of handling a wider range of applications than ever before. For example, while early GC-MS systems were large and required dedicated laboratory space, modern systems are often small enough to fit on a benchtop, making them accessible to a broader range of users and laboratories. The increased automation and ease of use of these systems have also made GC-MS more accessible to non-specialists, allowing a wider range of researchers and industries to benefit from its powerful analytical capabilities.

In conclusion, the development and evolution of GC-MS have transformed it into an indispensable tool in analytical chemistry, with applications spanning a wide range of scientific and industrial fields. The combination of gas chromatography and mass spectrometry offers unmatched capabilities for the separation, identification, and quantification of complex mixtures, making GC-MS a critical component in modern research and analysis. As technology continues to advance, the future of GC-MS looks promising, with ongoing innovations aimed at making the technique more accessible, efficient, and versatile. Whether through the development of portable systems, the exploration of alternative carrier gases, or the integration of advanced data processing techniques, GC-MS is poised to remain at the forefront of analytical science for years to come [5].

2.2 Mass Spectrometric detectors

When integrating MS with GC, the selection of the appropriate mass analyzer is crucial to achieving the desired analytical outcomes, particularly in terms of resolution, sensitivity, and specificity. GC is a powerful separation technique that excels at isolating compounds within complex mixtures, but it provides limited information about the chemical nature of the analytes. MS, on the other hand, complements GC by offering detailed structural information through the ionization and fragmentation of these separated compounds. This combination of GC-MS has become an indispensable tool in analytical chemistry, particularly for the qualitative identification and quantitative analysis of complex organic compounds [2][6]. This section is dedicated to commercially relevant systems available on the market. Additionally, ion mobility detectors will not be included in the discussion, as they do not provide diagnostic information on molecular structure. Hybrid ion mobility-MS solutions will also be excluded, as they are primarily dedicated to coupling with liquid chromatography. In this section, the various detectors will be presented according to a classification based on their resolution. Nevertheless, some instruments are borderline and could fit into multiple categories.

- **Quadrupole Mass Analyzer (QMS).** QMS analyzers are widely used due to their robustness, affordability, and ease of use. They function by filtering ions based on their mass-to-charge ratio (m/z) using a combination of direct current and radiofrequency

fields. Quadrupoles are particularly effective in targeted analyses where specific compounds are of interest. However, their relatively low resolution and mass accuracy can be limiting when it comes to distinguishing between compounds with very similar m/z values, which can be critical in complex mixtures.

- **Ion Trap Analyzer.** Ion traps, including both linear and quadrupole ion traps, offer the capability to perform multiple stages of mass spectrometry (MS^n), which is advantageous for structural elucidation. They trap ions within an electric field and selectively eject them for detection based on their m/z values. Ion traps are compact and versatile, making them suitable for portable GC-MS systems. However, they present certain disadvantages: ion traps generally offer lower resolution compared to TOF and QTOF analyzers, which can limit their effectiveness in applications requiring high-resolution data. Additionally, the lower acquisition frequency reduces the range of applications in which they can be effectively coupled like fast GC or GC $\times$ GC [7][8].

2.2.1. High-Resolution Mass Spectrometry

High-resolution mass spectrometry (HRMS) has revolutionized the field of analytical chemistry by offering unparalleled mass accuracy, sensitivity, and resolution, making it essential for a wide range of complex analytical tasks. In mass spectrometry, the resolution refers to the ability of the instrument to distinguish between ions with very similar mass-to-charge ratios. High-resolution systems are capable of distinguishing compounds that differ by a fraction of a Dalton, which is crucial for the accurate identification of compounds in complex matrices. This ability to differentiate between isobaric species, compounds with the same nominal mass but different exact masses, gives HRMS a distinct advantage over low-resolution instruments.

The high resolving power of HRMS makes it invaluable in fields such as proteomics, metabolomics, environmental science, and food safety, where the analysis often involves mixtures containing thousands of compounds. In these areas, the ability to provide exact mass measurements enables researchers to identify unknown compounds, confirm molecular formulas, and detect trace levels of contaminants [9], [10]. The first prototype of a high-resolution mass spectrometry analyzer was the magnetic sector, which operates based on the deflection of charged ions in a magnetic field. When ions enter the magnetic field, they are subjected to a force perpendicular to their velocity and the field direction, causing them to follow a curved trajectory; the radius of this curvature depends on their m/z , allowing for their separation and precise analysis. Although magnetic sector instruments are still commercially available, they are now used primarily for niche applications [11]. Today, the main high-resolution systems are the TOF and Orbitrap analyzers.

- **Time-of-Flight (TOF) Analyzer.** TOF analyzers offer significantly higher resolution and mass accuracy compared to quadrupoles. They measure the time it takes for ions to travel through a field-free region, with lighter ions reaching the detector faster than heavier ones. This characteristic allows TOF analyzers to handle a broad mass range, making them particularly useful in non-targeted analyses where the goal is to identify and characterize unknown compounds. While TOF analyzers are generally classified as high-resolution detectors, their actual performance depends on the specific

configuration of the instrument. The high-speed data acquisition capabilities of TOF analyzers make them also well-suited for high-throughput applications, such as proteomics or metabolomics, where large amounts of data are generated and need to be processed quickly [6].

- **Orbitrap Analyzers.** Orbitrap operates by trapping ions in an electrostatic field, measuring their oscillation frequencies, which correlate with the ions' m/z values. Orbitrap instruments are favored in proteomics, metabolomics, and pharmaceutical applications due to their high resolution and ability to perform full-scan analyses with high dynamic range, nevertheless, they present some limitations in terms of acquisition frequency, decreasing the range of application of these instruments [12].

MS detectors have evolved significantly, transitioning from classical single-stage systems to more advanced tandem configurations, allowing for greater analytical capabilities. Classical MS detectors focus on measuring the m/z of ions generated from the sample, providing reliable identification and quantification of analytes. However, their ability to resolve complex mixtures or elucidate molecular structures is often limited.

Tandem mass spectrometry (MS/MS) builds upon classical designs by incorporating multiple stages of mass selection and fragmentation, enabling the isolation of precursor ions, controlled fragmentation to generate product ions, and detailed analysis of these fragments. This approach enhances sensitivity and specificity, making MS/MS essential in fields requiring high precision and accuracy for the analysis of complex samples.

For instance, in metabolomics, the identification of metabolites in biological systems necessitates high resolution to differentiate between structurally similar compounds. Similarly, in environmental science, MS/MS is critical for identifying pollutants and their degradation products in water, soil, and air, where resolving closely related compounds is essential for accurate reporting. In food safety, tandem MS plays a vital role in detecting trace levels of contaminants, such as pesticides, mycotoxins, and veterinary drug residues, ensuring compliance with stringent safety regulations. The transition from classical to tandem MS represents a pivotal innovation, expanding the scope of analytical applications and addressing the increasing complexity of modern analytical challenges [10][13].

One of the first low-resolution tandem mass analyzers was the ion trap, whose operating principle has been previously described; today, it has largely been replaced by the triple quadrupole (QqQ). QqQ operates by using three quadrupoles in series, which enables them to perform multiple stages of mass analysis. The first quadrupole selects precursor ions based on their m/z , the second quadrupole which acts as a collision cell, fragments these ions, while the third quadrupole analyzes the resulting product ions. QqQ analyzers offer excellent sensitivity and selectivity, making them indispensable for applications where trace-level detection is required, such as in pesticide residue analysis in food, clinical diagnostics, and pharmaceutical research. The ability of QqQ instruments to focus on specific ion transitions through selected reaction monitoring (SRM) ensures that even low-abundance compounds can be detected with high accuracy. This high selectivity is particularly useful in cases where the sample matrix is complex, and matrix interference could otherwise hinder the analysis [14].

In high-resolution systems, tandem mass spectrometry is typically implemented using hybrid configurations. One of the most significant examples is the quadrupole-time-of-flight (QTOF) analyzers. QTOF analyzers are widely recognized for their capability to combine high mass accuracy with rapid data acquisition, making them particularly suitable for high-resolution analyses in complex sample matrices. QTOF instruments offer a dual advantage: they not only filter ions based on their m/z using a quadrupole but also provide high-resolution mass analysis using time-of-flight detection. This combination makes QTOF ideal for applications such as metabolomics, where non-targeted analysis is often required to detect unknown or unexpected compounds in a sample. One of the primary benefits of QTOF in high-resolution mass spectrometry is its ability to perform full-scan acquisition, which captures detailed mass spectra for each ion detected. This data-rich approach provides a comprehensive view of all the analytes present in a sample, enabling researchers to identify and characterize compounds with high precision. Additionally, QTOF analyzers are known for their high sensitivity and wide dynamic range, making them suitable for detecting trace levels of compounds, which is critical in environmental monitoring and forensic analysis. The QTOF's high resolving power allows for the separation of isobaric compounds, which is particularly important in the analysis of complex biological and environmental samples. For example, in environmental analysis, where pollutants and their degradation products may share similar masses, QTOF can differentiate between these closely related compounds, providing more accurate results [15]. However, these detectors, by generating such highly accurate information, produce big-size data, making the data processing less smooth and more time-consuming.

In conclusion, mass spectrometry analyzers differ in resolution, sensitivity, and functionality. Low-resolution systems are well-suited for routine analyses and quantitative measurements, with triple quadrupoles being the most commonly used in tandem configurations. High-resolution systems, including hybrid instruments like QTOF, offer greater accuracy and precision, making them essential for more complex analytical tasks. These advancements highlight the evolving capabilities of mass spectrometry, with each system designed to meet specific analytical needs.

2.3 Applications of GC-MS

GC-MS has become an indispensable tool in various scientific fields due to its ability to both separate and identify organic compounds within complex mixtures. Its versatility, high sensitivity, and the decreasing cost of equipment have significantly expanded its application across many sectors, including environmental monitoring, food safety, medical diagnostics, forensic science, and anti-doping testing. Despite its increasing accessibility, the effective use of GC-MS still requires qualified operators who are trained not only in operating the equipment but also in interpreting the complex data it generates.

- In environmental monitoring, GC-MS plays a critical role in detecting a wide range of organic pollutants. The method is particularly effective in screening for contaminants such as chlorophenols in water and soil, polycyclic aromatic hydrocarbons (PAHs), unleaded gasoline, dioxins, and dibenzofurans, among others. Additionally, the detection of trace-level contaminants is often required. In this case, the use of

advanced sample preparation techniques to effectively concentrate analytes before GC-MS analysis is mandatory. These extraction methods are essential for isolating volatile and semi-volatile organic compounds from complex matrices such as water, soil, and air. For example, the purge and trap technique is commonly used to concentrate volatile organic compounds (VOCs) like benzene, toluene, ethylbenzene, and xylene (BTEX), as well as halogenated solvents such as trichloroethylene and dichloromethane. These techniques enhance the sensitivity of GC-MS, making it particularly effective for screening low-concentration pollutants alongside more persistent contaminants like chlorophenols, PAHs, and organochlorine pesticides. A focus on this application is provided in the experimental section of the thesis where regulated VOCs as contaminants in surface and potable waters are monitored by purge-and-trap extraction and pre-concentration followed by GC-MS.

GC-MS also helps identify organochlorine pesticides, herbicides, phenols, and sulfur compounds in the air, making it highly versatile for both air and water quality assessments. The technique's precision makes it a powerful tool for environmental research, for example including tracking degradation products in bio-mass research or analyzing carbamazepine metabolites in treated wastewater without the need for derivatization. This broad applicability is essential for assessing the environmental impact of pollutants, identifying their sources, and ensuring compliance with environmental regulations [16].

- In bioanalysis, GC-MS is a powerful tool employed to detect a broad range of substances in biological samples, such as blood, urine, saliva, and other body fluids. Its application in clinical diagnostics and biomedical research has expanded significantly due to its high sensitivity, specificity, and ability to detect and quantify trace levels of compounds within complex biological matrices.

One of the primary uses of GC-MS in bioanalysis is the detection of pharmaceuticals and drugs of abuse. The technique is particularly effective for identifying anesthetics, anticonvulsants, sedatives, and hypnotics, as well as a wide array of narcotics and residual solvents [17].

Furthermore, GC-MS is increasingly applied to the discovery of novel biomarkers in biological samples. This aspect is particularly promising for the early detection of diseases and metabolic disorders. For instance, in the context of inborn errors of metabolism (IEM), GC-MS is used to detect abnormal metabolites in urine that may indicate rare genetic conditions affecting metabolic pathways. By identifying unusual compounds or altered levels of known metabolites, clinicians can diagnose metabolic diseases more effectively, allowing for earlier intervention and personalized treatment strategies [18], [19].

Other applications in the same field involve isotopic labelling, typically with carbon-13 (^{13}C). This labelling allows researchers to track metabolic activity, with the isotope ratio mass spectrometry (IRMS) component of GC-MS systems providing precise measurements of ^{13}C - ^{12}C ratios. These capabilities make the technique fundamental in studying metabolic processes and identifying deviations indicative of health issues [20].

In addition to its role in metabolic disorder diagnosis, biomarker discovery for chronic diseases is another growing area of application for GC-MS. For example, volatile organic compounds (VOCs) in breath samples have been studied as potential non-invasive biomarkers for diseases such as lung cancer, diabetes, and liver disease. Researchers have used GC-MS to analyze the composition of exhaled breath, identifying specific compounds that could be associated with pathological changes in the body. These compounds, once validated as disease-specific markers, could open the way for the development of new diagnostic tools that rely on breath analysis instead of invasive procedures like biopsies or blood tests [21].

- GC-MS is the gold standard for anti-doping tests in, the realm of sports, ensuring athletes' compliance with regulations regarding performance-enhancing substances. It is highly effective in detecting anabolic steroids and other prohibited substances in urine samples, even at trace levels, providing accurate and reliable testing that is crucial for maintaining fair competition. The high sensitivity and specificity of GC-MS ensure that performance-enhancing drugs are identified precisely, enabling sports organizations to enforce strict anti-doping regulations effectively [22].
- In forensics, GC-MS is an indispensable tool for investigating several scenarios. For example, it is used to analyze trace substances from the human body, clothing, or environmental samples to link suspects to criminal activity. Another prominent example is the analysis of fire debris to detect accelerants in arson investigations. Forensic samples are often complex, and the tandem use of mass spectrometry (GC-MS/MS) provides a high degree of accuracy, making it possible to isolate and identify specific compounds between interfering substances. The robust data generated by GC-MS/MS is admissible in court and is frequently used to establish scientific evidence that supports legal proceedings [23].

2.3.1 Food Applications of GC-MS

Food is a highly complex matrix, containing a wide range of compounds that can vary significantly in structure and concentration. To effectively investigate and analyze the components within food, advanced analytical tools are required. GC-MS has emerged as one of the most powerful techniques for this purpose, thanks to its exceptional sensitivity, specificity, and capacity to separate and identify complex mixtures. It plays a crucial role in the detection of volatile compounds, contaminants, pesticide residues, and fatty acids, offering detailed insights into food composition, quality, and safety. The following sections will explore the various applications of GC-MS in food analysis [24].

Component profiling is a crucial area where GC-MS demonstrates exceptional utility, particularly in the fields of metabolomics. In contemporary metabolomics studies, instruments such as GC-Q-MS, GC-QqQ-MS, and GC-TOF-MS are frequently employed [25]. For metabolite profiling, GC-TOF-MS is particularly advantageous due to its full-scan capability, enabling the analysis of a broad m/z range without constraints on scan time. On the other hand, Q-MS and QqQ-MS are more suited to targeted analyses, utilizing SIM or selected reaction monitoring (SRM) for quantification. This method proves highly effective, especially when paired with soft ionization, as it offers an extended dynamic range and enhanced specificity [26], [27]. The accurate quantification of these compounds is not only crucial for health claims but also for tracking metabolic pathways within food products [28].

A peculiar application is Lipidomics, which focuses on the comprehensive analysis of lipid profiles. In the past, this technique was not widely employed due to the complex derivatization required before separation [29]. This process can often remove critical structural information, particularly in polar lipids, making the analysis more challenging. Additionally, GC requires pre-separation steps that are both time-consuming and labor-intensive, involving hydrolysis and derivatization procedures. Despite these limitations, GC has become a standard tool for the quantitative analysis of complex lipid mixtures, especially for fatty acids. An early milestone in GC's development for bioanalytical applications was the separation, identification, and quantification of fatty acids, first described by James and Martin in 1952 [30]. Shortly after, Cropper and Heywood (1953) demonstrated the analytical separation of fatty acid methyl esters (FAME) using vapor-phase chromatography [31]. Following these pioneering studies, the characterization of fatty acid composition through esterification to FAME, followed by GC analysis, has become a routine and widespread method in lipidomics laboratories. Generally, the quantification of FAME is commonly performed using GC coupled with FID, which has long been established as a reliable quantitative approach, while GC-MS is primarily used for the qualitative analysis of FAME because it *“represents a potentially powerful tool for quantitative analysis of FAME, especially in the presence of a convoluted biochemical background”* [32].

Furthermore, GC-MS is widely used for metabolomic analysis of carbohydrates, amino acids, and secondary or specialized metabolites, providing a comprehensive view of the biochemical composition of food products [33]. This knowledge is crucial for verifying food authenticity and optimizing processing techniques. In addition, studies using metabolomics have demonstrated its potential in predicting quality attributes and informing strategic decisions in industry, enabling comprehensive quality assessment and optimization of storage and processing conditions for the food industry [34].

GC-MS is also a cornerstone technique in the domain of food quality; in fact, the analysis of various compounds in food products is essential for ensuring quality, primarily by meeting consumer expectations regarding sensory characteristics, health benefits, and accurate labeling of ingredients and origin. Consumers highly value the absence of contaminants and residues, along with the presence of desired components. Among the numerous analytical techniques used in food analysis, gas chromatography stands out due to its exceptional chromatographic resolution, enabling the simultaneous detection of several similar compounds. Moreover, advancements in extraction methods, combined with MS detection technologies, have significantly enhanced sensitivity, allowing for the identification of compounds even at trace levels.

In the domain of food safety and quality control, GC-MS can monitor a wide range of volatile and semi-volatile compounds in complex food matrices. Its precision and sensitivity make it ideal for detecting contaminants at trace levels, which is crucial in ensuring that food products meet both regulatory standards and consumer safety expectations. One of the primary applications of GC-MS in food safety is the detection of chemical contaminants and residues, including pesticides, mycotoxins, veterinary drugs, pollutants, and other harmful compounds that can compromise human health. These substances may be present due to

environmental exposure, agricultural practices, or unintentional contamination during food production [35][36]. Moreover, GC-MS is indispensable in ensuring compliance with international food safety standards, such as those set by the European Commission and the United States Food and Drug Administration (FDA). The development of advanced techniques such as tandem mass spectrometry (GC-MS/MS) further enhances the capabilities of GC-MS by improving selectivity and sensitivity, making it possible to detect even lower concentrations of contaminants in complex food matrices. The integration of automation and sophisticated data analysis software has also streamlined the process, allowing for high-throughput screening of food samples in both targeted and non-targeted workflows. Its ability to detect contaminants down to parts-per-billion (ppb) levels makes it one of the most reliable tools for verifying that food products are free from harmful substances [13][37][38].

Beyond contaminants, GC-MS is widely used in monitoring food additives, preservatives, and flavor compounds. Ensuring the accurate quantification of these substances is essential not only for compliance with food safety regulations but also for maintaining the sensory attributes of food that are important for consumer acceptance.

GC-MS is widely used to profile volatile compounds from different foods. Combined with headspace (HS) approaches like solid-phase microextraction (SPME) or dynamic headspace (D-HS), it enables high-throughput profiling and monitoring of many aroma-active compounds [39], [40]. For this reason, another crucial application in food analysis is the detection of adulteration and authenticity verification. This is particularly important in high-value products such as olive oil, honey, wine, and spices, where the intentional substitution or dilution of ingredients can have significant economic and health implications. Several papers underline the capacity of GC-MS to identify marker compounds that are indicative of specific geographic origins or production methods, thus verifying the authenticity of the product and protecting consumers from fraudulent practices [41], [42], [43].

GC-MS also can play an important role in another shape of quality control, *i.e.*, sensory properties. From the consumer perspective, this is the most important topic after food safety. The shape, color, size, freshness, consistency, taste, and smell (flavor) of the food are the main criteria for buying and consuming the products.

In particular, among the possible applications, three are key: the ability to detect off-flavors deriving from oxidative pathways and/or spoilage mechanisms, identification of flavour contaminations from the industrial processes, and, ultimately, the use of aroma compounds to guide industrial strategies (*e.g.*, guide the breeding of key raw materials).

Oxidative degradation is one of the leading causes of off-flavors in food products, particularly in foods rich in polyunsaturated fatty acids (PUFAs). When oils undergo oxidation, a variety of volatile compounds are produced, many of which contribute undesirable aromas and flavors that negatively impact product quality. According to the study by *Gómez-Cortés et al.*, lipid autooxidation leads to the formation of volatile compounds such as aldehydes, ketones, and alcohols, which are primary contributors to off-flavors in oxidized oils. Such compounds are produced in higher concentrations in oils with elevated levels of ω -3 fatty acids, more susceptible to oxidative processes, such as flaxseed and marine oils [44]. The use of GC-TOF-MS has proven effective in quantifying these oxidation products at low levels. This approach enables the early detection of off-flavors, allowing for the

implementation of quality control measures to mitigate the impact of oxidation during storage and processing.

In the context of food production, contamination from industrial processes can result in significant flavor alterations, often described as off-flavors or taints. These undesirable flavors can arise from various stages of manufacturing, including the use of materials and chemicals that come into contact with the food during packaging, storage, or transportation. For instance, during packaging, substances from the materials used, like inks, adhesives, or plastics, can migrate into the food, altering its flavor. Foods with high-fat content are particularly vulnerable, as fats can absorb these contaminants more readily. In addition, bad flavours can result from the water used for processing if it contains impurities such as phenols which, if chlorinated, can create unpleasant flavours.

For example, in the food and beverage industry, cork is a common material used for sealing wine bottles, but it can also be a source of off-flavors due to contamination. One of the primary contaminants associated with cork taint is chloroanisoles, particularly 2,4,6-trichloroanisole, which has a very low sensory threshold and can impart a musty or moldy odor [45]. This contamination occurs due to the microbial degradation of chlorophenols, compounds used in insecticides, herbicides, or chlorinated solutions that are applied during the bleaching process of cork. As cork is exposed to these treatments, microbial activity converts the chlorophenols into chloroanisoles, leading to taint even at trace levels.

The presence of such contaminants in cork underscores the importance of stringent quality control processes during cork production and storage. To address this issue, methods such as gas GC-MS have been largely employed in literature to detect chloroanisoles at very low concentrations, ensuring that tainted corks are identified before they reach the consumer [46], [47].

In industrial strategies, GC-MS allows for the high-throughput analysis of volatile compounds, which are key indicators of flavor. By identifying and quantifying these compounds, breeders can make informed decisions about which varieties to cultivate, enhancing the sensory qualities of the final product, such as sweetness, freshness, and aromatic intensity. This approach is not only important for maintaining desirable flavor profiles but also for balancing other key traits like durability, disease resistance, and shelf-life.

The ability of GC-MS to accurately detect and measure aroma compounds at trace levels makes it an auxiliary tool for breeding programs. It supports the selection of varieties that meet consumer demands while optimizing the economic value of crops. As such, the integration of volatile profiling through GC-MS into breeding strategies highlights the role of analytical chemistry in shaping the future of food production, where both sensory and agronomic qualities are finely tuned to meet market needs [48].

In conclusion, GC-MS has proven to be an indispensable tool in various scenarios. Further applications within the realm of foodomics will be discussed in greater detail in dedicated Chapters of the thesis.

2.4 Use of alternative carrier gasses in GC-MS

Helium has long been the preferred carrier gas in GC-MS due to its unique combination of advantageous chromatographic properties and chemical inertness. As an inert gas, helium prevents unwanted reactions during the ionization process, ensuring reliable and accurate fragmentation of analytes. This makes it ideal for applications where precision and sensitivity are of fundamental importance. In addition, its high diffusivity and low viscosity improve separation performance and contribute to faster analyses.

However, despite its widespread use, helium is a finite resource. It is primarily sourced from natural gas reserves and through the radioactive decay of certain minerals. However, it is a relatively rare element, making up only about 5.2 parts per million by volume in Earth's atmosphere. Given its scarcity, concerns about a future helium shortage have been raised, and the effects are already being felt in the form of escalating prices and supply uncertainties. These factors have significantly impacted its accessibility, especially for laboratories relying on its consistent availability [49].

The growing helium crisis has highlighted the need to identify and evaluate alternative carrier gases that match the performance of helium in GC-MS while offering better long-term availability. Researchers and laboratories worldwide are now focusing on other gases as potential replacements. However, each alternative brings its own challenges and trade-offs, searching for a suitable replacement both complex and essential for the future of GC-MS. For these reasons, several other gases have been explored in the literature as alternatives. Among these, hydrogen, nitrogen, argon, and, in some cases, compressed air has been employed. Argon and compressed air are less frequently applied, usually in more specific or niche applications [50][51].

The selection of a suitable carrier gas depends on the balance between performance, safety, and cost-efficiency. Hydrogen and nitrogen are undoubtedly the most commonly used alternatives, with hydrogen offering faster analysis times and nitrogen being a cost-effective option.

The evaluation between different gasses can be made by considering the Van Deemter equation (**Equation 2.1**), a key model that describes how various factors influence column efficiency.

Equation 2.1:

$$H = A + \frac{B}{\bar{u}} + C\bar{u}$$

Where H: height of a theoretical plate

$\bar{u}$: average linear velocity of the mobile phase

A: eddy diffusion term

B: longitudinal diffusion term

C: mass transfer term

The equation identifies three primary terms, Eddy diffusion (A), longitudinal diffusion (B), and mass transfer resistance (C), that together determine the overall plate height (H), a

measure of the separation efficiency. The optimal carrier gas is the one that minimizes the plate height by balancing these factors across a range of flow rates.

- **Eddy Diffusion (A term):** This refers to the multiple pathways that molecules can take as they travel through the packed column, leading to band broadening. The choice of carrier gas has little impact on this term, as it depends mainly on the uniformity of the packing material.
- **Longitudinal Diffusion (B term):** This describes the diffusion of analytes along the length of the column. The higher the diffusion coefficient of the carrier gas, the faster analyte molecules can spread longitudinally, reducing efficiency. Gases with lower diffusivity, such as nitrogen, exhibit a slower rate of longitudinal diffusion, which is advantageous at low flow rates but becomes problematic at higher velocities.
- **Mass Transfer Resistance (C term):** This refers to the resistance to mass transfer between the mobile phase (carrier gas) and the stationary phase (column coating). Faster-moving gases like hydrogen reduce the time analytes spend in equilibrium between these phases, enhancing efficiency at higher flow rates.

In the case of a capillary column, the A term in the van Deemter equation is equal to zero because there is no packing material to create multiple flow paths for the solute molecules. As a result, the absence of Eddy diffusion contributes to the high efficiency and superior resolution characteristic of capillary columns. Thus, **Equation 2.1** simplifies to the Golay equation (**Equation 2.2**) [52].

Equation 2.2:

$$H = \frac{B}{\bar{u}} + C\bar{u}$$

While the Golay equation is specifically formulated to describe the behavior of capillary columns, the van Deemter equation provides a more general model applicable to both packed and capillary columns. The following van Deemter curves illustrate the contribution of different terms to the height equivalent to a theoretical plate (HETP) as a function of the linear velocity of the mobile phase, highlighting the optimal conditions for maximizing chromatographic efficiency.

The optimal velocity is the point where the curve reaches its minimum, representing the best compromise between the three factors [53]. **Figure 2.1** shows Van Deemter curves for Helium, Nitrogen, and Hydrogen with a 30m × 0.25mm ID × 0.25µm.

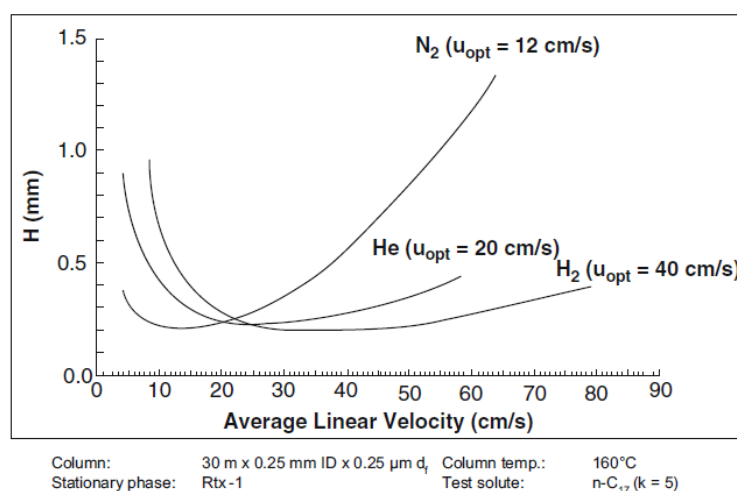


Figure 2.1: $H/\bar{u}$ plots for different carrier gases: nitrogen, helium, and hydrogen [27]

Figure 2.1 shows that helium offers a wide range of optimal linear velocities that allow flexible adjustment of the flow rate without major losses in efficiency. This is because its molecular diffusivity is a good balance; low enough to minimize longitudinal diffusion, but not so low that mass transfer becomes a significant problem. This gives helium a relatively flat Van Deemter curve, which makes it suitable for a wide range of analytical conditions. However, the optimum velocity of helium is lower than that of hydrogen, which means longer analysis times if you are aiming for maximum efficiency.

On the other hand, hydrogen has the steepest Van Deemter curve, but also the highest optimal velocity, meaning it can maintain high efficiency at much faster flow rates compared to helium and nitrogen. This leads to significantly shorter run times and a higher throughput. Hydrogen's high diffusivity reduces longitudinal diffusion, while its fast mass transfer minimizes band widening even at high speeds. Despite these advantages, the flammability of hydrogen and its potential reactivity with analytes or column materials limit its use in certain applications.

Nitrogen exhibits a very different behavior on the Van Deemter curve. It has the lowest diffusivity of the three gases, resulting in minimal longitudinal diffusion at low flow rates and making it very efficient at these velocities. However, at higher flow rates, the mass transfer term of nitrogen becomes dominant, causing the plate height to increase rapidly and efficiency to decrease. As a result, nitrogen has a much narrower range of optimal flow rates, and leaving this range leads to significant losses in resolution. This makes nitrogen unsuitable for high-speed separations, but a viable, cost-effective option for slower, routine analyses where resolution and analysis time are less critical [27].

To summarize, both hydrogen and nitrogen are excellent alternatives to helium as a carrier gas in gas chromatography. Each gas offers distinct advantages: hydrogen allows for shorter analysis times and higher efficiency at higher speeds, while nitrogen is a cost-effective and readily available option, especially for routine analysis, thanks in part to the ability to generate it on-site with suitable nitrogen gas generators, providing laboratories with flexibility and a continuous supply at a lower cost [54]. However, one of the biggest challenges with hydrogen is its flammability, which raises significant safety concerns in the laboratory

environment. Hydrogen has a wide flammability range from 4% to 74.2% in air, combined with a high burning rate and very low ignition energy. These factors increase the risk of fire or explosion when hydrogen accumulates in enclosed spaces such as the GC oven, the MS detector, or the laboratory itself.

Modern GC systems are often equipped with safety features, such as automatic shutdown of the hydrogen flow in case of leaks, hydrogen sensors in the GC oven, and explosion protection mechanisms, which help to mitigate these risks. Nevertheless, one of the most effective ways to reduce the hazards associated with hydrogen is to use hydrogen gas generators instead of conventional pressurized gas cylinders. These generators not only provide a safer, on-demand supply of hydrogen but also eliminate the risk of gas build-up and spontaneous combustion due to leaking cylinders. Thus, hydrogen generators provide a practical solution to the safety issues associated with the use of hydrogen in GC, making it a more viable alternative to helium for many laboratories [55].

2.4.1 Hydrogen in GC-MS

As mentioned in the previous section, hydrogen is a highly effective carrier gas with several key advantages. It increases the speed of analysis and improves resolution in gas chromatography. Its many advantages include faster separations, the ability to work at lower temperatures, longer column life, reduced environmental impact, and greater availability. These factors make hydrogen a compelling alternative to helium, especially in laboratories seeking a balance between performance and sustainability. However, despite these advantages, there are also some disadvantages associated with its use.

The first disadvantage of using hydrogen as a carrier gas becomes apparent in connection with the vacuum system required to operate the mass spectrometer. Due to the higher diffusivity of hydrogen compared to helium, both the rotary and turbomolecular pumps -the most widespread in MS systems- must operate at a higher power to maintain a stable vacuum. This increased diffusivity can lead to a less effective vacuum when using hydrogen, which can have a negative impact on the performance of the mass spectrometer [56].

In addition, an optimal operating vacuum can avoid the chemical noise sometimes seen when oil diffusion pumps are not working properly. The differences in vacuum performance between hydrogen and helium need to be carefully considered as they directly affect the quality and stability of MS measurements, highlighting the operational challenges associated with hydrogen in this context.

Another potential challenge in using hydrogen as a carrier gas in GC-MS arises from its lack of inertness. Unlike helium, hydrogen can undergo reactions during the ionization and fragmentation process that can lead to unreliable mass spectra. This reactivity can lead to deviations in the fragmentation patterns of the analytes. The result is spectra that are not fully comparable to those stored in commercial databases, which are predominantly recorded with helium. As a result, the use of hydrogen can complicate the interpretation of mass spectrometry data and reduce the accuracy of compound identification.

Information regarding the challenges of using hydrogen as a carrier gas in gas chromatography is not widely documented in the literature, and the studies available often present conflicting findings. According to Muñoz-Guerra *et al.* [56] a notable issue observed when using hydrogen as the carrier gas is the variation in the fragmentation pattern of steroids. The intensity of the signal generally decreases, with smaller, less specific fragments becoming more abundant. This shift in fragmentation behavior leads to a reduction in the signal-to-noise (S/N) ratio for most diagnostic ions. As a result, hydrogen introduces an increase in background noise and alters the relative abundance of the key diagnostic ions. This loss in the abundance of the more specific ions negatively impacts the detection capability of the system, as the contribution of noise becomes more pronounced in the selected signals. Contrary Lin *et al.* indicated that the fragmentation patterns of 170 adulterants in traditional Chinese medicines are similar using hydrogen as the carrier gas except for nitrobenzene. This effect derives probably from the reduction of nitrobenzene in the presence of hydrogen, thus resulting in the different fragmentation patterns and peak tailings. Nevertheless, they observed an increment of the fragmentation and a higher abundance of the diagnostic ions [57].

In another study conducted with an ion trap MS detector Impens and co-workers reported that *'the fragmentation patterns are similar whether helium or hydrogen is used as carrier gas'* for the MS calibration gas (perfluorotributylamine). In the same paper, examining the MS spectrum and selecting specific fragments ion of norgestrel they reported a higher intensity of the relative fragment ion when using hydrogen as carrier gas. They underline also that *'the combination of hydrogen (as carrier gas) and helium (as collision gas) in the ion trap results in a higher degree of fragmentation, and the ion trap spectra resemble more those obtained using a quadrupole mass analyzer'*. Impens *et al.* also evaluated full-scan MS-MS using hydrogen as the carrier gas and found that it produced a "cleaner" fragment ion spectrum, with slightly more component-specific fragmentation [58].

As already mentioned, another problem when using hydrogen is the increase in background noise, which leads to a reduction in the S/N ratio. The S/N values are typically 3 to 5 times lower with hydrogen compared to helium. This reduction means a decrease in sensitivity that may be caused by chemical interactions between hydrogen and certain analytes within the MS ion source, or by other contributing factors [57].

A comparison between helium and hydrogen was also conducted by Oshima *et al.*, specifically focusing on the S/N. They observed the highest S/N was for helium, followed by hydrogen, which was 55-93% lower than that with helium for the molecules they analyzed [59]. Nevertheless, Oshima and co-workers underlined that the limit of detection was 1/100th of guideline values and so they concluded that hydrogen is an optimal solution as alternative carrier gas in GC-MS.

In conclusion, the balance between the advantages and disadvantages of using hydrogen as a carrier gas in GC-MS must be carefully considered. While hydrogen shows promise for many applications, it cannot yet be regarded as a universal carrier gas without limitations. Both the scientific community and analytical instrument manufacturers will undoubtedly continue to explore ways to mitigate the challenges outlined above. However, at present, the

use of hydrogen requires a case-by-case evaluation, paying particular attention to the specific molecules being analyzed and the potential for cross-reactions that could affect the results.

2.5 Method translation

Method translation in GC is a theoretical framework that provides a systematic approach to optimizing chromatographic methods while preserving the integrity of the separation process. At its core, method translation relies on the concept of void time (t_M) as a universal time unit in chromatography. By expressing all time-related components in units of t_M , method translation achieves a high degree of predictability in chromatographic behavior. For instance, normalized temperature programs allow for proportional adjustments to heating rates and plateau durations, ensuring that methods remain mutually translatable when other factors, such as stationary phase type and phase ratio, are held constant. This normalization framework underlies the scalability of GC methods and enhances the predictability of changes in separation performance. This theoretical construct ensures that the proportionality of retention times for analytes is maintained across different methods, preserving the peak elution order regardless of changes in column geometry, carrier gases, or pneumatic conditions.

The fundamental principle of method translation is rooted in the proportional scaling of chromatographic parameters. This scaling ensures that the elution pattern of peaks remains consistent even when the analysis speed is increased or decreased. For instance, the retention times of solutes in two methods, t_{R1} and t_{R2} , are related by the speed gain factor S , expressed mathematically in the **Equation 2.3**.

Equation 2.3:

$$S = \frac{t_{R1}}{t_{R2}}$$

Where S : speed gain factor

t_{R1} : Retention time in method 1

t_{R2} : Retention time in method 2

This proportional relationship is central to translating methods between columns of varying dimensions or adapting them to different carrier gases.

A critical aspect of method translation theory involves classifying chromatographic changes into translatable and non-translatable parameters. Translatable changes include adjustments to column dimensions, carrier gas types, and temperature program rates, while non-translatable factors, such as the stationary phase type and initial temperatures, must remain constant during translation. This distinction is essential for maintaining the fidelity of chromatographic separations during method adaptation. For example, in temperature-programmed GC, scaling the temperature ramp rate proportionally to the t_M ensures that the elution order and resolution of peaks are preserved across methods [60].

The theory also emphasizes the impact of instrumental and physicochemical constraints on method translation. The relationship between column dimensions and carrier gas dynamics is particularly significant. For instance, smaller column diameters result in narrower peaks and faster analysis times but require higher pressures to maintain optimal flow rates. Theoretical models account for these trade-offs by optimizing parameters such as the carrier gas velocity and temperature ramp rates, ensuring that separations are both efficient and scalable. Additionally, the choice of carrier gas influences the separation dynamics, with hydrogen offering the fastest analysis speeds due to its superior diffusivity compared to helium or nitrogen. Theoretical calculations incorporate these factors to guide method development and ensure compatibility across different configurations.

The theoretical framework accommodates changes in column dimensions, such as moving from wide-bore to narrow-bore columns and vice versa. This adaptability ensures that separation quality is preserved even when transitioning to columns with vastly differences. These differences can be in terms of internal diameters or phase ratios, which is the ratio of the volume of the mobile phase (carrier gas) to the volume of the stationary phase within a capillary column. It is commonly expressed in terms of the column's geometry as reported in **Equation 2.4**.

Equation 2.4:

$$\beta = \frac{d_c}{4d_f}$$

Where d_c : the internal diameter of the column
 d_f : the thickness of the stationary phase film

The β influences key chromatographic properties, such as retention time, resolution, and separation efficiency. Columns with a high β offer shorter retention times and faster analyses but may compromise resolution for closely eluting compounds. Conversely, columns with a low phase ratio provide greater interaction between analytes and the stationary phase, enhancing resolution at the cost of longer analysis times [27][52].

As with the difference in column geometry, the choice of carrier gas like helium, hydrogen, or nitrogen is taken into account when translating the method by incorporating differences in diffusivity and viscosity of the gas into the scaling process. Hydrogen, for instance, provides higher diffusivity and lower viscosity, making it ideal for faster separations, while helium offers a balance between speed, safety and inertness. Nitrogen, though slower, is often preferred for its cost-effectiveness, especially in routine analyses. These gas-specific properties are critical in ensuring that the translated method maintains consistent performance and resolution across different setups.

Moreover, method translation addresses the distinct requirements of detectors operating under different pressures, such as vacuum conditions in MS versus atmospheric pressure in FID. By adjusting pneumatic conditions and temperature programs appropriately, method translation ensures that retention times and separation efficiency are maintained across these varied setups. This flexibility not only simplifies the adoption of diverse analytical

configurations but also enhances the reproducibility and robustness of chromatographic results

In summary, method translation in GC is supported by a robust theoretical framework that prioritizes the preservation of peak elution patterns through proportional scaling of chromatographic parameters. By leveraging the t_M as a universal normalization metric and classifying changes into translatable and non-translatable categories, method translation provides a systematic approach to adapting and optimizing GC methods [61].

In conclusion, these theoretical principles not only ensure the reproducibility and scalability of chromatographic separations but also play a central role in advancing green analytical chemistry. By enabling the reduction of analysis times, such as through the adoption of fast GC, they contribute to lower energy consumption and improved operational efficiency. Furthermore, method translation facilitates seamless transitions between detectors, such as from MS to FID or TCD or new alternative detectors, which often require less energy to operate. It also supports the replacement or reduction in the use of non-renewable resources like helium by incorporating more sustainable alternatives, such as hydrogen or nitrogen, without compromising analytical performance. This adaptability not only enhances the sustainability of chromatographic practices but also aligns with broader goals of minimizing the environmental impact of analytical methodologies.

2.6 *Fast GC*

The growing demand for higher analytical throughput in quality control and assurance, combined with an increasing emphasis on safety and environmental sustainability, has been a major driver in the evolution of fast chromatographic techniques. This shift is not only motivated by the need to enhance laboratory productivity by reducing analysis times and operational costs while maintaining high separation efficiency but also by the principles of green analytical chemistry. Fast GC exemplifies this synergy, offering the dual benefit of significantly shortening analysis durations and aligning with sustainable practices. By minimizing energy consumption, optimizing resource use, and enabling the adoption of more environmentally friendly methodologies, Fast GC addresses both the operational and ecological imperatives of modern analytical laboratories, ensuring reliable qualitative and quantitative results without compromising sustainability goals. Although the principles of high-speed GC were explored as early as the 1960s, it is only in recent decades that advancements in instrumentation and methodology have made it viable for routine use in laboratories [62].

The speed of GC analysis is typically assessed through parameters such as average peak width, analysis time, and the rate of temperature programming [63]. Currently, a GC analysis is considered "fast" if it is completed in under 10 minutes using columns with internal diameters between 0.25 and 0.1 mm, lengths ranging from 5 to 15 meters, temperature programming rates of 20–60°C/min, and peak widths of 0.5–5 seconds. In contrast, "ultra-fast" GC (UFM-GC) refers to analyses performed in under one minute, using shorter columns (2–10 m) with narrower diameters (0.1–0.05 mm), temperature programming rates exceeding

1°C/s, and peak widths as narrow as 0.05–0.5 seconds. Achieving temperature programming rates above 100°C/min typically requires the use of directly resistively heated columns [64].

Two primary strategies are used to accelerate GC analyses. The first, and more commonly adopted, involves the use of short columns with very narrow inner diameters. The second employs short capillary columns with conventional diameters. Both approaches rely on reducing column length to increase analysis speed while maintaining sufficient efficiency to achieve the required separations. For the first approach, efficiency losses due to shorter column lengths can be mitigated by using narrower diameters (e.g., 0.1 mm), higher carrier gas velocities, thinner stationary phase films, and optimized temperature programming rates of 20–60°C/min. This method can reduce analysis times by a factor of four to eight compared to conventional GC, making it suitable for complex matrices while maintaining compatibility with detectors such as QMS [61][63].

In contrast, the second approach capitalizes on the often-excessive efficiency of conventional capillary columns, recognizing that a calculated reduction in column length can significantly shorten analysis times without compromising separation quality. For moderate-complexity samples containing up to 30–40 components, analysis times can similarly be reduced by a factor of four to eight using 5-meter columns paired with appropriate temperature programming. While this approach reduces the system's theoretical plate count, sufficient separation can still be achieved by selecting stationary phases with enhanced selectivity tailored to the specific application [66].

Ultimately, advancements in fast and ultra-fast GC methodologies represent a major step forward in chromatographic science. By balancing speed and efficiency through careful optimization of stationary phase, column geometry, and operational conditions, these approaches align with the modern needs of high-throughput, cost-effective, and environmentally conscious analytical practice.

As previously discussed, method translation plays a central role in modern chromatographic advancements, particularly in the development and application of fast gas chromatography. This approach not only enhances the reproducibility and scalability of analytical methods but also supports the transition to greener, more sustainable practices. A significant milestone in the routine implementation of F-GC was achieved through the foundational work of Klee and Blumberg [61]. This innovation enables laboratories to seamlessly adapt existing methods to faster, more efficient workflows while maintaining analytical integrity. By simplifying the transition to F-GC, method translation has become an indispensable tool for addressing both the practical and environmental challenges of contemporary analytical chemistry.

2.7 Portable GC systems

In recent years, on-site analysis has garnered increasing attention due to its ability to provide qualitative and/or quantitative information directly at the location where the samples are collected, often far from analytical laboratories. This approach is particularly crucial in situations where the composition of the target analytes may change during transport or where analytical results are urgently needed to advise immediate actions, especially in the presence

of potential hazards or threats. Various analytical technologies have been employed for on-site analysis, spanning a broad range of applications, including the detection of pollutants, hazardous compounds, and other critical components in environmental, industrial, and biological samples.

The feasibility of on-site analysis largely depends on the development of both analytical instruments and sample preparation techniques. Instruments designed for field use differ substantially from those intended for laboratory settings. Portability is a key feature of on-site tools, which are often required to be compact, lightweight, and sometimes even handheld or smartphone-compatible. In some cases, mobile laboratories are used, where traditional laboratory instruments are housed in vehicles [67].

The necessity for portability and miniaturization imposes restrictions on which analytical techniques can be effectively employed in the field. Analytical workflows, including sampling, preparation, detection, and data processing, must be performed entirely at the collection site, often in challenging environmental conditions. This requirement renders many laboratory-based techniques unsuitable for on-site applications, especially those that demand large amounts of reagents, involve lengthy procedures or rely on complex auxiliary equipment. Traditional laboratory methods such as, for example, solvent extraction, rotary evaporation, and nitrogen blow-down concentration are impractical for field use. Instead, microextraction and sorbent-based techniques have emerged as viable alternatives for both liquid and gas samples and are now widely applied in on-site analysis, even though some sample preparation treatments may still be necessary.

In general, on-site instruments do not achieve the same performance as laboratory-based instruments. However, several key factors must be prioritized for successful on-site analysis. For instance, the sampling process should be simple, avoiding complex pretreatment procedures or the need for additional equipment. The instruments must be compact and lightweight, facilitating transport to the sampling location either manually or by vehicle, while also maintaining low energy consumption, as they typically operate on battery power or short-term supplies from vehicles.

Fast analyses are essential, as the entire analytical workflow must accommodate and meet the rapid response demands often required in field operations while maintaining high performance. These methods must deliver quick results without compromising on key analytical characteristics, such as sensitivity, selectivity, and detection limits, ensuring accurate detection and, when necessary, quantification of target components [8].

Additionally, instruments designed for on-site analysis must prioritize user-friendliness, ensuring that even individuals without extensive technical expertise can operate them effectively after minimal training. This usability aspect is critical for enabling fast and reliable analyses in field conditions, where access to specialized personnel may be limited. Furthermore, these instruments need to demonstrate a high degree of robustness, as they are often deployed in environments with challenging conditions. They must be capable of resisting temperature fluctuations, which can affect both the functionality of sensitive

components and the accuracy of results. In addition, they should be resistant to humidity, dust, and wind exposure, all of which can degrade performance over time. Lastly, the physical durability of the instruments is fundamental. They must be designed to endure the mechanical stresses that come with frequent transportation, including exposure to vibrations, impacts, and shocks, without compromising their operational integrity or calibration.

Nowadays, the market for commercial portable GCs has expanded significantly, with numerous options available for on-site analysis. A wide range of analytical instrument manufacturers now offer portable GC solutions, reflecting the growing need for flexible, field-deployable systems [68][69].

The miniaturization of GC systems can be achieved through two main approaches:

Reducing the size of individual components, such as the injector, column, and detector, without fundamentally altering the system's overall architecture or adopting microfluidic technologies. The first allows for a more compact design while retaining the core functionalities of traditional gas chromatographs while thanks to the recent advancements in microfluidic technologies, specifically through the use of microelectromechanical systems (MEMS), new concepts could be explored. By integrating MEMS technology, it is possible to fabricate highly compact and efficient microfluidic devices, which enable the precise control of gas flow and sample handling at a microscale level.

These innovations not only reduce the physical size of the chromatographic system but also improve its portability, making it more suitable for on-site analysis and field applications [70].

Concerning the first approach, i.e. size reduction of existing modules, creating a portable GC involves carefully designing and miniaturizing several key components, starting with the injector. In standard GCs, injectors are large, robust systems that provide precise sample introduction under controlled laboratory conditions. However, in a portable GC, the injector must be much smaller to reduce the overall size of the instrument while still delivering accurate and repeatable sample injections. Achieving this involves redesigning the injection system to minimize its footprint, often using simplified components that can handle smaller sample volumes but still maintain suitable precision. Additionally, the portable injector must be robust enough to handle field conditions, such as temperature fluctuations and potential mechanical stresses, while ensuring consistent sample introduction for reliable analysis [71].

The next critical component in a portable GC is the column, which is responsible for the separation of the sample's components. Traditional laboratory GCs use long capillary columns, often between 30 and 60 meters in length, to achieve high-resolution separations over extended analysis times. However, for a portable system, size constraints necessitate the use of much shorter columns, typically ranging from 1 to 5 meters. These shorter columns allow for faster analysis times, which is crucial for on-site applications where quick results are often needed. However, the reduction in column length can limit the separation efficiency, as shorter columns provide less interaction time between the analytes and the stationary phase. To compensate for this, column materials and coatings are optimized for specific applications, ensuring that, despite the shorter column length, adequate separation is achieved. The

columns in portable GCs are designed to be more durable, and capable of withstanding vibrations, temperature fluctuations, and other environmental factors that could affect their performance in the field. Another aspect of column durability relates to replacement; unlike traditional GCs, where columns can be easily swapped or modified for different analyses, portable systems typically do not offer the same flexibility. As a result, the columns must be more resilient and versatile, ensuring reliable performance over extended periods without the need for frequent replacements or adjustments [69], [72].

The final major consideration in designing a portable GC is the selection and miniaturization of the detectors. In laboratory settings, GCs often utilize highly sensitive and specialized detectors, which provide detailed qualitative and quantitative information. However, these detectors are typically large and require complex vacuum systems, making them unsuitable for portable applications. Instead, portable GCs are commonly paired with detectors that are both compact and capable of operating without the need for extensive infrastructure. TCDs and photoionization detectors (PIDs) are the most frequently used in portable GCs due to their simplicity, small size, and ability to provide rapid detection of a wide range of analytes. TCDs, for example, work by detecting changes in the thermal conductivity of the carrier gas as different analytes pass through the detector, making them ideal for general-purpose applications. PIDs, on the other hand, are highly sensitive to volatile organic compounds and are often used in environmental and industrial settings where detecting trace levels of pollutants is crucial. Both detectors are designed to be lightweight, easy to transport, and capable of operating under varying environmental conditions without the need for external power sources or complex maintenance procedures [73][74].

A detector of fundamental importance in many applications turns out to be the MS detector. The miniaturization of mass spectrometers, particularly for field-portable applications, presents significant challenges due to the complex requirements needed for effective analysis. Mass spectrometry relies on several critical components - such as the ionization source, mass analyzer, and vacuum system - that must be scaled down without compromising performance. One of the key obstacles is maintaining the required vacuum levels, which typically necessitate bulky, power-hungry pumps. Despite the presence of all these criticalities and challenges, according to Learny *et al.* [75] various stratagems were used for the miniaturization of these detectors.

To address this, many portable systems now use technologies like toroidal ion traps or quadrupole analyzers, which can operate at higher pressures, thus reducing the need for sophisticated vacuum systems. For instance, toroidal ion traps provide increased ion storage volume compared to traditional ion traps, minimizing ion-ion repulsions and improving performance under high-pressure conditions.

However, the miniaturization process often results in trade-offs. Reduced size can lead to limitations in mass resolution and sensitivity, as seen in some portable systems that may struggle with complex matrices or low-concentration analytes. Each component, from the injector to the columns and detectors, must be carefully designed to ensure that the instrument remains compact, lightweight, and durable enough for field use, all while

delivering accurate and reliable analytical results. These innovations have made portable GCs indispensable tools in a wide range of applications, including environmental monitoring, industrial safety, and emergency response, where rapid, on-site analysis is essential [74][75].

The approach that exploits the use of MEMS and micro-machining to realize portable GC bring some advantages. These technologies allow for the creation of highly compact and lightweight devices that are easy to transport and operate in challenging environments. Despite their smaller size, these instruments can perform rapid analyses by utilizing short columns and miniaturized detectors. However, this focus on speed often comes at the expense of separation power and resolution.

The first on-chip GC was developed in the late 1970s by Terry *et al.* [77], marking a significant advancement in miniaturization. This pioneering device consisted of an injection valve connected to a 1.5 m planar column and a thermal conductivity detector, all integrated into two coupled wafers. Terry and co-workers were the first to introduce planar columns, also known as microchannels, MEMS columns, or microfabricated columns, utilizing microelectromechanical systems (MEMS) technology. These columns are fabricated using a gas-phase reactive ion etching technique, which creates rectangular open channels on a silicon wafer. The wafer is then sealed by bonding it to a plate made of either silicon or Pyrex glass. A major innovation of these planar columns is their capability to resist rapid temperature changes, facilitated by thin-film resistance heaters integrated into the etched silicon channels. This design effectively removes the need for traditional convection ovens, streamlining the overall structure of the miniaturized GC system. Resistance heating also offers benefits such as reduced energy consumption and faster heating rates, which are particularly advantageous for increasing separation speed with shorter columns [78].

On-chip GC systems offer clear benefits in terms of energy savings, reduced material use, and space efficiency, making them ideal for compact analytical platforms. On-chip technology plays a crucial role in developing portable GC systems for in-field analyses, with successful applications in environmental monitoring, the mineral oil industry, and forensic science for the detection of toxicants and explosives [79]. When integrated with on-line sample preparation or micro-concentration systems, on-chip GCs can be highly effective in monitoring the evolution of volatile mixtures or specific markers in real-time in different applications. Notable uses include tracking plant maturation processes, studying ecological biochemistry and multitrophic interactions, conducting chemotaxonomic research, and controlling the production of essential oils or extracts [80][81].

Despite the advancements in miniaturized GC systems, their performance has not yet fully reached the level of traditional laboratory GCs. However, numerous studies in the literature report excellent results achieved by these portable instruments, demonstrating their potential for high-quality analysis [81][82]. As development continues, these systems are expected to increasingly rival their laboratory counterparts, offering reliable and efficient performance across various applications.

Although the two approaches, MEMS and non-MEMS, have been discussed separately, they are often combined in the commercially available portable GC systems today. The integration of both MEMS and non-MEMS components in a single system highlights a key

challenge: while MEMS technology enables significant miniaturization, non-MEMS parts, such as conventional columns and detectors, often limit further size reduction. Moreover, the stability of MEMS components has yet to fully match the reliability of their non-MEMS counterparts. At the same time, ongoing research is driving the development of next-generation portable GCs that aim to enhance performance while reducing power consumption and improving separation efficiency and detection sensitivity. These advancements are expected to bring portable GCs closer to the performance levels of traditional laboratory-based systems, pushing the boundaries of what is achievable in on-site analytical applications. By addressing the limitations posed by both MEMS and non-MEMS technologies, future devices will offer even greater potential for high-quality, portable gas chromatography in diverse fields [67].

References

- [1] R. Upadhyay, K. Patel, and U. Upadhyay, "A Review Article on Advancements in GC-MS," *Int. J. Pharm. Res. Appl.*, vol. 8, no. 2, p. 54, 2023.
- [2] B. Gruber, F. David, and P. Sandra, "Capillary gas chromatography-mass spectrometry: Current trends and perspectives," *TrAC - Trends Anal. Chem.*, vol. 124, p. 115475, 2020.
- [3] R. S. Gohlke, "Time-of-Flight Mass Spectrometry and Gas-Liquid Partition Chromatography," *Anal. Chem.*, vol. 31, no. 4, pp. 535-541, 1959.
- [4] S. Squara *et al.*, "Validation of a high-throughput method for the accurate quantification of secondary products of lipid oxidation in high-quality hazelnuts (*Corylus avellana* L.): A robust tool for quality assessment," *J. Food Compos. Anal.*, vol. 114, p. 104766, Dec. 2022.
- [5] O. D. Sparkman and F. G. K. Zelda E. Penton, *Gas chromatography and mass spectrometry : a practical guide*, vol. 27, no. 7. 2011.
- [6] A. El-Aneed, A. Cohen, and J. Banoub, "Mass Spectrometry, Review of the Basics: Electrospray, MALDI, and Commonly Used Mass Analyzers," vol. 5, no. 3, pp. 185-192, 2012.
- [7] D. Nolting, R. Malek, and A. Makarov, "Ion traps in modern mass spectrometry," *Mass Spectrom. Rev.*, vol. 38, no. 2, pp. 150-168, 2019.
- [8] G. Emanuela, Ed., *Green Approaches for Chemical Analysis*, vol. 19, no. 5. 2016.
- [9] X. Wang, S. Wang, and Z. Cai, "The latest developments and applications of mass spectrometry in food-safety and quality analysis," *TrAC - Trends Anal. Chem.*, vol. 52, pp. 170-185, 2013.
- [10] F. W. McLafferty, "High-resolution mass spectrometry," *Anal. Chem.*, vol. 151, no. 3711, pp. 641-649, 1966.
- [11] A. M. Haag, "Mass analyzers and mass spectrometers," *Mod. Proteomics – Sample Prep. Anal. Pract. Appl.*, vol. 919, pp. 157-169, 2016.
- [12] R. A. Zubarev and A. Makarov, "Orbitrap mass spectrometry," *Anal. Chem.*, vol. 85, no. 11, pp. 5288-5296, 2013.
- [13] M. L. Xu, Y. Gao, X. Wang, X. X. Han, and B. Zhao, "Comprehensive strategy for sample preparation for the analysis of food contaminants and residues by gc-ms/ms: A review of recent research trends," *Foods*, vol. 10, no. 10, 2021.
- [14] I. Ferrer and E. M. Thurman, *Advanced Techniques in Gas Chromatography-Mass Spectrometry (GC-MS-MS and GC-TOF-MS)*, vol. 15, no. 3. 2013.
- [15] F. Hernández, T. Portolés, E. Pitarch, and F. J. López, "Gas chromatography coupled to high-resolution time-of-flight mass spectrometry to analyze trace-level organic compounds in the environment, food safety and toxicology," *TrAC - Trends Anal. Chem.*, vol. 30, no. 2, pp. 388-400, 2011.
- [16] F. J. Santos and M. T. Galceran, "Modern developments in gas chromatography-mass spectrometry-based environmental analysis," *J. Chromatogr. A*, vol. 1000, no. 1-2, pp. 125-151, 2003.
- [17] A. Orfanidis *et al.*, "A GC-MS method for the detection and quantitation of ten major

drugs of abuse in human hair samples,” *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.*, vol. 1047, pp. 141–150, 2017.

- [18] D. Bressanello *et al.*, “Urinary metabolic fingerprinting of mice with diet-induced metabolic derangements by parallel dual secondary column-dual detection two-dimensional comprehensive gas chromatography,” *J. Chromatogr. A*, vol. 1361, pp. 265–276, 2014.
- [19] S. Squara *et al.*, “Extending the breadth of saliva metabolome fingerprinting by smart template strategies and effective pattern realignment on comprehensive two-dimensional gas chromatographic data,” *Anal. Bioanal. Chem.*, no. 0123456789, Jan. 2023.
- [20] J. Young, “Doping with testosterone and androgenic/anabolic steroids: Impact on health, screening tools and medical care,” *Ann. Endocrinol. (Paris)*, vol. 84, no. 3, pp. 401–405, 2023.
- [21] M. Bajo-Fernández, É. A. Souza-Silva, C. Barbas, M. F. Rey-Stolle, and A. García, “GC-MS-based metabolomics of volatile organic compounds in exhaled breath: applications in health and disease. A review,” *Front. Mol. Biosci.*, vol. 10, no. January, 2023.
- [22] L. Amendola, F. Molaioni, and F. Botrè, “Detection of beta-blockers in human urine by GC-MS-MS-EI: Perspectives for the antidoping control,” *J. Pharm. Biomed. Anal.*, vol. 23, no. 1, pp. 211–221, 2000.
- [23] A. Sampat, M. Lopatka, M. Sjerps, G. Vivo-Truyols, P. Schoenmakers, and A. van Asten, “Forensic potential of comprehensive two-dimensional gas chromatography,” *TrAC Trends Anal. Chem.*, vol. 80, pp. 345–363, Jun. 2016.
- [24] S. J. Lehotay and J. Hajšlová, “Application of gas chromatography in food analysis,” *TrAC - Trends Anal. Chem.*, vol. 21, no. 9–10, pp. 686–697, 2002.
- [25] D. J. Beale *et al.*, “Review of recent developments in GC–MS approaches to metabolomics-based research,” *Metabolomics*, vol. 14, no. 11, p. 152, 2018.
- [26] W. Weckwerth, “Unpredictability of metabolism-the key role of metabolomics science in combination with next-generation genome sequencing,” *Anal. Bioanal. Chem.*, vol. 400, no. 7, pp. 1967–1978, 2011.
- [27] D.-W. Katja, *Practical Gas Chromatography: A Comprehensive Reference*. 2014. .
- [28] M. Zhou *et al.*, “Lipidomics and GC-MS analyses of key lipids for volatile compound formation in crayfish meat under different thermal processing treatments,” *Lwt*, vol. 189, no. November, p. 115522, 2023.
- [29] K. Jurowski, K. Kochan, J. Walczak, M. Barańska, W. Piekoszewski, and B. Buszewski, “Analytical Techniques in Lipidomics: State of the Art,” *Crit. Rev. Anal. Chem.*, vol. 47, no. 5, pp. 418–437, 2017.
- [30] J. H. VAN DE KAMER, K. W. GERRITSMA, and E. J. WANSINK, “Gas-liquid partition chromatography: the separation and micro-estimation of volatile fatty acids from formic acid to dodecanoic acid,” *Biochem. J.*, vol. 61, no. 1, pp. 174–176, 1955.
- [31] F. R. Cropper and A. Heywood, “Analytical separation of the methyl esters of the C₁₂-C₂₂ fatty acids by vapour-phase chromatography,” *Nature*, vol. 172, no. 4389, pp. 1101–

- [32] E. D. Dodds, M. R. McCoy, L. D. Rea, and J. M. Kennish, "Gas chromatographic quantification of fatty acid methyl esters: Flame ionization detection vs. electron impact mass spectrometry," *Lipids*, vol. 40, no. 4, pp. 419–428, 2005.
- [33] M. M. Ulaszewska *et al.*, "Nutrimetabolomics: An Integrative Action for Metabolomic Analyses in Human Nutritional Studies," *Mol. Nutr. Food Res.*, vol. 63, no. 1, p. 1800384, Jan. 2019.
- [34] S. Squara *et al.*, "Artificial intelligence decision making tools in food metabolomics: Data fusion unravels synergies within the hazelnut (*Corylus avellana* L.) metabolome and improves quality prediction," *Food Res. Int.*, vol. 194, p. 114873, Oct. 2024.
- [35] J. Cho, J. Lee, C. U. Lim, and J. Ahn, "Quantification of pesticides in food crops using QuEChERS approaches and GC-MS/MS," *Food Addit. Contam. - Part A Chem. Anal. Control. Expo. Risk Assess.*, vol. 33, no. 12, pp. 1803–1816, 2016.
- [36] A. Arena, M. Zoccali, L. Mondello, and P. Q. Tranchida, "A method for the determination of 70 pesticides in extra virgin olive oil based on a limited-volume solvent extraction step prior to comprehensive two-dimensional gas chromatography-tandem mass spectrometry," *Anal. Bioanal. Chem.*, vol. 415, no. 13, pp. 2459–2469, 2023.
- [37] E. Asensio, J. Uranga, and C. Nerín, "Analysis of potential migration compounds from silicone molds for food contact by SPME-GC-MS," *Food Chem. Toxicol.*, vol. 165, no. May, p. 113130, 2022.
- [38] M. V. Russo, P. Avino, and I. Notardonato, "Extraction and GC-MS analysis of phthalate esters in food matrices : a review," pp. 37023–37043, 2015.
- [39] C. Cordero, J. Kiefl, P. Schieberle, S. E. Reichenbach, and C. Bicchi, "Comprehensive two-dimensional gas chromatography and food sensory properties: potential and challenges," *Anal. Bioanal. Chem.*, vol. 407, no. 1, pp. 169–191, Jan. 2015.
- [40] B. Quintanilla-Casas *et al.*, "Virgin olive oil volatile fingerprint and chemometrics: Towards an instrumental screening tool to grade the sensory quality," *LWT*, vol. 121, p. 108936, 2020.
- [41] E. Hong *et al.*, "Modern analytical methods for the detection of food fraud and adulteration by food category," *Journal of the Science of Food and Agriculture*. 2017.
- [42] N. S. Brillhante, H. R. Bizzo, A. Caratti, S. Squara, and C. Cordero, "Artificial intelligence tools and concepts give access to authenticity and quality information in Brazilian olive oil volatilome," *J. Food Compos. Anal.*, vol. 136, p. 106826, Dec. 2024.
- [43] A. Caratti *et al.*, "Beyond current quality indices: Quantitative volatilomics unrevealed cultivar traits, harvesting practices impact, and aroma blueprint of extra-virgin olive oils," *J. Food Compos. Anal.*, vol. 137, no. September 2024, p. 106975, Jan. 2025.
- [44] P. Gómez-Cortés, G. L. Sacks, and J. T. Brenna, "Quantitative analysis of volatiles in edible oils following accelerated oxidation using broad spectrum isotope standards," *Food Chem.*, vol. 174, pp. 310–318, 2015.
- [45] K. Ridgway, S. P. D. Lalljie, and R. M. Smith, "Analysis of food taints and off-flavours: A review," *Food Addit. Contam. - Part A Chem. Anal. Control. Expo. Risk Assess.*, vol. 27, no. 2, pp. 146–168, Feb. 2010.
- [46] P. Caldentey, M. D. Fumi, V. Mazzoleni, and M. Careri, "Volatile compounds produced

by microorganisms isolated from cork,” *Flavour Fragr. J.*, vol. 13, no. 3, pp. 185–188, 1998.

- [47] R. M. Callejon, A. M. Troncoso, and M. L. Morales, “Analysis for chloroanisoles and chlorophenols in cork by stir bar sorptive extraction and gas chromatography-mass spectrometry,” *Talanta*, vol. 71, no. 5, pp. 2092–2097, 2007.
- [48] A. Kende *et al.*, “High throughput quantitative volatile profiling of melons with silicone rod extraction – thermal desorption – GC–MS for plant breeding line selection,” *Food Chem.*, vol. 270, no. July 2018, pp. 368–374, 2019.
- [49] K. J. Margolin Eren, H. F. Prest, and A. Amirav, “Nitrogen and hydrogen as carrier and make-up gases for GC-MS with Cold EI,” *J. Mass Spectrom.*, vol. 57, no. 5, pp. 1–11, 2022.
- [50] E. Pella, “The use of argon as carrier gas in the microdetermination of oxygen in organic compounds,” *Anal. Chim. Acta*, vol. 35, no. i, pp. 96–108, 1966.
- [51] N. A. Zaidi, M. W. Tahir, M. J. Vellekoop, and W. Lang, “A gas chromatographic system for the detection of ethylene gas using ambient air as a carrier gas,” *Sensors (Switzerland)*, vol. 17, no. 10, pp. 1–11, 2017.
- [52] R. L. Grob and E. F. Barry, *Modern Practice of Gas Chromatography*, vol. 29, no. 9, 1978.
- [53] L. M. Blumberg, “Theory of fast capillary gas chromatography. Part 1: Column efficiency,” *J. High Resolut. Chromatogr.*, vol. 20, no. 11, pp. 597–604, 1997.
- [54] C. N. Nnaji, K. C. Williams, J. M. Bishop, and G. F. Verbeck, “Hydrogen as a GC/MS carrier and buffer gas for use in forensic laboratories,” *Sci. Justice*, vol. 55, no. 3, pp. 162–167, 2015.
- [55] J. Ellis and K. Kounovsky-Shafer, “Hydrogen or Helium Conservation in Gas Chromatography Mass Spectrometry: How We Chose What Was Right for Our Laboratory,” *ACS Chem. Heal. Saf.*, vol. 30, no. 4, pp. 151–155, 2023.
- [56] J. A. Muñoz-Guerra, P. Prado, and S. V. García-Tenorio, “Use of hydrogen as a carrier gas for the analysis of steroids with anabolic activity by gas chromatography-mass spectrometry,” *J. Chromatogr. A*, vol. 1218, no. 41, pp. 7365–7370, 2011.
- [57] Y. P. Lin, Y. L. Lee, C. Y. Hung, C. F. Chang, and Y. Chen, “Detection of adulterated drugs in traditional Chinese medicine and dietary supplements using hydrogen as a carrier gas,” *PLoS One*, vol. 13, no. 10, pp. 1–21, 2018.
- [58] S. Impens, K. De Wasch, and H. De Brabander, “Determination of anabolic steroids with gas chromatography-ion trap mass spectrometry using hydrogen as carrier gas,” *Rapid Commun. Mass Spectrom.*, vol. 15, no. 24, pp. 2409–2414, 2001.
- [59] N. Oshima, M. Takagi, S. Sakai, and Y. Ikarashi, “Comparison of Helium-Alternative Carrier Gases for Gas Chromatography/Mass Spectrometry of Standard Test Methods for Indoor Air Quality Guidelines in Japan,” *BPB Reports*, vol. 5, no. 4, pp. 84–87, 2022.
- [60] L. M. Blumberg and M. S. Klee, “Method Translation and Retention Time Locking in Partition GC,” vol. 70, no. 18, pp. 3828–3839, 1998.
- [61] M. S. Klee and L. M. Blumberg, “Theoretical and Practical Aspects of Fast Gas Chromatography and Method Translation,” *J. Chromatogr. Sci.*, vol. 40, no. 5, pp. 234–247, 2002.
- [62] C. A. Cramers, H. G. Janssen, M. M. Van Deursen, and P. A. Leclercq, “High-speed gas

chromatography: An overview of various concepts,” *J. Chromatogr. A*, vol. 856, no. 1–2, pp. 315–329, 1999.

- [63] L. M. Blumberg, “Theory of fast capillary gas chromatography part. 2: Speed of analysis,” *J. High Resolut. Chromatogr.*, vol. 20, no. 12, pp. 679–687, 1997.
- [64] C. Bicchi, C. Brunelli, C. Cordero, P. Rubiolo, M. Galli, and A. Sironi, “Direct resistively heated column gas chromatography (Ultrafast module-GC) for high-speed analysis of essential oils of differing complexities,” *J. Chromatogr. A*, vol. 1024, no. 1–2, pp. 195–207, 2004.
- [65] C. A. Cramers and P. A. Leclercq, “Strategies for speed optimisation in gas chromatography: An overview,” *J. Chromatogr. A*, vol. 842, no. 1–2, pp. 3–13, 1999.
- [66] C. Bicchi, C. Brunelli, M. Galli, and A. Sironi, “Conventional inner diameter short capillary columns: An approach to speeding up gas chromatographic analysis of medium complexity samples,” *J. Chromatogr. A*, vol. 931, no. 1–2, pp. 129–140, 2001.
- [67] R. Jian, Y. Huang, S. Lai, L. Sung, and C. Lu, “Compact instrumentation of a μ -GC for real time analysis of sub-ppb VOC mixtures,” *Microchem. J.*, vol. 108, pp. 161–167, 2013.
- [68] D. W. You, Y. S. Seon, Y. Jang, J. Bang, J. S. Oh, and K. W. Jung, “A portable gas chromatograph for real-time monitoring of aromatic volatile organic compounds in air samples,” *J. Chromatogr. A*, vol. 1625, 2020.
- [69] C. Duan, J. Li, Y. Zhang, K. Ding, X. Geng, and Y. Guan, “Portable instruments for on-site analysis of environmental samples,” *TrAC - Trends Anal. Chem.*, vol. 154, p. 116653, 2022.
- [70] M. Mittermüller and D. A. Volmer, “Micro- and nanostructures and their application in gas chromatography,” *Analyst*, vol. 137, no. 14, pp. 3195–3201, 2012.
- [71] A. You, M. Be, and I. In, “A micro gas chromatography column with a micro thermal conductivity detector for volatile organic compound analysis,” no. August 2022, pp. 4–9, 2016.
- [72] J. Sun *et al.*, “Compact Prototype GC-PID System integrated with Micro-PC and Micro GC Column,” *J. Micromechanics Microengineering*, 2018.
- [73] S. Zampolli, I. Elmi, G. Carlo, L. Masini, F. Bonafè, and F. Zardi, “Compact-GC platform : A flexible system integration strategy for a completely microsystems-based gas-chromatograph,” *Sensors Actuators B Chem.*, vol. 305, no. November 2019, p. 127444, 2020.
- [74] M. L. Woebkenberg, C. Editor, R. G. E. Burroughs, and M. W. Tabor, “Portable Gas Chromatographs,” vol. 14, pp. 159–163, 1999.
- [75] P. E. Leary, B. W. Kammrath, and J. A. Reffner, “Field-Portable Gas Chromatography–Mass Spectrometry,” *Encycl. Anal. Chem.*, pp. 1–23, 2018.
- [76] B. A. Eckenrode, “Environmental and Forensic Applications of Field-Portable GC-MS: An Overview,” *J. Am. Soc. Mass Spectrom.*, vol. 12, no. 06, pp. 683–693, 2001.
- [77] S. C. Terry, J. H. Herman, and J. B. Angell, “A Gas Chromatographic Air Analyzer Fabricated on a Silicon Wafer,” *IEEE Trans. Electron Devices*, vol. 26, no. 12, pp. 1880–1886, 1979.
- [78] P. A. Smith, “Person-portable gas chromatography: Rapid temperature program

operation through resistive heating of columns with inherently low thermal mass properties,” *J. Chromatogr. A*, vol. 1261, pp. 37–45, 2012.

- [79] I. Azzouz *et al.*, “Review of stationary phases for microelectromechanical systems in gas chromatography: Feasibility and separations,” *Anal. Bioanal. Chem.*, vol. 406, no. 4, pp. 981–994, 2014.
- [80] P. Rubiolo, B. Sgorbini, E. Liberto, C. Cordero, and C. Bicchi, *The Chemistry and Biology of Volatiles. Analysis of the Plant Volatile Fraction*. 2010.
- [81] Y. Pico, *CHEMICAL ANALYSIS OF FOOD. Techniques and Applications*, Second edi. 2020.
- [82] C. Cagliero *et al.*, “Conventional and enantioselective gas chromatography with microfabricated planar columns for analysis of real-world samples of plant volatile fraction,” *J. Chromatogr. A*, vol. 1429, pp. 329–339, 2016.
- [83] J. Luong, R. Gras, R. A. Shellie, and H. J. Cortes, “Applications of planar microfluidic devices and gas chromatography for complex problem solving,” *J. Sep. Sci.*, vol. 36, no. 1, pp. 182–191, 2013.

Chapter 3

Comprehensive Two-Dimensional Gas Chromatography-Mass Spectrometry: Principles and Tools

This chapter provides an overview of the principles and tools of comprehensive two-dimensional gas chromatography-mass spectrometry (GC×GC-MS). GC×GC addresses the limitations of one-dimensional gas chromatography (1D GC) and classical multidimensional gas chromatography (MDGC) by employing a modulator to enhance peak capacity and selectivity significantly. Advances in thermal and flow modulation, particularly with cryogenic modulators, have refined the technique. Coupled with mass spectrometry (MS), GC×GC-MS adds a separation dimension for analyzing complex mixtures. The chapter also discusses software advancements that support data processing and visualization, showcasing the evolution of GC×GC-MS technology.

3.1 Introduction

One-dimensional gas chromatography (1D GC) separations are governed by two fundamental factors: peak capacity (n_c) and selectivity. The latter is tightly related to the chemistry of the stationary phase, while the former to the available separation space. Peak capacity can be measured easily by calculating the theoretical number of peaks that can be stacked side-by-side [with a specific average peak base width (w_b) and resolution value (e.g., 1)], between the void time and the end of the temperature program. A “normal” 1D GC n_c value is in the range 300-500 (using a 30 m x 0.25 mm ID x 0.25 μ m df capillary column).

Heart-cutting (or classical) multidimensional GC (MDGC) was introduced rather shortly after the invention of GC itself [1]. In heart-cutting MDGC, a dedicated transfer device (e.g., a Deans switch) enables the passage of time-fractions of eluate from a first-dimension (1D) column to a second-dimension (2D) one. The two columns are characterized by a different stationary phase, and thus possess a different selectivity. The use of more than two columns is rare [2]. Heart-cutting MDGC is characterized by enhanced n_c and selectivity compared to 1D GC, and represented a significant advancement in GC.

Comprehensive two-dimensional gas chromatography (GC $\times$ GC) was first described in 1991 [3], and stemmed from research performed over previous years by its inventor J.B. Phillips, on thermal desorption modulation (TDM) [4]. The concept of TDM - a means of analyte entrapment (or accumulation), re-concentration (or focusing) and then release (or re-injection) - extended to a transfer device (named modulator), located between a 1D and 2D column with different stationary phases, enabled GC $\times$ GC analysis. The main difference, with respect to classical MDGC, is that the modulator performs continuous and sequential heart-cutting across the entire 1D eluate. The first modulator was heated electrically, was of simple construction and potentially a low-cost, efficient modulator. However, it was characterized by fragility due mainly to the difficulty of reliable temperature control. Over the decades different thermal modulators have been introduced, with cryogenic modulators (CMs) conquering the scene, also commercially. Flow modulators (FM) currently represent an interesting alternative, with various devices nowadays available, stemming from research performed by J. Seeley and co-workers [5].

The modulator works like a beating heart in the GC $\times$ GC system, inasmuch that it enables a sequence of brief events involving accumulation and release. Analyte re-concentration can occur (thermal modulation), or not (flow modulation). The different configurations of a 1D GC, a MDGC and A GC $\times$ GC instrument are shown in **Figure 3.1**. The two GC columns can be positioned in a single, or two GC ovens, with the latter being by far the best option. Separation times on the 1D column are altogether similar to those observed in 1D GC, while those on the 2D column are much faster, normally between 1-8 s. For such a reason, 1D columns are usually of a conventional length (e.g., 30 m), while 2D ones are much shorter, and deliver very narrow peaks (w_b values are normally in the range 100-400 ms, in particular when using thermal modulation) to the detector. In such a respect, the detector used must ideally have no, or a

very limited effect on band broadening, and a 50-Hz minimum acquisition frequency (for satisfactory peak construction).

The time required to perform a modulation event is defined as the modulation period (P_M), and remains constant during the analysis (in particular cases, a variable P_M can be used). The P_M , which is equivalent to the ^2D analysis time, should enable 3-4 modulations per peak - a lower number (under-sampling) can cause a loss of resolution already achieved on the first column (due to remixing during the modulation process), while over-sampling can lead to a reduction of analyte detectability and an overly short modulation period. Detection is performed at the outlet of the ^2D column, thus "untransformed" GC $\times$ GC chromatograms are one-dimensional; considering a P_M of 6 s, and an elution window of 1 h, then a chromatogram containing six hundred 6-s sub-chromatograms would be formed. As will be seen, such chromatograms are most-commonly visualized in a two-dimensional manner, with ^1D and ^2D retention time (t_R and ^2t_R) values positioned along an x- and y-axis, respectively. Peaks are represented with an oval/circular shape (unless overloading occurs), with size and colour related to signal intensity. Apart from the enhanced selectivity, the peak capacity generated becomes very high, greatly increased compared to classical MDGC, and of about an order-of-magnitude higher with respect to $1D$ GC. A considerable enhancement in analyte detectability is also obtained, specifically when using thermal modulation. A further advantage relates to the chromatogram formation of group-type analyte patterns, when homologous series of compounds are analyzed. As for classical MDGC, the use of more than two columns is rarely reported [6].

Mass spectrometry (MS) is by far the most important form of detection in GC (and not only), and consists in its own right as a further dimension of separation, on a mass-to-charge (m/z) basis (in the mass analyzer). The technology GC $\times$ GC-MS, though extremely powerful, is not fully orthogonal. The three analytical dimensions are, in part, correlated due to the fact that separations in both GC and MS are dominated by vapour pressures, which in turn are related (though not entirely) to molecular mass. Comprehensive $2D$ GC chromatograms are often characterized by 100s, if not 1000s, of peaks. In such a respect, the use of MS is nearly always mandatory.

The objective of the present chapter is to provide a critical view on state-of-the-art GC $\times$ GC-MS and related software packages (there has been great technological evolution over the last three decades).

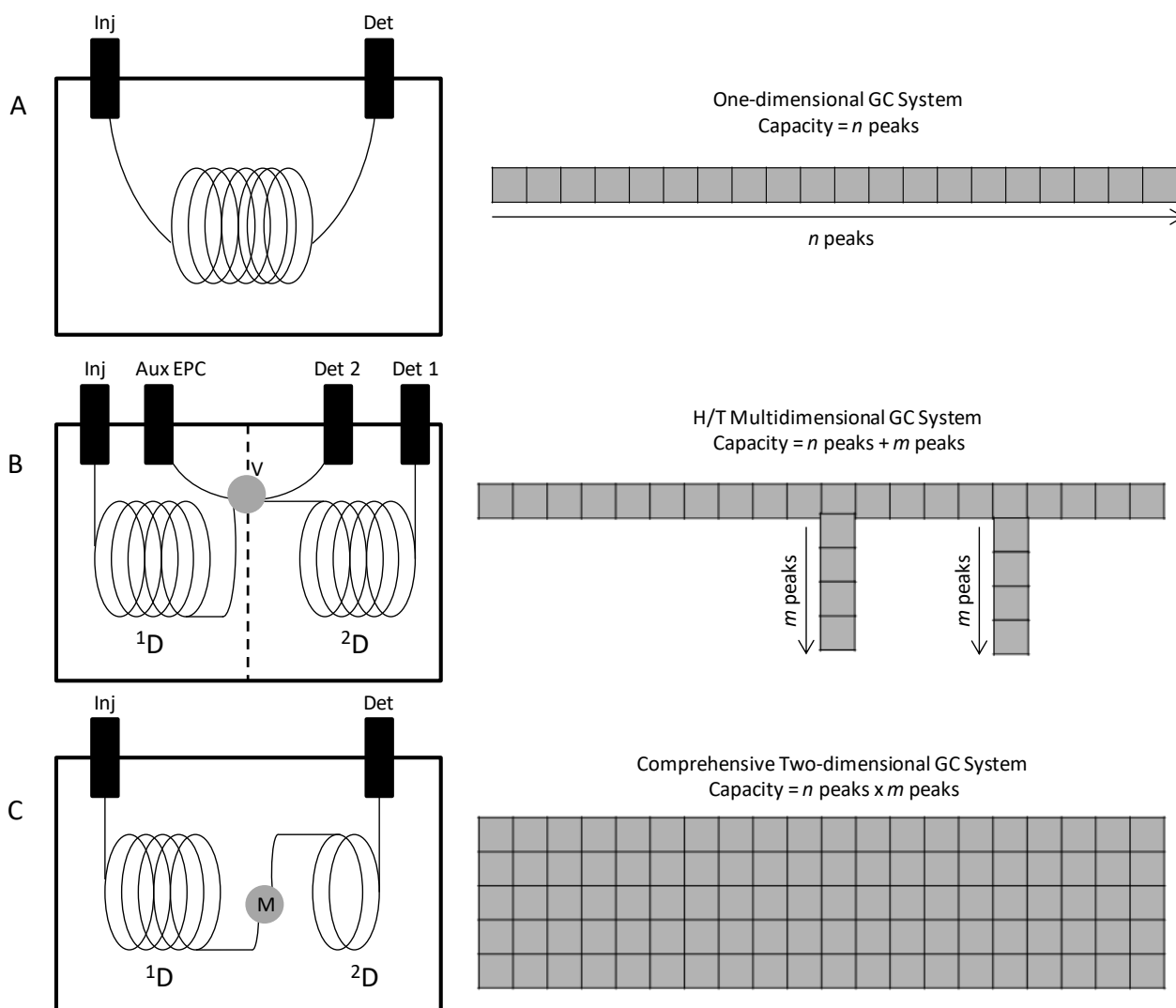


Figure 3.1: Schematic diagrams of a 1D GC and multidimensional GC instruments. 1A shows a typical 1D GC; 1B a H/C MDGC while 1C shows a comprehensive two-dimensional system. Inj: injector; aux EPC: auxiliary gas Electron Pressure Controller; Det 1: primary detector; Det 2: monitor detector; V: switching interface; M: modulator; 1D: first dimension column; 2D: second dimension column. Reproduced with permission from [7].

3.2 Experiments set-up in different application fields

The basic set-up for a GC $\times$ GC-MS instrument is illustrated in **Figure 3.1C**. Injection conditions are the same as in 1D GC, and will not be described. There can be some differences related to the sample preparation process, worthy of note. Other important instrumental aspects - column selection, modulation, mass spectrometry and data collection - relate specifically to the three-dimensional technology.

3.2.1 Sample Preparation

In GC $\times$ GC, as for 1D GC, the direct injection of liquid samples is possible provided that the presence of non-volatile interferences is suppressed and/or minimized through appropriate strategies. For raw fragrances and cosmetic products, where GC $\times$ GC is employed for targeted

quantitative determinations (e.g., volatile suspected allergens)[8], [9] and fingerprinting (e.g., authentication, fraud detection, potent odorant identification/qualification [10]), non-volatile components can be trapped inside the injector liner, packed with inert materials. In alternative, programmable temperature vaporizer injectors with cryo-focussing options can accommodate large sample volumes, thus reaching low detection limits, and selectively retain the non-volatile interferences. Another application area where direct sample analysis is feasible is petrochemistry, specifically middle distillates (C₉-C₂₀), such as diesel and jet-fuels [11]. Heavier fractions (e.g., C₂₀+) may sometimes require pre-fractionation by liquid-liquid extraction, solid-phase extraction and/or dedicated treatments [12] (e.g., deasphalting, column fractionation, supercritical fluid extraction). In some cases, to safeguard the capillary column, the back-flushing option is helpful. With such an approach, heavy compounds can be back-flushed through the split port of the injector, by a carrier gas stream delivered in a counter-current manner.

Gas samples and volatile analytes

Gas samples, as well as the fraction of volatile and semi-volatile components from condensed samples (liquid or solid), can be analyzed by GC×GC through headspace (HS) approaches. By definition, HS extraction enables the sampling of the gaseous or vapor phase in equilibrium with a solid or liquid matrix [13]. It can be conducted in static (S-HS) or dynamic (D-HS) conditions. Moreover, to achieve an adequate analyte detectability, while compensating for discrimination toward semi-volatile compounds, analyte pre-concentration onto suitable extraction materials either by sorption or adsorption is largely adopted [i.e., high-concentration capacity (HCC) techniques] [14]. Solid-phase microextraction (SPME) with conventional devices (30-100 µm film thickness × 1-2 cm length) or with higher amounts of sorbent material, such as for SPME Arrow (100-250 µm film thickness × 1-2 cm length) or HiSorb™ [e.g., an HS sorptive extraction approach available for automated HS and in-solution sampling] are the most common. In volatilomics, HS SPME enables comprehensive profiling and fingerprinting of volatile and semi-volatile components with minimal sample manipulation. It enables accurate quantification, through multiple headspace extraction or standard addition, performed with fully-automated robotic stations that guarantee high sample throughput and sampling repeatability [15].

The SPME process can be conducted at sub-ambient pressure, speeding up analyte partitioning in the HS while reducing the interfacial resistance to mass transfer [16]. In such a respect, vacuum-assisted HS SPME (or Vac-HS SPME) has demonstrated its effectiveness in the enrichment of semi-volatile compounds or in the low-temperature sampling of perishable food volatiles (e.g., fish [16]). Alternatively, multi-cumulative trapping SPME or D-HS approaches that accumulate analytes onto sorbent/adsorbent materials (e.g., solid-phase dynamic extraction) represent valid options to maximize the information potential of a GC×GC run. Application areas where HCC techniques in combination with GC×GC play a crucial role are food volatilomics, sensomics, forensics, and clinical applications [17]–[19].

Non-volatile analytes

Metabolomics targets polar/non-volatile analytes and require derivatization to allow for GC amenability. The primary metabolome signature in biological fluids (e.g., urine, plasma, saliva, cerebrospinal fluid, etc.), tissues, and food samples can be profiled by

converting free amino acids, low molecular weight (MW) saccharides, organic acids, and amines into more volatile derivatives. The most common procedure includes protein precipitation, solvent extraction to remove lipids and/or high-MW specialized metabolites (e.g., in vegetal foods), followed by methoximation with *O*-methylhydroxylamine hydrochloride to stabilize ketones and silylation with trimethylsilyl reagents [e.g., *N*, *O*-bis(trimethylsilyl)trifluoroacetamide, *N*-Methyl-*N*-(trimethylsilyl)trifluoroacetamide] [20].

The comprehensive fatty acid (FA) composition of lipid fractions can be very effectively elucidated by using GC×GC due to the generation of structured retention patterns. Acid- or base-catalyzed transesterification in the presence of an alcohol, mainly methanol [21], is a common procedure. Esterified FAs are hydrolyzed and converted in the corresponding methyl esters (FAMES) with high yields and minimal risks for polyunsaturated homologs that are prone to autoxidation and decomposition. Based on the matrix characteristics (i.e., plasma, tissues, oils, intact food, sediments, etc.) and compositional features, the derivatization protocol can be adjusted to avoid artifact formation and analyte degradation.

Hyphenation with liquid chromatography

Comprehensive 2D GC can be also online connected to liquid chromatography (LC), by using dedicated interfaces. This coupling is a solution of choice for pre-fractionation and clean-up of complex samples when the focus is group-type analysis. Moreover, operating by online sequential separations, sample manipulation is minimized, analysis time is shorter and cross-contamination or artifact formation due to atmospheric oxygen or moisture exposure is limited, if not eliminated. Different interfaces are available, mostly developed for LC-GC coupling [22]. Applications demonstrating the usefulness of such a combination include speciation of mineral oil hydrocarbon contamination in edible oils and food, and in-depth characterization of essential oil fractions [23].

3.2.2 Column Geometries and Orthogonality

With regard to the topic of this sub-section a distinction must be made between cryogenic and flow modulation, in relation to column dimensions. The stationary phases obviously relate to all modulator types. In cryogenic-modulated devices, a popular column combination is: low-polarity (e.g., “5% diphenyl”) 30 m x 0.25 mm ID x 0.25 µm df ¹D column and a medium-polarity [e.g., polyethylene glycol (PEG)] 1 m x 0.10 mm ID x 0.10 µm df ²D (*n_c* 20-30) one (considerations on column selection for FM analyses are provided in the next sub-section). For convenience, such a column set is defined as orthogonal (defined by some Authors as “normal-phase”) because after a ¹D separation relying mainly on vapour pressures, potentially-overlapping isovolatile compounds will then be separated on the ²D column on the basis of specific functional group-stationary phase interactions. However, an inversion in the polarity of the stationary phases (medium-polarity × low-polarity; defined by some Authors as “reversed-phase”) can often provide a better result [24]. Moreover, stationary-phase chemistries with intermediate polarity (e.g., PEG + “50% diphenyl”) have been combined with success in flavor & fragrance applications. Even though there is no general rule, it is preferable to use as, ¹D column, the stationary phase with the highest selectivity (and thus separation power) for the specific sample. In such a manner, the extent of compound

overlapping at the ^2D inlet will be reduced [24]; such an effect is evident in **Figure 3.2**, which illustrates the GC $\times$ GC-MS [low-resolution time-of-flight (LR TOF) MS was used] separation of coffee volatiles on two column sets.

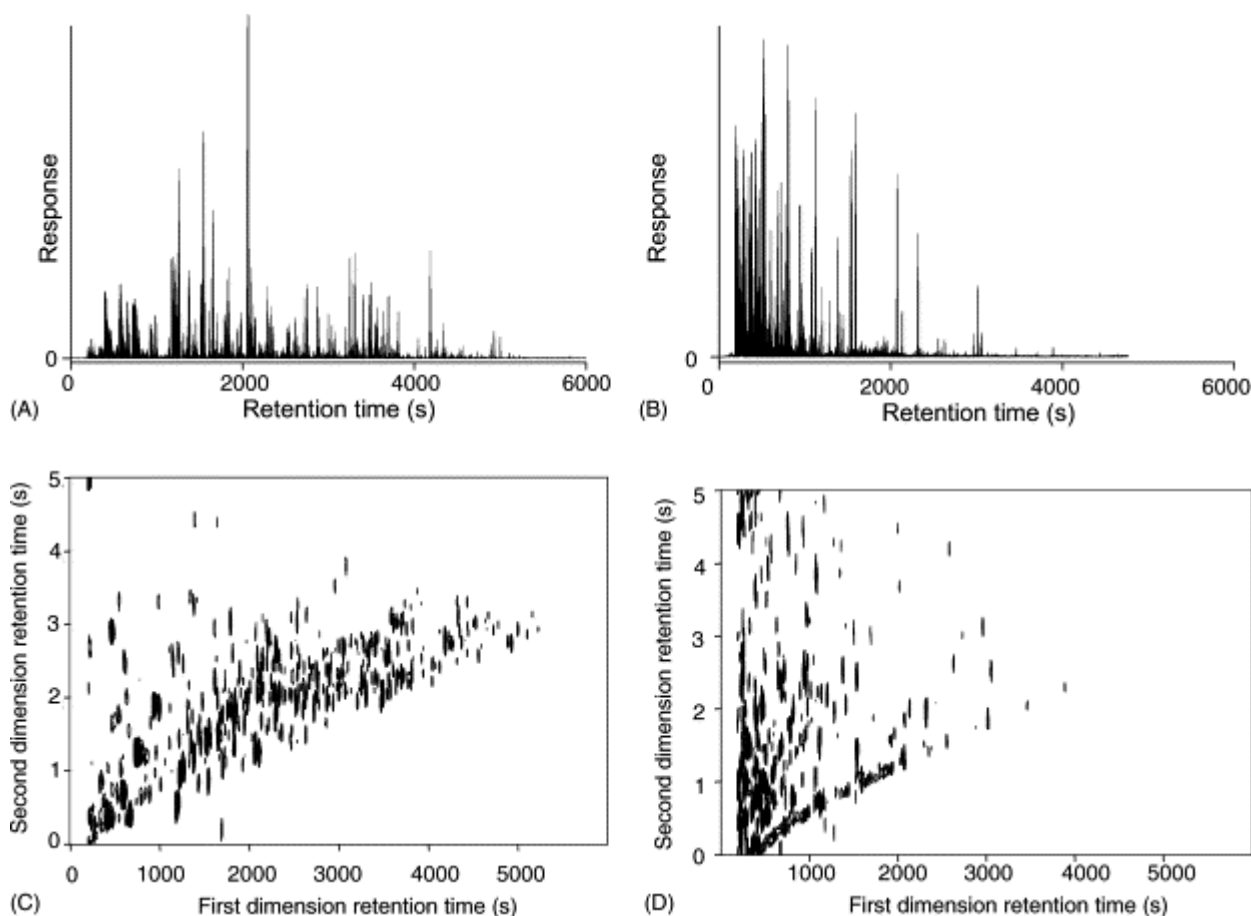


Figure 3.2: Headspace solid-phase microextraction GC $\times$ GC-LR TOFMS analysis of coffee volatiles by using a (A and B) medium-polarity + low-polarity and a (C and D) low-polarity + medium-polarity column set. Untransformed chromatograms are shown in A and C. Reproduced with permission from Elsevier [24].

A further issue to be considered is the organized nature of 2D chromatograms. For example, PEG is a highly selective stationary phase for the separation of FAMES. Hence, the 5% diphenyl + PEG combination provides an outstanding result if the formation of group-type patterns is considered [25]. In terms of CM enhancement in analyte detectability, it is not possible to provide specific information due to the many variables involved (i.e., gas flow, ^2D analyses temperatures, P_M , analyte-stationary phase interactions, MS acquisition frequency, etc.). However, when using a 1 m x 0.10 mm ID x 0.10 μ m df ^2D column, an increase in signal-to-noise ratio (s/n) values in the range 10-20 can be expected. The chromatogram signals shown in **Figures 3.2A** and **3.2B** are so narrow that they resemble a sequence of lines, and not peaks.

The main disadvantages of using a 0.25 mm + 0.10 mm ID combination are: a) frequent occurrence of ^2D column overloading when high-intensity ^1D peaks are modulated; b) mismatch between ^1D and ^2D gas linear velocities. With regard to the latter issue, a ^1D average gas velocity of 25-30 cm s $^{-1}$ will lead to an excessively-high ^2D one, a factor that can be counterbalanced by using independent temperature programs. An interesting alternative

consists in the use of columns with the same ID (i.e., 0.25 mm), allowing for a diminished possibility of ^2D column overloading, and the possibility of generating ideal gas velocities in both dimensions. Furthermore, when using MS, low-pressure conditions will extend across the entire ^2D column, a positive factor for high-speed separations [26].

Whenever the ^2t_R of an analyte (e.g., 7 s) exceeds that of the P_M (e.g., 6 s) then it will elute during the successive modulation event, with a ^2t_R of 1 s. Such a phenomenon is called “wrap-around”, and can be resolved easily by increasing the P_M or the ^2D analysis temperature (it is usually not necessary to shorten the ^2D column). However, wrap-around is fully acceptable if it is not excessive (also in terms of band broadening) and it does not interfere with group-type patterns. Moreover, if wrap-around analytes elute in the void time space of the ^2D chromatogram then it can be considered as beneficial. The void time space is usually limited upperly by a (more-or-less) horizontal streak of ^1D column bleed extending across part of the ^2D chromatogram. The isolation of column bleed is beneficial on the general mass spectral quality.

3.2.3 Modulators for GC $\times$ GC

There has been considerable evolution in the field of modulation over the past three decades. The first commercial device - the thermal sweeper - was a moving system (a slotted rotating heater), based on thermal desorption, and it rapidly became obsolete [27]. A major breakthrough came with the introduction of the first CM (cooled by using CO_2), the longitudinally modulated cryogenic system (LMCS) [28]. The LMCS was a moving device, initially used as a means to increase analyte detectability, and later exploited to perform modulation. A static dual-jet CM, cooled through CO_2 , was introduced by Beens et al. [29], with the absence of moving parts representing an advantage. In both the LMCS and dual-jet CM, the release event was enabled by the GC oven heat. Further evolution was made with the introduction of the quad-jet and loop-type CMs, with these currently the most popular forms of modulation [6]. The quad-jet CM is formed of two couples of cooling/heating jets (C/H I and C/H II), positioned at the head of the ^2D column and distanced by about 5 cm. The jets direct their emissions onto an upstream and downstream point: jets C I and H II are activated at the same time, while the same occurs for jets C II and H I, in an alternate fashion. In a manner similar to TDM, two entrapment and release events occur, thus enabling dual-stage modulation. The loop type (dual-stage) CM is characterized by a pair of C/H jets, positioned perpendicularly one to the other. A delay loop (approx. 1 m length of uncoated/coated column) - positioned between the ^1D and ^2D columns - is coiled inside a metal assembly in such a manner to create two overlapping loops, coinciding with the initial (upstream) and final (downstream) part of the delay loop. The C jet, positioned vertically, operates continuously and directs its flow onto a small central segment of the overlapping loops. The H jet, positioned horizontally, operates at specific time intervals (e.g., every 5 s) for a brief period (e.g., 200 ms), heating the same central segment. Both the quad-jet and loop-type CMs perform cooling with liquid N_2 , enabling the entrapment of highly volatile compounds (e.g., C_4). Cryogenic-fluid free refrigeration units are also available (reducing safety issues related to laboratory air), enabling an entrapment temperature down to -90°C , sufficient for the

focusing of C8-C9 and above. The use of a heating gas (e.g., 300 °C), normally N₂, enables a more efficient release step compared to the exploitation of the GC oven heat.

A variety of FMs have been developed, and can be divided into in-line and out-line valve systems. Auxiliary pressure is used at the connection point between the analytical dimensions to enable modulation. The first GC×GC FM was introduced in 1998 by Bruckner et al., and was an in-line valve device [30].

Out-line valve systems (no direct connection between the columns and a valve) are currently the most commonly used, and can be further classified in high and low duty-cycle FMs. The duty cycle relates to the percentage of 1D eluate that exits the modulator. In out-line high duty-cycle FMs the volume of eluate introduced onto the 2D is usually high and does not undergo focusing. For such a reason, it is convenient to use a 2D column with a wider bore (e.g., 0.32 mm), compared to the 1D one, also because high 2D gas flows are normally used (e.g., 15-25 mL min⁻¹). Gas flows in the 1D column are usually “normal”. Accumulation occurs in some form of sampling chamber/loop, with this flushed through a brief pressure pulse (e.g., 150 ms). Such modulators are single stage. The use of high gas flows can be an issue when using MS, especially when pumping capacities are exceeded. In out-line low duty-cycle FMs, only a minor part of the 1D eluate is transferred to the 2D column at specific time intervals, corresponding to the P_M . The 1D eluate is directed to waste during the time interval between sequential re-injection events; as such, there are no accumulation-release processes, but rather waste-transfer events. If there are a sufficient number of transfer events per peak (i.e., 3), then there will be no loss of 1D analytical information due to the use of a waste line, apart from that related to the considerable reduction of analyte detectability. Column configurations can be altogether similar to those used in CM investigations, due to the narrow chromatography bands subjected to 2D analyses. Furthermore, 2D gas flows are much lower compared to high duty-cycle FMs.

3.3 Detection requirements in GC×GC

Comprehensive 2D GC detectors should be “fast” enough to collect a suitable number of data points per unit time, in relation to the 2D peak widths (very narrow when using thermal modulation). Detectors can be classified on the basis of the dimensionality of the data stream generated: single-channel and multi-channel detectors. Single-channel detectors generate one-dimensional, time-ordered data, consisting of intensity values for each chromatographic data point; for multi-channel detectors the data stream consists of an intensity value array (multi-dimensional array) for each chromatographic data point.

3.3.1 Single-channel detectors

Within the single-channel type, the flame ionization detector (FID) is the most common due to its response stability, linearity over wide concentration intervals, and the possibility of adopting structure-specific response factors for the reliable quantification of analytes, without external calibration [31]. In high duty-cycle FM GC×GC, an FID can be adopted in parallel with MS to accommodate part of the high 2D gas flows [32]–[34], also to enable more straightforward analyte quantification. The sulfur-chemiluminescence detector has been used for the speciation of sulfur-containing compounds in middle distillates and

diesel samples, although nowadays the availability of high-resolution MS (combined with GC) offers a larger flexibility and superior information capacity in heteroatom speciation [35]. Other selective single-channel detectors, such as the nitrogen-phosphorus, the electron-capture and the nitrogen-chemiluminescence detectors, have been used with various degrees of success, even though none have found widespread application.

3.3.2 Multi-channel detectors

Besides MS, the vacuum ultraviolet (VUV) detector enables the collection of information on analyte structure-specific features (i.e., related to UV absorption in the range 120-240 nm). Gas chromatography-VUV has been used for the qualitative analysis of complex mixtures (e.g., petrochemical, food, and environmental samples) since in the case of co-elution, time interval deconvolution supports single-analyte/class identification [36]. Fuel composition and other relevant properties (e.g., temperature dependence of density, viscosity, thermal conductivity, and heat capacity) have been estimated on the basis of the GC×GC-VUV direct analysis of fuel samples [37].

The first published paper on GC×GC-MS appeared in 1999, and it involved the use of single-quadrupole mass spectrometry (QMS) [38]. The Authors of that manuscript emphasized the technological limitations of the QMS used, proposing LR TOFMS as an “ideal solution”. In fact, at that time the QMS scanning speeds were far too low, and provided an insufficient number of spectra s^{-1} for satisfactory peak construction (and hence analyte quantification). Nowadays, QMS instrumentation has advanced considerably, with a reported scan speed of 30,000 amu s^{-1} , and the capability of providing consistent spectral profiles across narrow peaks (absence of, or limited peak skewing). The scan mode is used for untargeted analysis, while quantification is performed through selected-ion-monitoring (SIM) or extracted-ion-chromatograms (EICs). Current-day QMS instrumentation can provide both scan and SIM data in a single analysis, as the quadrupole can switch between both operational modes very rapidly.

Low-resolution TOFMS is characterized by extremely high acquisition frequencies (e.g., 1000 Hz), provides highly consistent spectral profiles, a factor tightly related to its deconvolution capabilities. Whereas QMS analyzes ions one at a time, true ion separations occur in the TOF mass analyzer, because it receives packets of accelerated ions, then detected across a specific mass range. The full-spectrum mode is used for untargeted analysis, while quantification is performed through EICs.

Low-resolution MS has gained a dominating role in the GC×GC-MS field, with LR TOFMS occupying the prime position, followed at a distance by QMS. Other forms of MS, i.e., triple quadrupole MS, high-resolution (HR) TOFMS, though used, are minor actors. Apart from economical requirements, the main reason relates to the analytical necessity (or not) to combine highly powerful forms of both GC and MS. It is highly difficult to find a sample for which GC×GC-LR MS is an insufficient tool, especially if cryogenic modulation is used.

3.4 Raw data interpretation

The amount of data collected by comprehensive multidimensional separations (in this case GC×GC) necessitates computer-assisted and automated processing [39]. Primary goals for computer processing include visualization, raw data/chromatogram processing, and sample analysis. Information technologies implemented in most of the commercial software packages guarantee full automation of sequence processing (and pre-processing) including artifact correction, 2D peak detection, analyte identification, and quantification. However, GC×GC is sometimes used for explorative analysis on just a few/selected samples. In this case, semi-automated processing is useful, as it allows visual interpretation and interactive analysis. As GC×GC transforms chemical samples into raw data, computer-assisted processing transforms GC×GC data into chemical information [40].

3.4.1 Data visualization and representation

Visualization of chromatogram data is a preliminary, yet fundamental step for qualitative analysis. By visualizing GC×GC data, the analyst can troubleshoot the chromatography, deduce general sample characteristics, examine specific 2D peaks, and compare chromatograms. Two-dimensional images allow a comprehensive overview of the GC×GC data array, while three-dimensional visualization gives access to quantitative relationships over a large dynamic range; tabular views show numeric values of the data, and annotations (graphical or text) provide additional information.

For visualization, the 1D time-ordered data generated by single channel detector(s) is rasterized for the two chromatographic dimensions according to the P_M . In the case of MS or VUV detection, spectral data require indexing for access and visualization. The GC×GC data is visualized as a 2D image by reorganizing raw data as a raster - a 2D array, matrix, or grid of pixels where pixel values correspond to the detector signal. To note, with MS and/or VUV detection, each pixel corresponds to a detector event, i.e., a spectrum. Commercial software packages use a right-handed Cartesian coordinate system to arrange data values collected during a single modulation cycle as a column of pixels; therefore, the ordinate (y axis, bottom-to-top) is sourced from the ^2t_R , and the abscissa (x axis, left-to-right) is sourced from the ^1t_R .

Pseudocolorization enables image presentation, with it mapping the pixel values (i.e., detector intensities) into colors. Simple colorization as a grayscale is the simplest way to render the scalar values collected from single-channel detectors into images. From multispectral data (e.g., MS and VUV), scalars can be extracted in various ways. If all the values of each spectrum are computed, the total intensity count (TIC) of each data point is used to generate the 2D image; if only selected channels are extracted for the computation, an extracted intensity count (XIC) 2D image is generated. Noteworthy is the possibility of extracting selected channels (i.e., m/z values) or multiple channels according to specific rules [i.e., common fragments generated by electron ionization (EI) in standardized conditions] by scripting into the data array, with it enabling specific visualizations of chemical classes or homolog series [41]. Examples of selective visualization by scripting are of great help in the group-type analysis of petrochemical samples [42]. The computer-based system of scripting is also used in data processing to identify and/or locate over the 2D retention space chemical classes based on known/common fragmentation rules [43].

Since humans show different sensitivities to light frequencies, trichromatic color models like the RGB (with values for red, green, and blue) modulated on HSV (hue, saturation, and brightness) are widely used for visualization. **Figure 3.3A-E** show 1D-2D-3D visualizations of a GC×GC-MS chromatogram. The technique GC×GC-MS has dedicated visualization tools that explore the MS metadata and generate images based on these data.

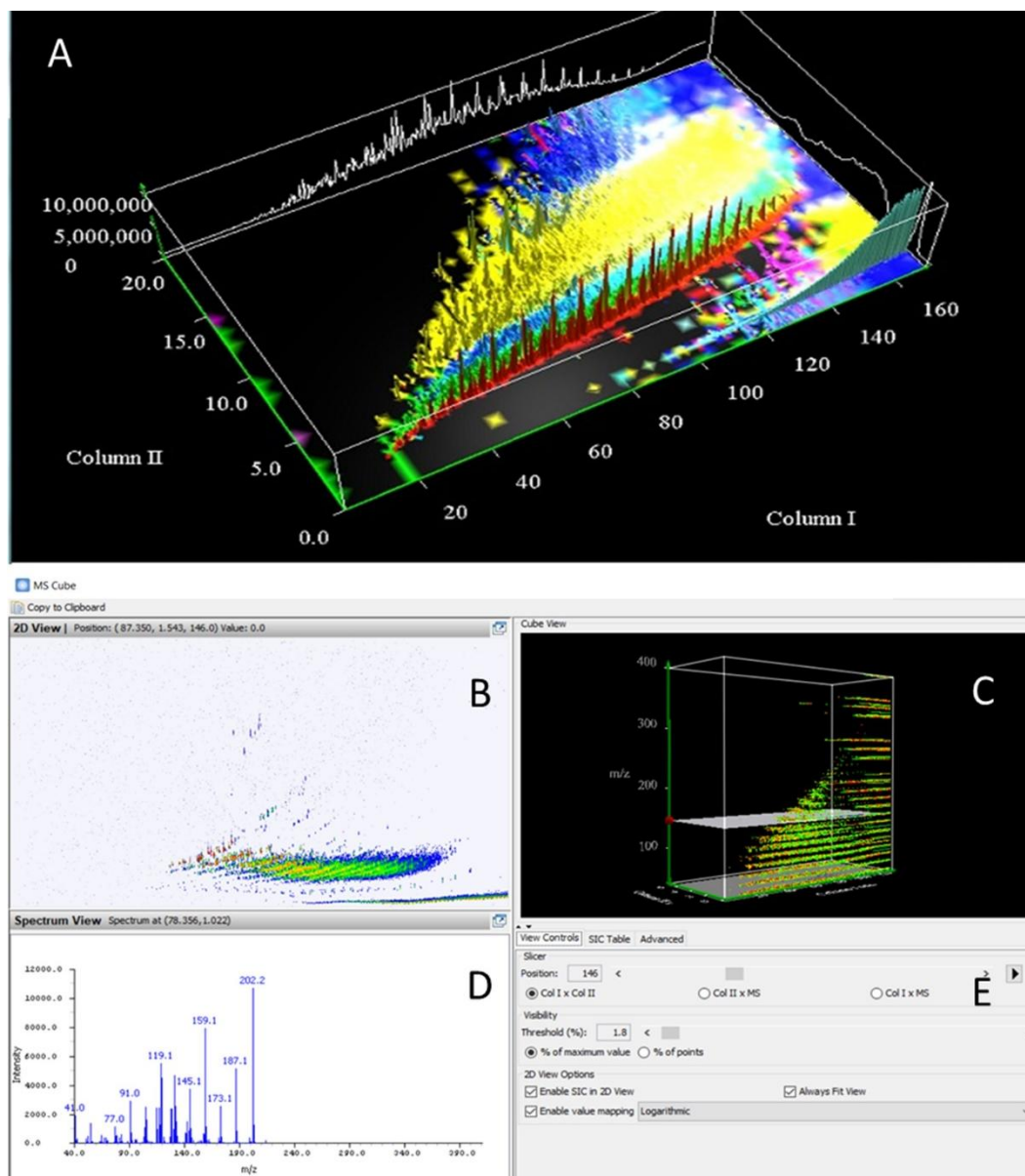


Figure 3: **3A** Spectral colorization of a diesel chromatogram highlighting different chemical classes; **3B**) contour plot of a diesel sample: XIC view at channel m/z 146; **3C**) data cube with spectral slice (light grey); **3D**) selected data point spectrum; **3E**) control panel. Reproduced and modified with permission from Elsevier [40].

Spectral colorization maps can be generated by selecting spectral channels of interest (e.g., selected m/z values or patterns) while setting user-specified colors to be displayed on a 3D image as shown in **Figure 3.3A**. Mass spectrometry cube options enable data visualization as a cube (t_R , t_R , spectral dimension) with the possibility of slicing to create customized 2D images. **Figure 3.3** also shows a 2D image based on an MS XIC at m/z 146 (**Fig. 3B**), a 3D cube

with spectral slice (**Fig. 3C**), the spectrum at a selected data point (**Fig. 3D**), and control panel tunable parameters (**Fig. 3E**). From HR MS data, mass defect (i.e., the difference between nominal and exact mass) and Kendrick mass defect plots support visualization and identification of homologous compounds differing only by some base units in high-resolution mass spectra [44].

Comparative visualization, one of the earliest approaches for cross-sample comparison, is supported in many commercial software platforms. The most advanced tools include automated chromatographic alignment through reference 2D peaks, accompanied by response intensity normalization and dedicated visualization rules to highlight compositional differences. Such an approach, very useful as a preliminary tool for cross-comparative analysis, is insufficient when dealing with complex patterns/samples, with large sample sets, and when quantitative differences have to be evaluated. On the other hand, being very intuitive and straightforward for pair-wise comparative analysis of 2D patterns, it is the basis of the conceptualization of computer vision tools based on GC×GC and comprehensive 2D LC (LC×LC) data.

3.4.2 Data processing: From raw-data to results

Chromatogram processing consists of several steps including data pre-processing (i.e., modulation-phase adjustment and baseline correction), 2D peak detection including deconvolution and unmixing, analyte identification, and chromatogram analysis (i.e., profiling and fingerprinting).

Pre-processing: Phase-shift and baseline correction

Modulation-phase adjustment is applied if the initial data point of each 2D chromatogram is not synchronized with the modulation cycle. The operation is a phase correction and its setting is user-defined based on specific needs. Two-dimensional peaks characterized by wrap-around could potentially elute in the 2D void time, with their virtual shift to the top of the chromatogram facilitating peak detection and further processing.

Baseline correction has the primary objective of smoothing the steady-state current baseline of the detector or eliminating/suppressing the signal of an unresolved complex mixture. Commercial software platforms implement different algorithms that can be reconducted into two main categories: (I) baseline estimation around single 2D peaks, and (II) baseline estimation across the entire chromatogram. For multi-channel detectors, the baseline can be estimated on each channel and should be performed at the same time as spectral centroiding. Further insights on these pre-processing steps are available in specialized literature [45].

Two-dimensional peak detection and analyte identification

In GC×GC data, 2D peak detection can be approached with traditional 1D algorithms by detecting peaks along 2D chromatograms and then aggregating those belonging to the same analyte. Alternatively, peak detection can be run simultaneously in both chromatographic dimensions. Image processing literature identifies the term blob as a cluster of pixels that are darker or brighter than their surroundings. For GC×GC data the terms blob or 2D peak are not interchangeable since a blob might be formed by several co-eluted

compounds while a 2D peak might be detected as several blobs due to background fluctuations (e.g., false minima) or chromatographic distortions (e.g. fronting or tailing). This phenomenon is referred to as over-segmentation.

Approaches based on 1D algorithms when applied to multi-channel detection data, perform spectral deconvolution to provide suitable specificity for 2D peak detection and spectral reliability for confident identification in cases of co-elution. The most common spectral deconvolution algorithm is AMDIS, developed by NIST [46], [47]. It is implemented in many software packages dedicated to GC×GC and LC×LC data. Alternative algorithms, developed for high-speed TOFMS and incorporated in proprietary software, perform deconvolution by identifying at least one target mass spectrum from a peak with minimal chemical noise. Once found, the residual peak spectra are deconvolved by a system of simultaneous equations [48].

When peak detection is performed by examining the signal in the 2D space, the drain algorithm provides the best performance in blob detection utilizing all relevant information deriving from both analytical dimensions [49]. In the case of under-segmentation, i.e., when multiple analyte peaks are detected as one blob, the process of unmixing is of help. In this case, it can be conducted with numerical methods if co-eluting peaks have symmetrical shapes by deconvolution of separable, bilinear single-valued data array or with tri-linear models for multichannel data. Better performances are achieved in GC×GC-MS by combining blob detection with dedicated 2D algorithms with the unmixing performed through spectral deconvolution. The 2D peak detection or blob detection is supported by a large set of tunable parameters that support consistent multi-sample cross-comparative analysis.

Once 2D peaks or blobs are detected, statistical features can be extracted for further processing. The most relevant one is the integration or sum of all peak intensity values since it is indicative of the relative amount of the analyte in the sample. Extractable quantitative values (volumes or sum of areas) can be defined by the analyst as a function of the detector adopted or 2D pattern complexity. In the case of MS detection, TIC or XIC intensities can be chosen, with the latter supporting accurate quantification even in the presence of co-eluting compounds.

Analyte identification can be based on retention times or on the combination of retention times and spectral features. The first approach is for single channel detectors where the relative position of analytes over the Cartesian plane can be compared to that of reference standards. In GC×GC, thanks to the additional dimension, this approach is more reliable and consistent than in 1D GC. Moreover, it can be corroborated by retention index (I^T) measurements (along the 1D) by adopting well-established models like Kovat's or Van den Dool and Kratz's [50], [51]. Commercial software supports the automatic calculation of I^T and combines this information in the spectral search. Mass spectrometry detection offers additional, yet specific information on analyte identity through EI fragmentation spectra. Various types of spectra can be extracted from 2D peaks or blobs (i.e., average spectra, peak spectra, pointwise spectra) and used for matching using MS databases. Of the tools that combine constraints on retention times positioning and spectral features, the smart template

approach is the most effective and flexible [52]. It is available in commercial software and allows consistent 2D peaks/blobs tracking and alignment across many chromatograms [10], [53].

Chromatogram analysis

Due to its enhanced chromatographic resolution and efficiency, GC×GC(-MS) has exceptional information power, enabling detailed profiling of samples while, at the same time, generating high-order data fingerprints for cross-comparative analysis (i.e., fingerprinting) [10]. The terms profiling and fingerprinting were coined within metabolomics, and literature in the field provides additional insights [54], [55]. In GC×GC(-MS), profiling can be conducted on analytes selected a priori from a list of known sample components. This approach, referred to as targeted profiling [56], is the most common yet simplest to reveal compositional differences between samples. However, since the analytical process generates individual yet distinctive features for all detectable components, profiling can be conceptually extended toward non-targeted components for a comprehensive mapping of all detectable analytes. In this case, it is referred to as untargeted/non-targeted or combined untargeted/targeted profiling [57].

Independently from the investigation approach, profiling or fingerprinting, 2D chromatograms can be indexed by different feature types before their cross-comparative analysis. Features can be data points, regions, peaks, and peak regions [58]. The most popular approaches are those based on peak features, where data for each analyte peak are summed and connected with associated metadata (MS spectra, retention times, I^T , chemical name, etc) before their alignment across many chromatograms. The process of matching peak features across multiple chromatograms requires target/non-target analyte peak detection in each chromatogram of the set. Then, the peaks for each target analyte are explicitly matched (targeted profiling) and non-target peaks are tracked based on the available metadata (untargeted profiling). Ambiguities may arise due to retention times and spectral inconsistencies with errors resulting in mismatched or unmatched peaks. To overcome such issues, region and peak-region features approaches were introduced.

Among region features, tile-based approaches have evolved over the years and are nowadays available in commercial software packages. They use rectangular tiles dimensionally (in times scale) congruent with chromatographic peak widths over imposed to the 2D chromatogram. Signal binning is then performed on each tile computing m/z values to locate 2D peak apexes and enable reliable re-alignment across many chromatograms. Further insights on this very effective approach are available in specialized literature [59], [60].

Template matching with peak region features offers an alternative to the tile-based region features approach. Peak regions are more specific than rectangular tiles, defining a 2D contour (i.e., region) around 2D peak/blob footprints on the retention time plane. In such a manner, peak regions achieve the one-feature-to-one-analyte selectivity, and they implicitly match the same peak region across chromatograms reducing the incidence of false negative matches. Co-elutions and retention time variability are circumvented by incorporating deconvolution, constraint functions based on MS spectral similarity, or by extracting

quantitative data from discriminant/selected ions [53]. Random and systematic retention-time variations are addressed by geometric transformations of the template in the retention-time plane [61], [62].

Recent overviews of the use of both basic and advanced software to do the previously described data mining approaches for GC×GC-MS chromatogram analysis should be examined for the latest developments. The primary goal of the data mining is to go from the raw data to useful chemical information. This process begins with a suitable experimental design so the appropriate samples are run by GC×GC, so the obtained raw data can be successfully analyzed to meet the analytical requirements. The various data mining methods can be categorized as targeted (simpler data mining) or non-targeted (more complex and often comprehensive in nature). In targeted methods, the analytes of interest are known a priori and can be quantified by traditional approaches or by chemometric deconvolution. In non-targeted analysis methods, the identification of class-distinguishing analytes is desired. If class information is known and used (e.g., “treatment” vs “control”) the method may be further defined as a supervised method, whereas methods that do not use class information are referred to as unsupervised methods. It is in the arena of non-targeted data analysis methods that has been a major focus in recent years.

3.5 Limitations and platform selection

Even though comprehensive 2D GC is the most powerful GC technique, yet, it is obviously a GC-based method and it suffers from all the intrinsic GC disadvantages. The molecules to be analysed need to be sufficiently volatile and thermally-stable, and thus sample preparation is needed to protect the vulnerable column system, and data processing is time-consuming. Such drawbacks are further exacerbated in GC×GC. Nowhere is this more evident than in handling the significantly increased volume of data. Whereas significant progress is expected in the data processing field, the other limitations are largely fundamental and unresolvable through further instrumental or software improvements. In terms of the application window, and considering CM GC×GC, the volatility range amenable to analysis is narrower since volatile molecules need to be trapped in the modulator rapidly and completely, while at the same time high boiling species need to be released efficiently. This reduces the upper end of the molecular weight in GC×GC to roughly C₆₀, versus approximately C₁₁₀ in regular GC. Variable modular settings over a run reduce the impact of such a drawback, and slight extensions have been demonstrated with carefully selected column sets. The presence of column connectors or flow switches in the modulator poses an increased risk for labile analytes. Differences exist between the various modulator designs, but coupling-free systems so far are not available. The impact of these drawbacks can significantly differ depending on the specific application. By carefully considering optimization strategies, selecting appropriate sample preparation methods, and implementing robust data processing techniques, one can effectively mitigate these limitations as is clearly demonstrated by the many successful applications.

Laboratory effectiveness and efficiency

The primary output of analytical laboratories is data to be used for, e.g., quality control, assessing regulatory compliance or to advance scientific knowledge. The data must be accurate and precise, without confounding errors arising from poor resolution, mixed mass spectra, or low signal-to-noise ratios. Comprehensive 2D GC increases the likelihood of meeting such analytical requirements and reduces the risk of providing false data. Additionally, it minimizes the chance of overlooking unexpected compounds. Hence, it significantly contributes to the analytical laboratory success in achieving the strategic goal of drawing accurate conclusions about sample specifications, individual health status, or, e.g., food product safety. When dealing with unknown or new samples, GC×GC should therefore be the method of choice.

While laboratory effectiveness is the first performance indicator of a laboratory, efficiency, i.e., speed and cost, is a clear second. When it comes to equipment and methods, simplicity is the key to enhanced instrument reliability and reduced cost of operation. The mere fact that a technique provides more information is irrelevant if that information is superfluous or possibly even distracting. The first GC×GC analysis of a new sample will immediately show whether this technique is truly needed, or that regular GC(-MS) would suffice. For the area of food analysis, it has been shown that the latter will often be the case. It is likely that future developments in mass spectrometry will further reduce the need for chromatographic resolution. The choice between GC×GC with lower-resolution MS versus GC with high-resolution MS remains unresolved. Eventually this will be determined by the value and cost of the information provided rather than the spectacularity of the chromatograms and data sets.

Instruments and parameter selection

Cryogenic modulation is the standard in GC×GC because of its superior resolution, peak capacity and analyte detectability; however, it is a technique with high cost in use and a significant environmental footprint. Consumption of liquid nitrogen can be very high, especially if highly volatile species such as food flavours are analysed. Flow modulation serves as an alternative method if one can tolerate the associated trade-offs in resolution and analyte detectability. More retentive stationary phases in the modulation capillary could provide efficient trapping at more favourable (low) temperatures but this should not cause slow re-injection nor increased column bleed. Ionic liquid stationary phases are potential options worth considering.

Peak capacity in GC×GC, to a first approximation, equals the product of the peak capacities in the two dimensions. Unfortunately, because the columns in the two dimensions usually have different inner diameters, albeit with the same gas flow (thermal modulators), they cannot be both in their ideal operational conditions at the same time. Either a somewhat higher flow is selected where the 1D is at its optimum, or a lower flow is applied where the 2D provides the maximum plate number. Careful inspection of the contour plot will indicate which dimension is most critical and conditions can be selected to optimize the separation. Care should be taken to avoid too low flows as this would yield poor separations, at long analysis times. Flow rate issues can be reduced by employing identical column diameters in both dimensions, albeit only at the expense of using sufficiently high modulation frequencies.

Carrier gas removal at the modulator would resolve the issue but this would raise questions regarding the reliability of quantitative data.

Orthogonality in GC×GC is often overestimated. The contour plots obtained contain peaks from bottom left to top right, suggesting full orthogonality. Retention in both dimensions is, however, strongly correlated, because largely determined by vapour pressures. To enhance orthogonality and increase peak capacity, the retention time difference between the two dimensions must be maximized. Such a factor can be achieved by using columns with the highest possible film thickness in the second dimension. An additional advantage of the increased film thickness is the improved column loadability. Limitations to such a choice are related to wrap around or by an excessive contribution of the stationary phase to 2D peak widths. The use of a separate ²D column oven (commercially available) is certainly advisable. Such an add-on greatly enhances the versatility of a GC×GC set-up and allows fine tuning of 2D retention times, without the need to replace the ²D column. True temperature-programming with rapid shifting gradients further aids in maximizing compound coverage and system versatility.

3.6 Conclusions and Perspectives

Comprehensive 2D GC-MS has been on the analytical scene since the end of the 1990s, hence it can be considered as well-established. Even so, its use at the academic, industrial, clinical, and regulatory agency levels, has been less than what was probably expected a quarter of a century ago (for reasons later discussed). In terms of instrumentation, there will be almost certainly less evolution on the GC×GC side, with respect to the MS one. In fact, there is currently a wide variety of GC columns available, in relation to dimensions, stationary-phase selectivity and thermal stabilities. There is also a vast range of highly efficient modulators (both cryogenic- and flow-based), which enable the coverage of the GC analyte range. Perhaps the TDM, proposed by Phillips, but never produced industrially, could represent an interesting alternative, of limited economical and operational cost. Mass spectrometry, on the other hand, will continue to evolve rapidly, as it has done over the last two decades. In fact, if one focuses on the GC×GC-MS literature since 1999, then the technological evolution mismatch becomes evident. For instance, the most common forms of cryogenic modulation were introduced over 20 years ago, and are basically unaltered (in truth, there has been more activity and progress in the field of flow modulation). On the other hand, TOFMS and QMS instrumentation (and not only) have evolved greatly.

Comprehensive 2D GC is without any doubt more powerful than its one-dimensional predecessor, offering clear advantages in terms of peak capacity, group-type separation and analyte detectability. However, for a technique to really thrive it needs unique applications. Comprehensive 2D GC has those, in particular in group-type separations in the petrochemical field. Its much higher analyte detectability also results in a unique position in, e.g., flavour and food chemistry. However, with continuous developments in the sensitivity and mass resolution of MS as outline above, the latter uniqueness is slowly fading. Regular GC is a fierce competitor to its descendent GC×GC. Properly optimised GC methods will often suffice. The

key aspects here are the efforts needed to optimize a one-dimensional GC method and the ability to unveil unexpected shortcomings. Comprehensive 2D GC chromatograms uncover issues such as overlapping peaks or on-column thermal degradation that might easily remain unnoticed in one-dimensional separations. Comprehensive 2D methods provide the highest chance of success if immediate answers are needed for new samples, or for situations where the risk of incorrect answers should be rigorously minimized.

Liquid chromatography is the other threat to GC methods, in general. Liquid chromatography systems are generally more reliable than GC systems. This holds both for first time use after a period of stand-still, as well as for use in daily routine. Despite many years of research in sample preparation methods and injection techniques, system contamination and column aging remain almost fundamental weak points in GC. This unreliability drives industrial preferences towards LC. Gas chromatography methods might indeed be more prone to contamination or thermal artifacts, but their speed, sensitivity and reliability of compound identification are unsurpassed. Moreover, recent trends such as the desire for green and sustainable analytical methods push laboratories back to GC. Comprehensive 2D GC is a powerful start in GC method development, but it is hardly ever the only feasible technique for an application and must earn its credits in a fair comparison to other GC and LC options.

The application of GC×GC necessitates computer-assisted and automated processing due to the vast datasets produced. Commercial software packages offer full automation capabilities, including artifact correction, two-dimensional peak detection, analyte identification, and quantification. However, for explorative analysis of a few select samples, semi-automated processing is advantageous. It allows for visual interpretation and interactive analysis, transforming raw GC×GC data into meaningful chemical information. This dual approach underscores the necessity for robust automated systems complemented by flexible semi-automated tools to meet diverse analytical needs.

Despite all current advancements, challenges remain in handling complex sample patterns, large sample sets acquired over large time frames. This necessitates the refinement and/or the development of existing yet new algorithms, categorized under Artificial Intelligence (AI) concepts and tools for profitable investigation of samples chemical dimensions and extend the existing knowledge.

Image pattern recognition and computer vision have advanced the interpretation of 3D/4D data arrays for deriving information and predicting sample properties. If compositional complexity challenges conventional data processing, the comprehensive 2D data's multidimensional nature offers unique opportunities.

Robust AI tools can compensate for retention-time shifts to facilitate comparative visualization, target unknown-knowns by exploiting mass spectral signatures amid co-elutions and misalignments, identify contamination sources in complex mixtures via the chromatographic fingerprints. This enables the analyst to focus on predicting aroma signatures by identifying key food odorants despite interferents (e.g., AI smelling [15]), or other emerging challenges in the analysis of fuels, feed stock evaluation (e.g., biomass to fuel), forensics and impurity profiling, environmental analysis, food quality, safety and security,

metabolomics, industrial discovery and quality control (e.g., biotech), and health biomarker discovery such as clinical studies.

The most powerful systems integrate multiple concepts and provide analysts with intuitive tools in commercial software platforms. Artificial Intelligence enhances the exploitation of comprehensive 2D data, and future advancements through generative learning will explore the full information content from diverse sources, linking it to additional, potentially unrelated properties.

References

- [1] A. T. JAMES and A. J. MARTIN, "Gas-liquid partition chromatography; the separation and micro-estimation of volatile fatty acids from formic acid to dodecanoic acid.," *Biochem. J.*, 1952.
- [2] G. De Grazia *et al.*, "Preparative three-dimensional GC and nuclear magnetic resonance for the isolation and identification of two sesquiterpene ethers from *Dictyota Dichotoma*," *J. Sep. Sci.*, 2023.
- [3] Z. Liu and J. B. Phillips, "Comprehensive Two-Dimensional Gas Chromatography using an On-Column Thermal Modulator Interface," *J. Chromatogr. Sci.*, vol. 29, no. 6, pp. 227–231, Jun. 1991.
- [4] Z. Liu, M. Zhang, and J. B. Phillips, "High-speed gas chromatographic analysis of a simulated process stream using on-column thermal desorption modulation for sample preconcentration and introduction," *J. Chromatogr. Sci.*, vol. 28, no. 11, pp. 567–571, 1990.
- [5] J. V. Seeley, F. Kramp, and C. J. Hicks, "Comprehensive Two-Dimensional Gas Chromatography via Differential Flow Modulation," *Anal. Chem.*, vol. 72, no. 18, pp. 4346–4352, Sep. 2000.
- [6] H. D. Bahaghighat, C. E. Freye, and R. E. Synovec, "Recent advances in modulator technology for comprehensive two dimensional gas chromatography," *TrAC - Trends in Analytical Chemistry*, vol. 113. Elsevier Ltd, pp. 379–391, 2019.
- [7] C. Cordero, C. Cagliero, E. Liberto, B. Sgorbini, P. Rubiolo, and C. Bicchi, "Chromatography: Focus on Multidimensional GC," in *Encyclopedia of Food and Health*, 2015, pp. 85–92.
- [8] C. Cordero, C. Bicchi, D. Joulain, and P. Rubiolo, "Identification, quantitation and method validation for the analysis of suspected allergens in fragrances by comprehensive two-dimensional gas chromatography coupled with quadrupole mass spectrometry and with flame ionization detection," *J. Chromatogr. A*, vol. 1150, no. 1–2, pp. 37–49, May 2007.
- [9] E. Belhassen *et al.*, "Routine quantification of 54 allergens in fragrances using comprehensive two-dimensional gas chromatography-quadrupole mass spectrometry with dual parallel secondary columns. Part I: Method development," *Flavour Fragr. J.*, vol. 33, no. 1, pp. 63–74, 2018.
- [10] F. Stilo, C. Bicchi, A. M. Jimenez-Carvelo, L. Cuadros-Rodriguez, S. E. Reichenbach, and C. Cordero, "Chromatographic fingerprinting by comprehensive two-dimensional chromatography: Fundamentals and tools," *TrAC Trends Anal. Chem.*, vol. 134, p. 116133, Jan. 2021.
- [11] M. K. Jennerwein, M. Eschner, T. Gröger, T. Wilharm, and R. Zimmermann, "Complete group-type quantification of petroleum middle distillates based on comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry (GCxGC-TOFMS) and visual basic scripting," *Energy and Fuels*, vol. 28, no. 9, pp. 5670–5681, 2014.
- [12] B. J. Pollo, G. L. Alexandrino, F. Augusto, and L. W. Hantao, "The impact of

comprehensive two-dimensional gas chromatography on oil & gas analysis: Recent advances and applications in petroleum industry,” *TrAC - Trends in Analytical Chemistry*. 2018.

- [13] C. Bicchi, C. Cordero, E. Liberto, B. Sgorbini, and P. Rubiolo, “Headspace sampling in flavor and fragrance field,” in *Comprehensive Sampling and Sample Preparation*, vol. 4, Elsevier Inc., 2012, pp. 1–25.
- [14] C. Cordero *et al.*, “High concentration capacity sample preparation techniques to improve the informative potential of two-dimensional comprehensive gas chromatography–mass spectrometry: Application to sensomics,” *J. Chromatogr. A*, vol. 1318, pp. 1–11, Nov. 2013.
- [15] S. Squara *et al.*, “Artificial Intelligence decision-making tools based on comprehensive two-dimensional gas chromatography data: the challenge of quantitative volatilomics in food quality assessment,” *J. Chromatogr. A*, vol. 1700, p. 464041, Jul. 2023.
- [16] N. Delbecque, S. Mascrez, E. Psillakis, and G. Purcaro, “Sub-ambient temperature sampling of fish volatiles using vacuum-assisted headspace solid phase microextraction: Theoretical considerations and proof of concept,” *Anal. Chim. Acta*, 2022.
- [17] L. M. Dubois *et al.*, “Thermal desorption comprehensive two-dimensional gas chromatography coupled to variable-energy electron ionization time-of-flight mass spectrometry for monitoring subtle changes in volatile organic compound profiles of human blood,” *J. Chromatogr. A*, vol. 1501, pp. 117–127, 2017.
- [18] C. Cordero, J. Kiefl, P. Schieberle, S. E. Reichenbach, and C. Bicchi, “Comprehensive two-dimensional gas chromatography and food sensory properties: potential and challenges,” *Anal. Bioanal. Chem.*, vol. 407, no. 1, pp. 169–191, Jan. 2015.
- [19] C. Cordero, J. Kiefl, S. E. Reichenbach, and C. Bicchi, “Characterization of odorant patterns by comprehensive two-dimensional gas chromatography: A challenge in omic studies,” *TrAC Trends Anal. Chem.*, vol. 113, pp. 364–378, Apr. 2019.
- [20] O. Beckonert *et al.*, “Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts,” *Nat. Protoc.*, vol. 2, no. 11, pp. 2692–2703, 2007.
- [21] G. Purcaro, P. Q. Tranchida, and L. Mondello, “Comprehensive Gas Chromatography Methodologies for the Analysis of Lipids,” in *Handbook of Advanced Chromatography/Mass Spectrometry Techniques*, Elsevier Inc., 2017, pp. 407–444.
- [22] G. Purcaro, S. Moret, and L. Conte, “Sample pre-fractionation of environmental and food samples using LC-GC multidimensional techniques,” *TrAC - Trends Anal. Chem.*, vol. 43, pp. 146–160, 2013.
- [23] M. Biedermann, C. Munoz, and K. Grob, “Epoxidation for the analysis of the mineral oil aromatic hydrocarbons in food. An update,” *J. Chromatogr. A*, vol. 1624, 2020.
- [24] D. Ryan, R. Shellie, P. Tranchida, A. Casilli, L. Mondello, and P. Marriott, “Analysis of roasted coffee bean volatiles by using comprehensive two-dimensional gas

chromatography-time-of-flight mass spectrometry," *J. Chromatogr. A*, vol. 1054, no. 1–2, pp. 57–65, 2004.

- [25] A. X. Zeng, S. T. Chin, and P. J. Marriott, "Integrated multidimensional and comprehensive 2D GC analysis of fatty acid methyl esters," *J. Sep. Sci.*, 2013.
- [26] C. Cordero, C. Bicchi, M. Galli, S. Galli, and P. Rubiolo, "Evaluation of different internal-diameter column combinations in comprehensive two-dimensional gas chromatography in flavour and fragrance analysis," *J. Sep. Sci.*, vol. 31, no. 19, pp. 3437–3450, Oct. 2008.
- [27] J. B. Phillips *et al.*, "A robust thermal modulator for comprehensive two-dimensional gas chromatography," *HRC J. High Resolut. Chromatogr.*, vol. 22, no. 1, pp. 3–10, 1999.
- [28] P. Marriott and R. Kinghorn, "Cryogenic solute manipulation in gas chromatography - The longitudinal modulation approach," *TrAC - Trends Anal. Chem.*, 1999.
- [29] J. Beens, M. Adahchour, R. J. J. Vreuls, K. Van Altena, and U. A. T. Brinkman, "Simple, non-moving modulation interface for comprehensive two-dimensional gas chromatography," *J. Chromatogr. A*, vol. 919, no. 1, pp. 127–132, 2001.
- [30] C. A. Bruckner, B. J. Prazen, and R. E. Synovec, "Comprehensive Two-Dimensional High-Speed Gas Chromatography with Chemometric Analysis," *Anal. Chem.*, vol. 70, no. 14, pp. 2796–2804, 1998.
- [31] J. Y. De Saint Laumer, E. Cicchetti, P. Merle, J. Egger, and A. Chaintreau, "Quantification in gas chromatography: Prediction of flame ionization detector response factors from combustion enthalpies and molecular structures," *Anal. Chem.*, vol. 82, no. 15, pp. 6457–6462, 2010.
- [32] F. Magagna *et al.*, "Advanced fingerprinting of high-quality cocoa: Challenges in transferring methods from thermal to differential-flow modulated comprehensive two dimensional gas chromatography," *J. Chromatogr. A*, vol. 1536, pp. 122–136, 2018.
- [33] C. Cordero *et al.*, "Method translation and full metadata transfer from thermal to differential flow modulated comprehensive two dimensional gas chromatography: Profiling of suspected fragrance allergens," *J. Chromatogr. A*, vol. 1480, pp. 70–82, 2017.
- [34] C. Cordero *et al.*, "Potential of the reversed-inject differential flow modulator for comprehensive two-dimensional gas chromatography in the quantitative profiling and fingerprinting of essential oils of different complexity," *J. Chromatogr. A*, vol. 1417, pp. 79–95, 2015.
- [35] T. M. Gröger, U. Käfer, and R. Zimmermann, "Gas chromatography in combination with fast high-resolution time-of-flight mass spectrometry: Technical overview and perspectives for data visualization," *TrAC Trends Anal. Chem.*, vol. 122, p. 115677, Jan. 2020.
- [36] I. C. Santos and K. A. Schug, "Recent advances and applications of gas chromatography vacuum ultraviolet spectroscopy," *Journal of Separation Science*. 2017.
- [37] J. Heyne, D. Bell, J. Feldhausen, Z. Yang, and R. Boehm, "Towards fuel composition and properties from Two-dimensional gas chromatography with flame ionization and vacuum ultraviolet spectroscopy," *Fuel*, vol. 312, p. 122709, Mar. 2022.

- [38] G. S. Frysinger and R. B. Gaines, "Comprehensive Two-Dimensional Gas Chromatography with Mass Spectrometric Detection (GC $\times$ GC/MS) Applied to the Analysis of Petroleum," *J. High Resolut. Chromatogr.*, vol. 22, no. 5, pp. 251–255, May 1999.
- [39] S. E. Reichenbach, "Chapter 4 Data Acquisition, Visualization, and Analysis," *Comprehensive Analytical Chemistry*, vol. 55, pp. 77–106, 2009.
- [40] S. E. Reichenbach, "GC $\times$ GC data visualization, processing, and analysis," in *Comprehensive Analytical Chemistry*, 1st ed., vol. 96, Elsevier B.V., 2022, pp. 185–229.
- [41] S. E. Reichenbach, V. Kottapalli, M. Ni, and A. Visvanathan, "Computer language for identifying chemicals with comprehensive two-dimensional gas chromatography and mass spectrometry," in *Journal of Chromatography A*, 2005, vol. 1071, no. 1–2, pp. 263–269.
- [42] B. A. Weggler, T. Gröger, and R. Zimmermann, "Advanced scripting for the automated profiling of two-dimensional gas chromatography-time-of-flight mass spectrometry data from combustion aerosol," *J. Chromatogr. A*, vol. 1364, pp. 241–248, 2014.
- [43] W. Welthagen, S. Mitschke, F. Muhlberger, and R. Zimmermann, "One-dimensional and comprehensive two-dimensional gas chromatography coupled to soft photo ionization time-of-flight mass spectrometry : A two- and three-dimensional separation approach," *J. Chromatogr. A*, vol. 1150, pp. 54–61, 2007.
- [44] M. Ubukata *et al.*, "Non-targeted analysis of electronics waste by comprehensive two-dimensional gas chromatography combined with high-resolution mass spectrometry: Using accurate mass information and mass defect analysis to explore the data," *J. Chromatogr. A*, vol. 1395, pp. 152–159, 2015.
- [45] S. R. Sternberg, "Grayscale morphology," *Comput. Vision, Graph. Image Process.*, vol. 35, no. 3, pp. 333–355, Sep. 1986.
- [46] K. Sichilongo and D. Banda, "GC-MS Determination of Targeted Pesticides in Environmental Samples from the Kafue Flats of Zambia," *Bull. Environ. Contam. Toxicol.*, vol. 91, no. 5, pp. 510–516, 2013.
- [47] N. I. of S. and Technology, *NIST/EPA/NIH Mass Spectral Library with Search Program: (Data Version: NIST 08, Software Version 2.0f)*, vol. 2010, no. 1/22. Gaithersburg MD: National Institute of Standards and Technology (NIST), 2005.
- [48] P. Livengood, R. Maciejewski, W. Chen, and D. S. Ebert, "OmicsVis: an interactive tool for visually analyzing metabolomics data.," *BMC Bioinformatics*, vol. 13 Suppl 8, no. Suppl 8, p. S6, 2012.
- [49] S. E. Reichenbach, M. Ni, V. Kottapalli, and A. Visvanathan, "Information technologies for comprehensive two-dimensional gas chromatography," *Chemom. Intell. Lab. Syst.*, vol. 71, no. 2, pp. 107–120, 2004.
- [50] H. van Den Dool and P. Dec. Kratz, "A generalization of the retention index system including linear temperature programmed gas—liquid partition chromatography," *J.*

Chromatogr. A, vol. 11, pp. 463–471, Jan. 1963.

- [51] E. Kováts, “Gas-chromatographische Charakterisierung organischer Verbindungen. Teil 1: Retentionsindices aliphatischer Halogenide, Alkohole, Aldehyde und Ketone,” *Helv. Chim. Acta*, vol. 41, no. 7, pp. 1915–1932, 1958.
- [52] S. E. Reichenbach, P. W. Carr, D. R. Stoll, and Q. Tao, “Smart Templates for peak pattern matching with comprehensive two-dimensional liquid chromatography,” *J. Chromatogr. A*, vol. 1216, no. 16, pp. 3458–3466, 2009.
- [53] S. E. Reichenbach, C. A. Zini, K. P. Nicolli, J. E. Welke, C. Cordero, and Q. Tao, “Benchmarking machine learning methods for comprehensive chemical fingerprinting and pattern recognition,” *J. Chromatogr. A*, vol. 1595, pp. 158–167, Jun. 2019.
- [54] O. Fiehn, “Extending the breadth of metabolite profiling by gas chromatography coupled to mass spectrometry,” *TrAC - Trends Anal. Chem.*, vol. 27, no. 3, pp. 261–269, 2008.
- [55] A. Scalbert *et al.*, “Mass-spectrometry-based metabolomics: Limitations and recommendations for future progress with particular focus on nutrition research,” *Metabolomics*, vol. 5, no. 4, pp. 435–458, 2009.
- [56] D. J. Beale *et al.*, “Review of recent developments in GC–MS approaches to metabolomics-based research,” *Metabolomics*, vol. 14, no. 11, p. 152, 2018.
- [57] F. Magagna *et al.*, “Combined untargeted and targeted fingerprinting with comprehensive two-dimensional chromatography for volatiles and ripening indicators in olive oil,” *Anal. Chim. Acta*, vol. 936, pp. 245–258, Sep. 2016.
- [58] S. E. Reichenbach, X. Tian, C. Cordero, and Q. Tao, “Features for non-targeted cross-sample analysis with comprehensive two-dimensional chromatography,” *J. Chromatogr. A*, vol. 1226, pp. 140–148, Feb. 2012.
- [59] P. E. Sudol, G. S. Ochoa, C. N. Cain, and R. E. Synovec, “Tile-based variance rank initiated-unsupervised sample indexing for comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry,” *Anal. Chim. Acta*, vol. 1209, no. December 2021, p. 339847, 2022.
- [60] T. J. Trinklein and R. E. Synovec, “Simulating comprehensive two-dimensional gas chromatography mass spectrometry data with realistic run-to-run shifting to evaluate the robustness of tile-based Fisher ratio analysis,” *J. Chromatogr. A*, vol. 1677, p. 463321, Aug. 2022.
- [61] S. E. Reichenbach *et al.*, “Alignment for Comprehensive Two-Dimensional Gas Chromatography with Dual Secondary Columns and Detectors,” *Anal. Chem.*, vol. 87, no. 19, pp. 10056–10063, 2015.
- [62] F. Stilo *et al.*, “Delineating unique and discriminant chemical traits in Brazilian and Italian extra-virgin olive oils by quantitative 2D-fingerprinting and pattern recognition algorithms,” *J. Food Compos. Anal.*, vol. 115, no. September 2022, p. 104899, Jan. 2023.

Chapter 4

Artificial Intelligence in *food-omic* applications

The integration of artificial intelligence (AI), particularly through machine learning and computer vision algorithms, has become essential for processing and interpreting complex data from multidimensional analytical techniques like advanced chromatographic separations. These techniques generate vast amounts of detailed information, which can be challenging to analyze using traditional methods. AI-driven tools allow researchers to efficiently manage and interpret this data, uncovering patterns and insights that might otherwise be missed. This chapter investigates the current uses of AI in food omics application.

The unceasing evolution of analytical instrumentation determines an exponential increase of data production, which in turn boosts new cutting-edge analytical challenges, requiring a progressive integration of artificial intelligence (AI) algorithms into the instrumental data treatment software. Machine learning, deep learning, and computer vision are the most common techniques adopted to exploit the information potential of advanced analytical chemistry measures. In this paper, our primary focus is on elucidating the remarkable advantages of leveraging AI tools for comprehensive two-dimensional gas chromatography data (pre)processing. We illustrate how AI techniques can efficiently explore the complex datasets derived from multidimensional platforms combining comprehensive two-dimensional separations with mass spectrometry in the challenging application area of food-omics. Pattern recognition based on image processing, computer vision, and AI smelling are discussed by introducing the principles of operation, reviewing available tools and software solutions, and illustrating their potentials and limitations through selected applications.

This chapter was published as Review Paper in Trends in Analytical Chemistry:

A. Caratti, S. Squara, C. Bicchi, E. Liberto, M. Vincenti, S.E. Reichenbach, Q. Tao, D. Geschwender, E. Alladio, C. Cordero, Boosting comprehensive two-dimensional chromatography with artificial intelligence: Application to food-omics, TrAC Trends Anal. Chem. 174 (2024) 117669. DOI: 10.1016/j.trac.2024.117669.

4.1 Introduction

Artificial intelligence (AI) is an exponentially expanding field poised to transform the landscape of analytical chemistry [1]. Within this expansive domain, various specialized branches have emerged to confront intricate challenges. One way to divide AI into branches is based on the type of learning involved [2]. Machine learning (ML) represents the cornerstone of AI, since it empowers computers to learn and predict from large amounts of data (Big Data) [3]. ML models and algorithms are trained on the data collected in the lab and then used to make predictions on novel samples. Another significant field of AI is deep learning (DL), an advanced facet of ML, which harnesses artificial neural networks (ANN) inspired by the complexity of the human brain, enabling tasks such as nuanced image recognition [4]. Other AI tools involve natural language processing (NLP), which facilitates seamless communication between computers and human languages, fuelling advancements in speech recognition and machine translation [3], computer vision (CV), which imparts machines with the ability to comprehend images and videos [5], and robotics, a pivotal arm of AI, is dedicated to the design and operation of robots, utilizing algorithms to handle tasks ranging from navigation to sophisticated manipulation [6]. These specialized branches collectively sculpt the intriguing terrain of AI, reshaping industries and heralding the future of technology. It is important to note that these branches of AI are not mutually exclusive. For example, many deep learning algorithms are also machine learning algorithms. Additionally, many AI applications use techniques from multiple branches of AI such as, for instance, open-source or commercial software [7], which might use DL algorithms to identify regions of interests (ROIs) contour images within the data acquired from comprehensive two-dimensional gas chromatography coupled to mass spectrometry (GC×GC-MS) [8], CV algorithms to evaluate the chromatographic fingerprint of the acquired samples images [9,10], and ML algorithms to develop supervised predictive models. All these strategies find applications in several research fields, such as environmental monitoring [11], exposome profiling [12], pharmaceutical research [13] and food quality [14–16], thus facilitating informed decision-making and accelerating the identification of innovative target compounds (*e.g.*, biomarkers).

Dividing AI into branches is useful for organising and understanding the different types of AI techniques. However, it is essential to remember that these branches are not mutually exclusive and that many AI applications use techniques from multiple branches of AI [17]. Undoubtedly, the most promising application of AI in analytical chemistry is ML, whose algorithms and models are exploited for data (pre)processing [18], unsupervised data analysis (explorative models) [19,20], supervised data analysis (classification [21] and regression models [22]) and design of experiment (DoE) [23,24]. AI algorithms serve as powerful tools, automating tasks, enhancing the precision of analytical methods, and pioneering novel techniques, especially in the *-omics* research [25]. Foodomics, in particular, is a complex field that analyses a wide range of food components, including primary metabolites, secondary/specialized metabolites, and other non-nutrient compounds (metabolomics) [26] to connect with biological outcomes in the nutrition and food science domains [27].

Multidimensional chromatography (MDC) techniques are powerful tools for separating and identifying these components, but it generates large and complex datasets whose interpretation can be challenging [26]. Therefore, AI algorithms are usually exploited to ease or possibly automate the analysis of these datasets and identify patterns and trends that would be substantially impossible to detect by manual inspection.

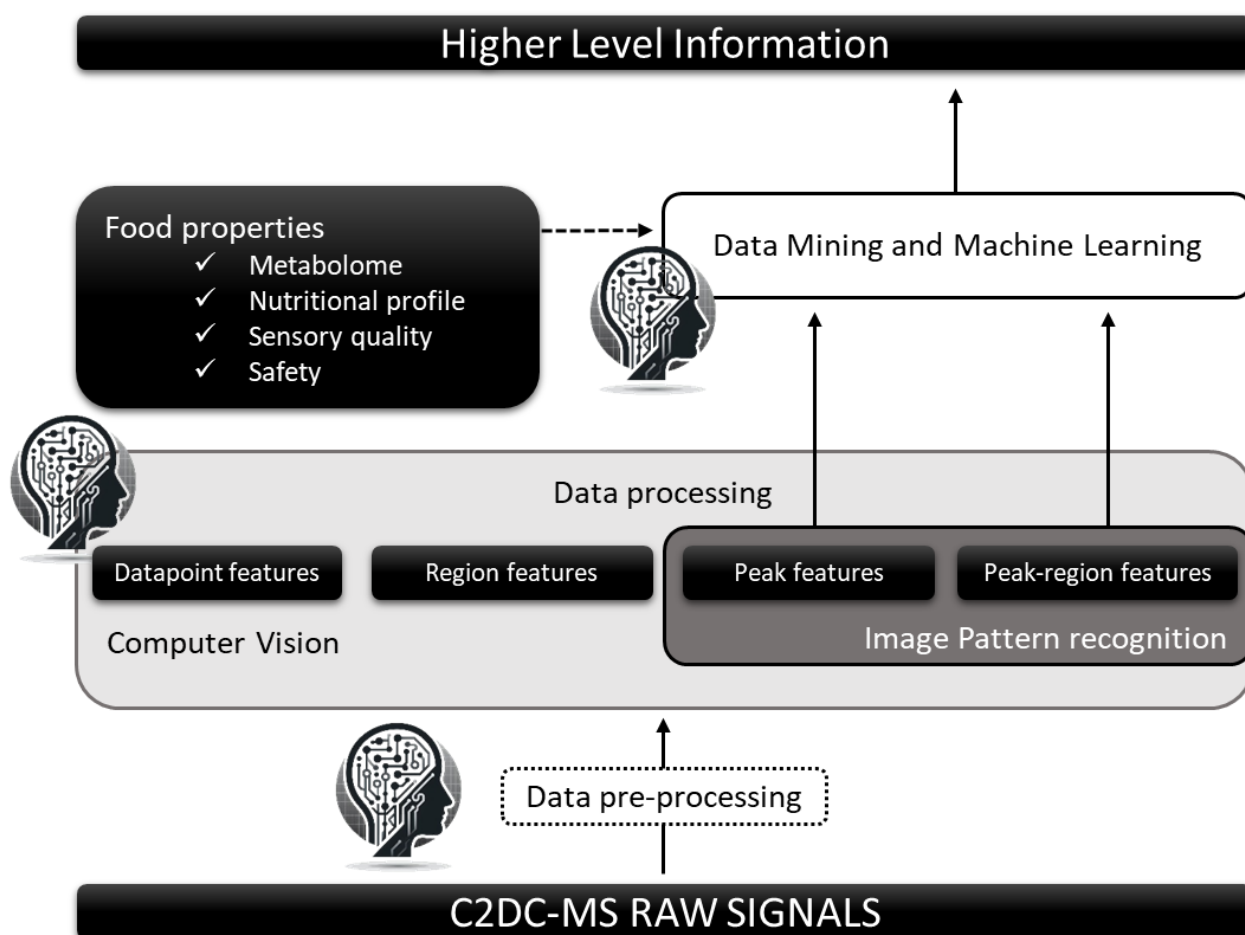


Figure 4.1. Logic workflow to access higher level information. Domains where AI tools are adopted: data mining and machine learning; data processing, data pre-processing. Further details are provided in the text.

In the domain of MDC, the targeted and untargeted approaches delineate methodological disparities, both designed with specific objectives and scientific applications. The targeted strategy is meticulously crafted to precisely examine predetermined compounds or compound classes within a given sample, necessitating calibration curves or standards tailored to the specific compounds under exam [28]. In contrast, the untargeted approach adopts a more expansive stance, enabling the comprehensive exploration of the entire chromatographic domain devoid of predefined targets [9,16]. This methodology involves the judicious selection of ROIs in both dimensions of the chromatogram, facilitating the discovery of unforeseen compounds or intricate patterns within the data [8,29]. ML techniques shine in this domain, leveraging their ability to recognize patterns and correlations within large datasets. ML algorithms, particularly those in unsupervised learning such as principal component analysis (PCA) [30,31] or t-distributed stochastic neighbour embedding (t-SNE)[32], can uncover previously unidentified biomarkers or unique biological patterns,

enhancing our understanding of complex biological systems. In contrast, while it is invaluable for confirming hypotheses and specific biological investigations, targeted approaches require fewer advanced computational methods due to their narrower focus, typically devoted to regression or classification tasks. Therefore, the selection between targeted and untargeted approaches, as well as the use of specific ML algorithms, is contingent upon the research objectives, wherein targeted methods offer precision and specificity, while untargeted approaches provide versatility and the opportunity for discovery novel target markers.

In this paper, our primary focus is not on ML *per se*. Instead, we focus on elucidating the remarkable advantages of leveraging AI tools for data (pre)processing. Our aim is to illustrate how AI techniques can efficiently explore the complex datasets derived from comprehensive two-dimensional chromatography (C2DC) in the application area of food-omic. By harnessing the power of AI, we highlight the potential for streamlined and insightful data analysis which includes data pre-processing, data processing based on different kind of features, and data mining as illustrated in the schematic diagram of **Figure 4.1**. Through this exploration, we showcase the transformative impact of AI in deciphering the complexities of GC×GC data, emphasizing efficiency and precision in data processing. The application of data mining and ML to collected peak features and peak-region features proves immensely useful in foodomics. This integration of techniques adds significant value to the analysis of complex datasets generated from advanced analytical technologies. One primary advantage lies in pattern recognition and classification, where these methods help identify distinctive peak features or patterns serving as biomarkers for specific food components or contaminants. This capability is particularly vital for quality control, authentication, and safety assessment within the food industry. Additionally, ML algorithms play a crucial role in feature selection, focusing on the most relevant peak features or regions and effectively reducing dimensionality. This not only streamlines the analysis process, but also enhances the efficiency of subsequent modeling efforts. Furthermore, ML techniques enable predictive modeling for quantitative analysis, allowing for the accurate prediction of the quantity of specific compounds present in food samples, contributing to a more in-depth understanding of food composition, quality, and safety through the extraction of meaningful insights from complex molecular datasets. Machine learning is seamlessly integrated into the field of data mining, playing a crucial role in various stages of the process. In terms of data pre-processing, machine learning algorithms contribute to handling missing data by predicting and filling values, and they assist in data cleaning through techniques like clustering and anomaly detection. Moreover, feature selection benefits from ML's statistical prowess, employing algorithms to identify the most relevant features for predictive models. Dimensionality reduction, often achieved through methods like principal component analysis (PCA) and unsupervised algorithms, utilizes machine learning to reduce complexity while preserving essential information. Furthermore, ML involves the application of classifiers and regressors to categorize and predict data in a supervised way. Therefore, ML's role extends to addressing the interpretability of models, with features integrated to enhance transparency in the data mining process. Consequently, the integration of ML into data mining amplifies the capacity to extract meaningful insights from vast and complex datasets

4.1.1 Nature's complexity and analytical system dimensions

As intriguing test bench for AI tools and concepts, food applications offer many challenges due to the compositional complexity of samples (e.g., primary metabolites, secondary/specialized metabolites, processing derivatives, exogenous compounds from microbial communities, presence of xenobiotics and contaminants etc.) and the properties connected to specific yet unique chemical patterns. Multidimensional analysis (MDA) systems, combining physico-chemical discrimination of individual components by chromatography with the structure-elucidation potential of MS, are the ultimate solutions for comprehensive investigations. However, the analytical data generated by MDA platforms have to be fully exploited by applying non-conventional approaches supported by the AI potential, to make a further step ahead and generate new knowledge on food properties, going beyond current quality indexes. Biological systems are characterized by a high degree of complexity due to multi-level interactions of molecular patterns on multiple targets [33]. To reveal and understand the higher-order network structures and relationships between conditions that generate biological phenomena, system biology has developed and validated investigation strategies and tools capable of tackling Nature's complexity while giving access to higher-level information [16,34]. The so-called *integrationist approach* of system biology necessitates MDA systems to perform untargeted analysis [33] while requiring dedicated yet effective data processing to fully exploit the information potential and generate new knowledge.

Food science can take great benefits by implementing such comprehensive investigations (i.e., *integrationism*) since the generation of new knowledge would overcome current protocols/markers used in quality and authenticity assessment [35,36], sometimes ineffective or unreliable, by defining new robust markers capable of predicting key properties of food (e.g., origin, sensory features, quality level, nutritional density, bioactivity etc) [9,37,38]. Nowadays, we can affirm that analytical platforms adopted in *omics* have reached the maturity and effectiveness foreseen by R. Wilson in 1991 “*A persistent research frontier is the analytical chemistry of the mixtures of chemical substances generated by living organisms. The problems of separation, molecular identification, and quantification of these mixtures are enormous. They are the ultimate molecular mishmash. The challenges they offer have demanded, and produced, many new concepts in chemical measurement science*” [39]. In the field of GC, the conceptualization and realization of the first system for GC×GC marked a turning point in the approach to volatiles and semi-volatile mixtures analysis, while also revolutionized the means to process and interpret the resulting data. In 1991 Liu and Phillips showed the first comprehensive 2D separation of a mixture by connecting in-series two capillary columns. The system, equipped with a modulator operating on a thermal principle (i.e., open tubular capillary modulator), adopted a first dimension (1D) polyethylene glycol (PEG) column [21 m × 0.25 mm id × 0.25 µm df] connected to a second dimension (2D) apolar column with dimethyl polysiloxane stationary phase [1 m × 0.10 mm id × 0.10 µm df] [40]. The analytical run lasted 2.5 min and the modulation period (P_M) was set at 2 s. The resulting chromatogram was displayed in a 2D time domain and authors could state that “*each substance forms a peak on the two-dimensional plane at a location determined by the interaction of the substance with the two different stationary phases*” paving the way toward

new concepts of chromatographic interpretation and investigation, *i.e.*, 2D chromatographic fingerprinting [34]. Moreover, the potential of the new technique was immediately evident, since the authors commented in their concluding remarks that “*potential applications for this method take advantage of its speed, its power to separate complex mixtures, or its more reliable identification of substances through two independent retention times*”.

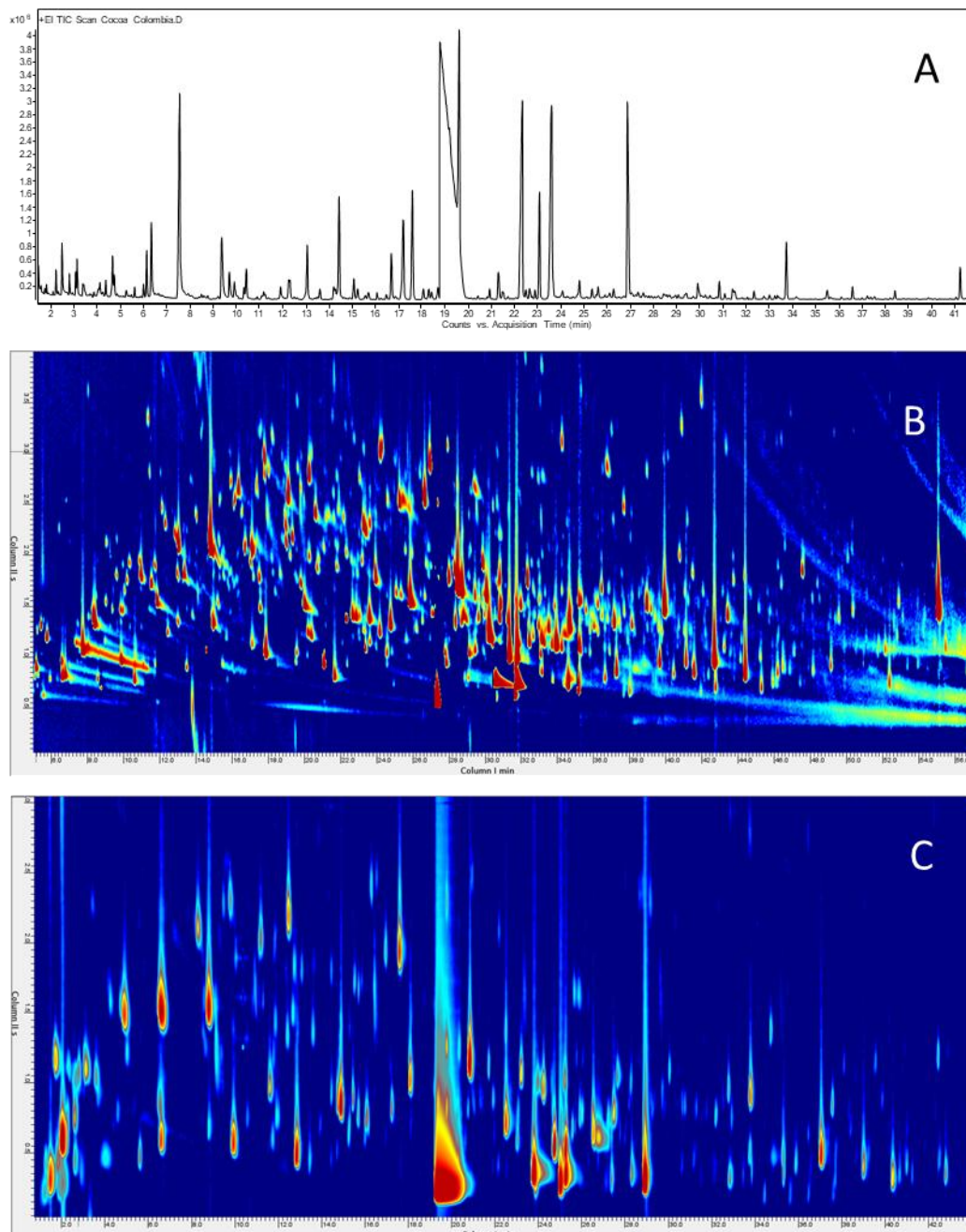


Figure 4.2. 1D-GC profile (**2A**) and GCxGC contour plots (**2B** and **2C**) of volatile compounds from roasted cocoa beans from Colombia origin. Sampling was by HS-SPME with a DVB/CAR/PDMS fiber on 1.50 g of finely grounded sample in 20 mL HS vial kept at 50°C for 40 min. Split/splitless injection was with a 1:20 split ratio in all analytical systems. Further details on the application and analytical set-up can be retrieved in reference study by Magagna *et al.* [41]. Detected peaks (resolution threshold 1.5) above a S/N of 50 were: 180 in the 1D-GC (**2A**); 1,130 in the GCxGC with loop-type thermal modulation (**2B**); and 290 in the GCxGC with reverse-inject differential-flow modulation with translated conditions (**2C**).

It took a while to overcome the skepticism of some experts, but then the technique was adopted in several fields as a complement to one-dimensional (1D)GC or by replacing it [42–51]. The improved separation capacity due to greater efficiency and resolution was accompanied by a significantly enhanced sensitivity, especially when platforms implement cryogenic modulation which produces band compression *in-space* [52]. These combined improvements made GC×GC the analytical tool of choice for investigating highly complex samples [53]. **Figure 4.2** compares the 1D-GC profile (2A) vs. GC×GC contour plots (2B and 2C) of the volatile fraction of a Colombian cocoa sample. Volatiles, sampled by headspace solid-phase micro extraction (HS-SPME) with a validated protocol [41], were analyzed in close-to optimal conditions by 1D-GC and GC×GC with loop-type thermal modulation and reverse-inject differential-flow modulation. The GC×GC methods were mutually translated to keep the original method resolution and preserve elution order [54]. The separation power and sensitivity can be roughly evaluated by the number of detectable peak features with a threshold resolution (*R*) of 1.5 above a signal-to-noise ratio (S/N) of 50. By 1D-GC a 180 chromatographic peaks were detected (2A); 1,130 in the GC×GC with loop-type thermal modulation (2B); and 290 in the GC×GC with reverse-inject differential-flow modulation with translated conditions (2C). To achieve suitable information capacity, MDA should nevertheless include MS detection, as it provides orthogonal information on the properties of the analytes. It supports structure elucidation [55], identity confirmation, and accurate quantification in presence of authentic or surrogated standards completing the information inferred by the differential retention in the C2DC system. MS, operating at unit-mass resolution or by exact mass assignments (*i.e.*, high-resolution MS - HRMS), allows access to fundamental molecular features and provides additional selectivity and specificity for data processing.

4.1.2 Food – omics domains and the role of comprehensive two-dimensional chromatography

The term "omics" describes the collective characterization and quantification of pools of biological molecules that translate into the structure, function, and dynamics of an organism or many organisms [56]. The first omics field was genomics, which was theoretically opposed to genetics and focused on examining individual variations or single genes rather than complete genomes. Numerous omics have been described to date, spanning a wide range of genetics, biology, and chemical domains. It's interesting to note that the majority of food-related omics attempt to link chemical patterns and composition with biological aspects of food, such as nutrient content and how it affects humans, sensory qualities, biological activity, and reactions to external stimuli, including climate change on crops, processing methods on semi-finished and finished goods, plant pathologies, and effects of bacterial and mold metabolism [56].

With a thorough and reliable definition of the chemical code of food, various phenomena are better understood, including (i) quality – by markers in raw materials and finished products, (ii) bioactivity - by patterns of nutrients and active ingredients, (iii) sensory perception - by taste and odor-active compounds that are receptor ligands, and (iv) safety - by contaminants and toxic compounds with concentrations exceeding predetermined limits. Within the general concept of *foodomics* [27], a discipline that investigates the interrelations

between food and nutrition domains by applying omics technologies, more specific omic-related disciplines have been conceptualized in the last decade [57]. They have common investigation approaches inasmuch as all looks for the interrelation between chemical patterns and different biological phenomena.

Food metabolomics, for example, studies the impact of external stimuli/variables on the primary metabolome that, for an edible crop, has relevance concerning nutritional density and sensory features (*e.g.*, the presence of taste-active primary metabolites or aroma precursors) [58–60] or might have diagnostic capacity in relation to genetic and phenotype traits. A good understanding of the effect on food composition of climate variations, harvesting practices, post-harvest treatments, storage conditions, and shelf-life, enables proactive strategies to drive quality toward desirable targets [28,37,61]. Food metabolomics has implemented GC×GC-MS as an essential technique, complementary to liquid chromatography (LC-MS) and nuclear magnetic resonance (NMR).

Food volatilomics focuses the investigation on the volatilome [62], also referred to as volatome [63], the fraction that includes volatile metabolites and all the other volatile organic and inorganic compounds originating from an organism [64] super-organism, or ecosystem. Volatile metabolites are part of the sample's metabolome, although degradation compounds and/or exogenous metabolites (*e.g.*, environmental contaminants, compounds formed by bacteria and molds - *microbial cloud* [65]) may also be present within this fraction. Better understanding of primary materials phenotyping and adaptation/reaction mechanisms is achieved by integrating the information brought by both volatilome and metabolome. Due to the physico-chemical properties of volatile metabolites, volatilomics is crucial in many food applications [50]. The easy access to the volatile fraction by gas sampling techniques and the availability of complete automation solutions for sample preparation steps have boosted the investigations in this area [37,66,67]. When nutrition sciences intercept the trajectory of food metabolomics, the nutrimetabolomics perimeter is delineated. Within its boundaries, the impact exerted by specific dietary patterns on human health is investigated [68]. By nutrimetabolomics, the impact of diet and food components/nutrients patterns on the human physiology [68] are deepened and better mechanistic understanding of the phenomena is likely achieved by providing nutritionists with effective predictive tools for early diagnosis or preventive strategies. Nutrimetabolomics focuses on both: (i) nutrients and essential nutrients influencing the metabolic response in humans, and (ii) non-nutrients having various bio-activities and impacting the human metabolism with various effects.

As an essential quality component, food hedonic properties are a challenging bench for analytical chemistry. Besides the actual potential of analytical platforms in terms of sensitivity and information capacity, human senses should be integrated into the analytical workflow. Olfaction, taste, and trigeminal perception should drive the investigation toward ligands (*i.e.*, sensory active compounds) screening and targeting, while validating the outcomes by recombination and omission experiments, as the final stage. Sensomics is the discipline that “*molecularizes flavor entities of nature*” by identifying the unique yet peculiar pattern of odorants and tastants that evoke food sensory identity [69]. Sensomics and flavoromics [70]

provide solid foundations for predicting food hedonic quality by rationalizing the chemoreceptive events occurring in our noses and oral cavity [71]. Food safety and authenticity are other crucial elements of food quality, although they have not yet been established as omics-related disciplines. Food safety is a prerequisite and a mandatory criterion for all food items intended for human consumption. Due to the current role played by GC×GC in safety evaluations of food (*e.g.*, mineral oil contamination assessment [72] and non-intentionally added substances detection [73]) and its information capacity for food *identification* [74], food safety and quality will be also covered by this review article.

4.2 Machine Learning techniques: general overview

Machine learning techniques can be broadly categorized into two primary groups: (i) pattern recognition techniques and (ii) regression/calibration models. In a nutshell, pattern recognition models further break down into two types: (i) unsupervised models (where a priori classification information for each object under examination is absent) and (ii) supervised/classification models (where the a priori classification of each object under examination is known). Well-known unsupervised methodologies, also referred to as exploratory analyses, include PCA [75], cluster analysis [76], t-distributed stochastic neighbor embedding (t-SNE) [77], and uniform manifold approximation and projection (UMAP) [78]. Conversely, supervised/classification modeling techniques can be subdivided into discrimination methods and class modeling techniques. Discrimination models, such as Fisher ratio [79], k-nearest neighbors (kNN) [80], logistic regression [81], naïve Bayes [82], linear discriminant analysis (LDA) [83], quadratic discriminant analysis (QDA) [84], and partial least squares – discriminant analysis (PLS-DA) [85], aim to distinguish objects within their respective classes by assessing features specific to each class. For instance, models like kNN evaluate distances between objects to estimate class boundaries, while LDA, QDA, and PLS-DA calculate new class boundaries in multidimensional space to separate objects within their classes. These models consistently yield outcomes related to object classification. Furthermore, they can also be categorized as probabilistic (parametric) methods because their classification algorithms rely on estimating parameters describing probability density functions, such as arithmetic mean, variance, and covariance, along with their respective distributions. Another set of supervised models for classification tasks relies on experience-based approaches. Specifically, these models employ iterative classification strategies, minimizing classification errors based on specific training sets used to construct the models. Traditionally labeled as ML techniques, examples include support vector machines (SVM) [4], random forest (RF) [86], eXtreme gradient boosting (XGBoost) [87], and ANN [88]. In contrast, class modeling techniques assess the similarity among objects within the same class. Consequently, different categories are modeled individually, allowing objects to be assigned to one, more than one, or even no category. Examples of class modeling techniques, also known as distance-based techniques, include soft-independent modelling of class analogy (SIMCA) [89] and unequal dispersed class (UNEQ) [90]. SIMCA and UNEQ classify objects based on different distance metrics, considering a measure of multivariate outlyingness of observations.

On the other hand, the primary goal of supervised regression/calibration modeling is to examine relationships between a matrix X of predictors/variables and one or more responses in a matrix Y. ML regression techniques include multivariate linear regression (MLR) [91], principal component regression (PCR) [92], and partial least squares regression (PLS-R) [93], but also most of the models mentioned previously for supervised classification can be converted to solve regression tasks (such as, for instance, kNN, SVM and RF). It's important to note that sometimes data cannot be structured into a matrix X but rather into an N-dimensional (N-way) array. This situation arises, for example, when considering information about sampling times or the opinions of different judges. A 3-way array, for instance, can be organized as follows: observations $\times$ parameters $\times$ sampling times. To explore such complex arrays, various N-way chemometric procedures can be applied, with parallel factor analysis (PARAFAC) [94] and Tucker3 [95] being the most popular. These models serve both classification and calibration purposes. A concise description of all the mentioned models is reported in the **Supplementary Material**.

4.2.1 Current applications of AI in correlating analytical data and food properties

Artificial intelligence algorithms (particularly ML and DL models) emerged as a powerful tools in food science, especially by correlating analytical data with various food properties (*i.e.*, data mining in **Figure 4.1**). In mass spectrometric and spectroscopic data, AI models can decipher intricate patterns associated with food composition, aiding in identifying specific compounds, additives, or contaminants [9]. By correlating this information with sensory attributes, nutritional content, or shelf life, researchers gain valuable insights into food products' quality, safety, and authenticity [96]. In particular, GC $\times$ GC-MS experienced rapid evolution under the boost of AI improvements. In data processing, the ROIMCR approach (regions of interest multivariate curve resolution), developed by the R. Tauler group [97], represents a sensitive and versatile technique to identify and quantify a wide range of metabolites in complex samples using comprehensive two-dimensional liquid chromatography (LC $\times$ LC) [98]. Specific ROIs are identified in the mass spectrum to highlight metabolites present at low levels, resolve overlapping peaks, discover new biomarkers and for the study of metabolic pathways [29].

In the discovery of new markers, multiple hypotheses testing is used to compare the levels of several metabolites between two sample groups and notice the significant differences, surpassing the limits of univariate testing. The Bayesian approach proposed by de Sousa *et al.* [99] explicitly incorporates the compositional constraint of metabolomic data, consisting in minor variations in numerous small-molecule metabolites within the investigated biological material, by representing them in terms of metabolite ratios. A new Bayesian parameter (b-value) was introduced, measuring the distance between a posterior distribution and the null distribution. The b-value can be used to rank the metabolites according to the strength of evidence against the null hypothesis (no significant differences) [100]. Also spectroscopic methods like Fourier transform infrared (FTIR) [101,102], Raman [103,104], and NMR [105,106], when used to investigate complex food compositions, develop

their true potential when coupled with ML and DL algorithms [107,108]. Once trained on big data, ML models distinguish subtle spectral patterns enabling rapid and accurate predictions of critical food properties, such as moisture content, fat composition, and protein levels [109], certifying origin and quality across diverse food products, from agriculture [110] to beverages [111].

Spectroscopic imaging technologies like hyperspectral imaging (HSI) provide detailed visual insights into foodomics [112,113]. In the food industry, AI-driven imaging technologies pinpoint defects [114], contaminants [115], and even predict fruit ripeness [116]. For instance, in fruit sorting processes, hyperspectral images are scrutinized by AI models to assess sugar content and peel features, facilitating precise sorting and quality control [117,118]. AI integration into real-time monitoring and quality control systems on food processing lines revolutionized the food industry [119], by prompt signalling deviations from quality standards, thus minimizing waste and assuring reliable answers/results.

4.3 AI and comprehensive two-dimensional chromatography data processing

Artificial Intelligence algorithms are currently used in many steps of an analytical workflow that have C2DC as core platform. Signal pre-processing is primarily concerned with individual chromatographic signals. Its focus is to refine raw data from the instrument, dealing with issues like modulation-phase adjustment [17,18], peak detection [120,121], noise removal [122], correction of baseline drifts [122–124] and peak distortions [24]. Therefore, these algorithms are expected to be applied directly to the signals of each sample, ensuring the accuracy of individual peaks. Several AI tools have been developed to perform and optimize signal pre-processing strategies, enhance the quality of individual signals before any data integration or analysis, and address the intricacies within each chromatographic trace, preparing them for subsequent processing steps [7,20,123,125]. One fundamental aspect of signal pre-processing is the precise identification of peaks within the raw chromatographic data, commonly known as peak detection, whose goal is to recognize those peaks that signify distinct compounds or analytes separated by the chromatographic system. Peak detection algorithms, employing techniques such as thresholding, derivative-based methods, and wavelet transforms, are instrumental to the accurate peaks location, forming the foundation for subsequent analyses [126,127]. In parallel, chromatographic signals often exhibit a baseline drift, a phenomenon induced by temperature fluctuations or detector noise: baseline correction algorithms come into play to eliminate or reduce it [123,128]. A clear and stable signal is obtained by removing this unwanted variation, providing a solid foundation for further in-depth analysis. Furthermore, random or systematic noise might arise from electronic or environmental sources, obscuring the genuine peaks within the data and hampering accurate analysis. Effective noise reduction techniques, including algorithms like smoothing, sparsity correction, and wavelet denoising are employed to this purpose [21,122,129]. These methods selectively diminish the noise levels while preserving the integrity of the actual signal, ensuring a more straightforward and accurate representation of the underlying data.

2D chromatography offers new opportunities for data exploration and interpretation, due to the high degree of multidimensionality of the data arrays collected during the analytical run. For the scope of this review, the potential of AI in the different data processing approaches is analyzed in detail. C2D chromatograms may be represented as digital images by rasterizing, *i.e.*, arranging the detector data values from individual modulation periods (or cycles) as a column of pixels (picture elements). It should be noted that when MS is adopted as the detector, each pixel corresponds to a spectrum, providing information on feature identity and amount, generating a 3D/4D data matrix [34]. Rasterized pixel columns are then arranged from left to right along the X-axis of a Cartesian plane according to 1D retention time, while the 2D data and related retention times are plotted from bottom to top on the Y-axis [130]. Chromatographic images can then be processed in a variety of ways, using multiple feature types and associated tools [34,130–133] as well as by applying image pattern recognition (PR) and CV tools [5,134,135]. Data arrays and fingerprints (*e.g.*, images, signals, textual data, or chemical features) can be effectively explored by PR to extract information about their chemical properties. Moreover, neural networks, ML or other AI algorithms can then support data classification into categories [136]. When applied to chromatographic data, PR drives sample discrimination and classification on the basis of the differential distribution of common chemical features or the presence of unique yet distinctive features [16,131,136]. Computer vision, a subfield of AI and computer science, extracts meaningful information from visual data (*e.g.*, digital images) and enables computer or systems to take actions or make recommendations [137]. CV applies techniques such as image processing, ML, and 3D image reconstruction [138]. Image PR and CV are, therefore, two strongly interconnected fields that share common algorithms and tools, and are possibly integrated into hybrid systems. As hybrid systems, CV and image PR offer a new perspective in the “observation” of the compositional differences between samples (*i.e.*, augmented visualization [139]) which is more intuitive and immediate than any other classical approach applied in chromatographic data analysis. Their integration in dedicated software and commercial platforms facilitates comparative analysis of large sets of samples while reducing the computational time even for high information density data such as those collected with HRMS.

4.3.1 Image pattern recognition and computer vision applied to C2DC data: approaches and tools

Image PR algorithms for 2D peak patterns matching in C2DC were implemented in commercial software many years ago [140,141]. Inspired by fingerprint recognition technology used in forensics, image PR based on template matching considers 2D chromatograms as the sample's unique fingerprint. Fingerprint verification systems localize and extract *minutiae* information (*i.e.*, ridge endings and bifurcations) from digitalized images of human fingertips and then perform cross-matching with stored templates for identification [142]. In C2DC, 2D peaks or other kinds of features (*e.g.*, datapoint or regions – **Figure 4.1**) are treated as *minutiae* features located over the 2D temporal space by tracking them across multiple images. When comparative analysis is aimed at samples discrimination or classification, it is referred to as chromatographic fingerprinting [34] which makes the process of sample *identitation* possible as previously conceptualized for 1D chromatographic fingerprinting [74]. Independently from

the feature type adopted, chromatographic fingerprinting could also provide visual evidence of compositional differences between samples, by comparative visualization [143]. In this perspective, it can be accounted a CV approach.

Computer vision by datapoint features

Comparative visualisation was one of the earliest approaches adopted to highlight compositional differences between pairs of chromatograms. As for other CV tools, it provides prompt evidence of pattern differences by comparing datapoint feature responses between pairs of chromatogram images. This pixel-by-pixel comparison could lead to erroneous conclusions when it is performed without any image/pattern re-alignment and transformation. Earlier applications based on datapoint features explored single-channel detector responses (*e.g.* flame ionization detector FID) with Fisher's *f* ratio and PCA [144]. Later, the same approach was extended to multi-channel detectors (*e.g.*, MS) by selecting specific *m/z* signals (mass-chromatograms) from the available spectra [31,145,146].

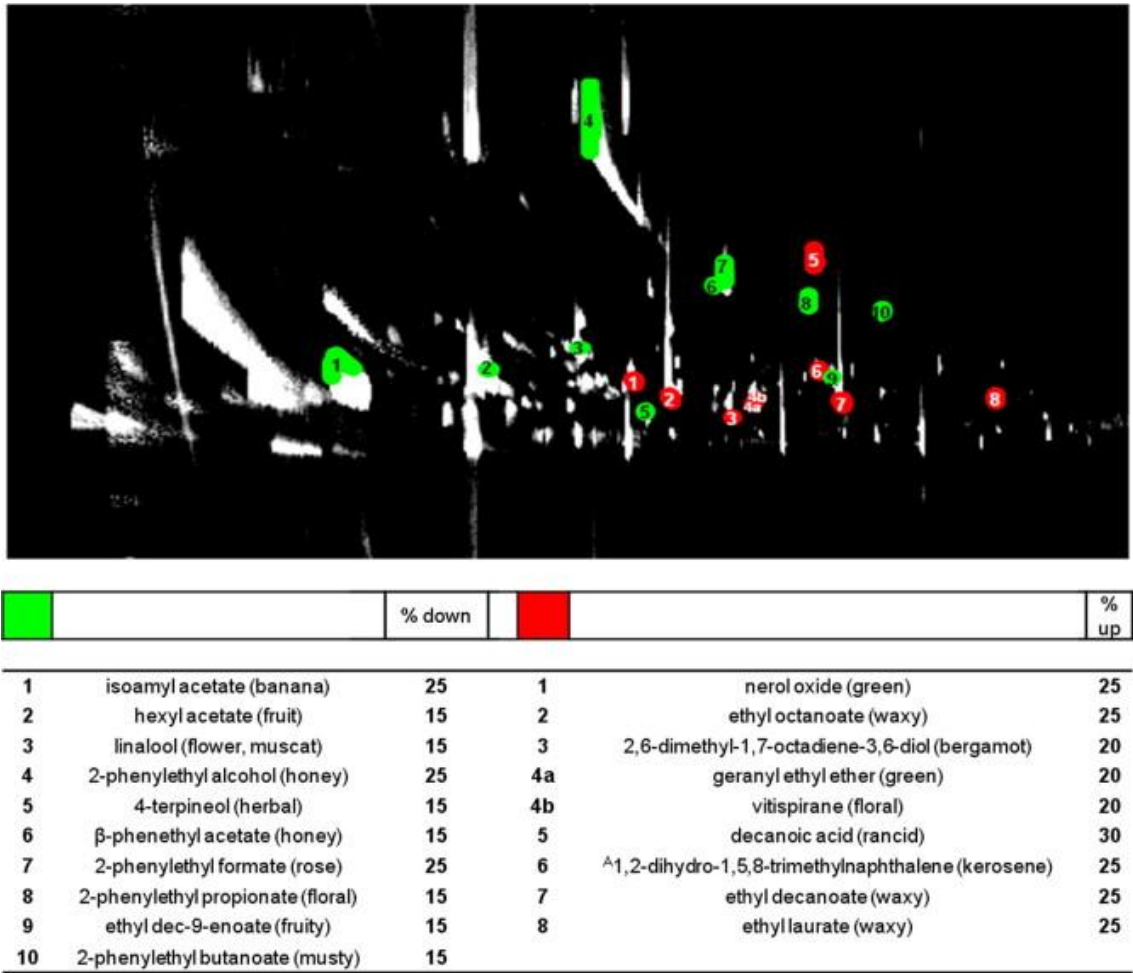


Figure 4.1. Computer Vision approach based on datapoint features. Comparative visualization among bidimensional chromatograms of Asti spumante wines (perfect one vs. imperfect one). Pixels in green correspond to analytes with decreasing trends while in red are represented those with increasing content. From Ref [135].

CV based on datapoint features without image re-alignment was adopted by Bordiga *et al.* [135] in a study focused on the evolution of the volatile fraction of “Asti Spumante” and “Moscato d’Asti” during aging. Based on Muscat grapes, both wines were screened by HS-

SPME to comprehensively map odorants and non-odorants across 6 months of storage at different temperatures (5, 15, and 25 °C). Objective of the research was the detection of variations in the distribution of potent odorants and the identification of haloanisoles which impart off-flavors to the wine. To enable effective comparative visualization of the 2D pattern differences, authors developed a Mathematica™ software (Wolfram Research Inc., Champaign, IL, USA) carrying out a multi-step process consisting of a first graphical reconstruction of the contour plot from raw data recorded by the GC instrument. The pixels size of resulting 2D-chromatogram images was 251 × 673. In the second step, chromatograms alignment on the time domain was checked by locating the internal standard reference pixels. The authors did not detect any “dimensional shifts” in the GC chromatograms, therefore the alignment was not applied. On the other hand, a logarithmic scaling on the intensity domain was applied to better appreciate the different pixel contrasts in the graphical representation. In the last step, re-scaling was performed by arbitrarily choosing a *reference* vs. a *test* sample, to produce comparable chromatogram pairs. The scaling based on a pixel-by-pixel correction intensity dependent generated a differential image illustrated in **Figure 4.3** for Asti spumante wines. By colorization, it was possible to highlight pattern differences between *reference* and *test* chromatograms, one of which corresponds to a wine classified as *perfect*, the other showing off-flavors and classified as *imperfect*. Differences are visualized in green for analytes with an increasing trend in the *perfect* class, and red for components with decreasing trends in wines with sensory defects (i.e., *imperfect*). As shown for the wine aroma application, datapoint features are comprehensive and provide the highest precision for fingerprint comparative visualization. However, many duplicative features per analyte might result in greater computational complexity while random shifts in retention times complicate consistent feature matching between pairs of chromatograms. These issues compromise the systematization of CV into decision-making processes; misinterpretation of compositional differences between sample pairs could produce wrong actions. Vial *et al.* proposed a method named *the discriminant pixel approach*, combining peak features with datapoint features discrimination for comparative analysis of a large set of samples [147]. The application on tobacco extracts, rich in volatiles and odor-active compounds, was a challenging test bench for their tool completed on Matlab™ and C platforms. It consisted of a pre-processing step with background correction and raw data intensity normalisation. Then, normalized signals were aligned on the temporal domain by dynamic time warping (DTW) applied with some constraints on the slope of the warping pipe and centering on the matrix diagonal using a windowing of 40 pixels. The temporal alignment was performed on each column of the matrix corresponding to each chromatogram as illustrated in **Figure 4.4**. For each class of the three tobacco types investigated (i.e., Burley, Virginia, and Oriental), one chromatogram was taken as a reference for the alignment. As the last step, the discriminant pixels were revealed by applying supervised PR models. Working in analogy to the Fisher Ratio method, first proposed by Pierce *et al.* [146], the authors were able to rank pixels corresponding to over-represented (or under-represented) compounds in each of the analyzed classes vs. all the others.

Very recently Ferreira *et al.* [10] proposed a comparative approach for the authentication of cachaça samples, a Brazilian distillate obtained from sugarcane, based on

the detectable volatilome. In particular, the approach included a pre-processing of the raw total ion current (TIC) signals for response normalization and retention times fluctuations compensation. The latter worked on the shift of the absolute retention times in both temporal axes and pattern transformation by affine linear functions available in commercial software (*i.e.*, GC Image [148]). After re-alignment, the chromatograms were converted into grayscale images and imported in Matlab™ for unfolding into vectors. Samples authentication, by brand and by adulteration, was achieved by applying data-driven soft independent modeling of class analogies (DD-SIMCA) with good performances. In particular, the model specificity achieved 98% for cachaça adulteration and 100% for branding distinction; in both cases, model sensitivity was 100% with no false negatives found (0.05 significance). A step ahead in making these hybrid systems (*i.e.*, combining image PR and CV) more specific was achieved by exploring the GC×GC-MS data array of regions and the corresponding peak region features. Higher specificity can be achieved by actively using the MS spectral signature during the comparative analysis or the CV. MS metadata can be exploited for discrimination - using the tile-based Fisher ratio analysis [149–151] - or effective re-alignment of the chromatographic patterns in the time domains, followed by features matching, as with the combined untargeted and targeted (*UT*) fingerprinting based on image PR algorithms [152–154].

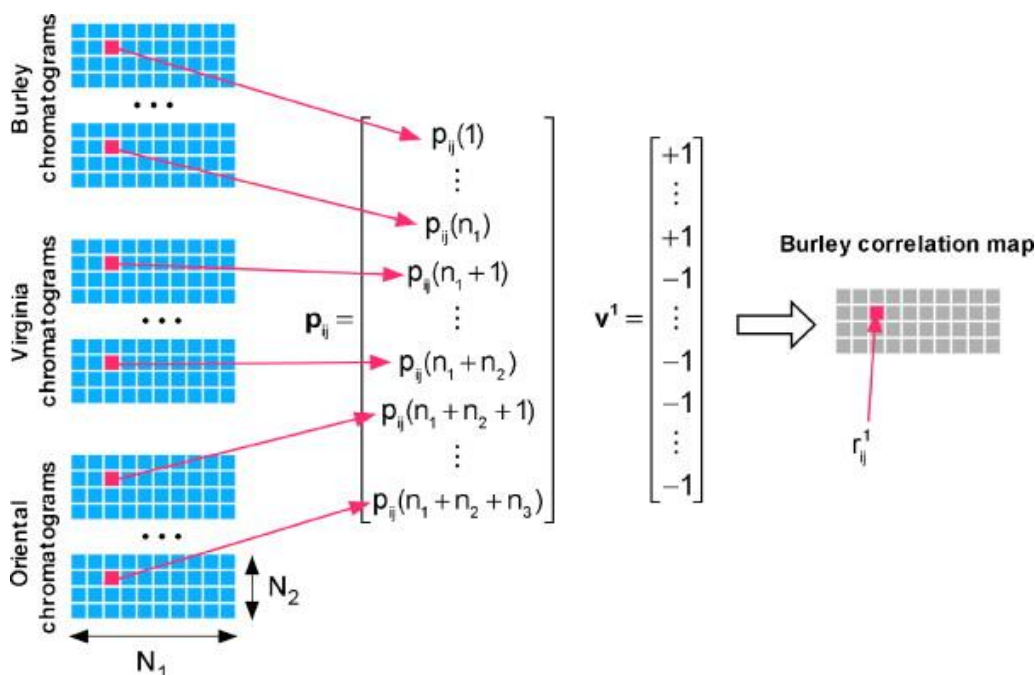


Figure 4.2. Example of the computation for one of the pixels of Burley tobacco correlation map. The ability of a pixel to discriminate between classes was defined as Pearson's linear correlation coefficient; if the correlation is close to 1, the pixel corresponds to a compound over-represented in Burley tobacco extracts, and if it is close to -1 , the pixel corresponds to a compound under-represented in Burley samples. Details are available in the Ref [147] where the original figure was reported.

Computer vision by region features

Regions are typically identified as small rectangular tiles overlaid on the contour plot of a 2D chromatogram. They were employed by Marney *et al.* [150] and Parsons *et al.* [149] to generate features for cross-sample analysis. Summing the m/z signal within a defined

number of datapoints on the first and second chromatographic dimensions into a single tile resulted in some advantages with respect to datapoint feature approaches, including reduced computational time and better compensation for temporal misalignments. These studies showed that PR operated with region features (*i.e.*, tiles) increased the rate of true positive matches and decreased the chance of false positives. Whenever MS constraints imposed no selectivity, the same approach showed limitations both in cases where tiles are populated by multiple analytes (*e.g.*, tiles larger than 2D peaks footprint) and when the analytes were split across multiple tiles. Tiles (*i.e.*, bins) are sized according to the chromatographic performances (*e.g.*, peak widths) while also considering possible random fluctuations of the retention times in the two dimensions. To improve the performance of the tile-based analysis, a four-grid tile scheme is used, as illustrated in **Figure 4.5**, with which each detected peak is likely to be included/captured within a tile rather than split into multiple tiles. The process, now implemented in the commercial software ChromaTOF™Tile (LECO, St. Joseph, MI, USA), overlaps a four-grid tile scheme over a chromatogram to enable signal binning on each tile on the basis of m/z values. The redundant hits generated by the four overlapped grid tiles are then removed by a “pinning and clustering” process, which is fully automated in the software. Once the “pin location” corresponding the 2D peak apex is determined, pins with similar retention in the 2D space are re-aligned and consolidated at the position of the pin with the highest intensity or response. The process is then repeated for all features and redundant hits are removed before peak centering. Details on tile-based analysis and its performances in case of signal saturation and co-elutions are provided in a recent review paper by Synovec and co-workers [136]. The tile-based approach was applied by Sudol *et al.* in a study focused on wine volatiles [155]. The geographical-based differences of “Grillo” wines from Sicily (Italy) were investigated by profiling the volatile organic compounds after HS-SPME sampling followed by differential-flow modulated GC×GC combined with time-of-flight MS (FM GC×GC-TOFMS). The analytes’ discriminant for the five wine classes were discovered by tile-based Fisher ratio analysis. For the specific application based on the actual 2D peak widths generated by the system, tile sizes were [7 s × 300 ms] corresponding to 10 modulations on the 1D and 45 spectra on the 2D, respectively. ANOVA test with a p -value threshold of 0.01 was applied to retain the most important class-distinguishing analytes, which were then used for unsupervised analysis with PCA. The PCA revealed the natural clustering of wine samples according to their geographical origin, supporting the adoption of tile-based analysis in challenging applications where the components span over a wide dynamic concentration range in the samples.

In food safety applications, contaminants consisting of complex fractions of chemicals are challenging for 1D chromatography-based methods. GC×GC with its capability of generating rational retention patterns for homologs and structurally related compounds, has great potential in groups/classes quantification and fingerprinting, as it is required for some contamination sources. Mineral oil hydrocarbon (MOH) contamination [73] is an example where CV based on region features was successfully applied. Grob and collaborators adopted for the first time on-line and off-line GC×GC coupled to LC to detect mineral oil-saturated hydrocarbons (MOSHs) and mineral oil-aromatic hydrocarbons (MOAHs) in food samples

[156]. By applying a medium polar $\times$ apolar column combination (OV-17 in the ^1D and PS-255-dimethylpolysiloxane in the ^2D) different classes of contaminants result separated within the 2D retention space with a retention logic driven by their volatility/polarity. Class distribution over the 2D space facilitates the identification of the contamination source [157]. Diagnostic patterns of contaminant mixtures were included in a contour (*i.e.*, region feature) which was stored as a template for comparative analysis. In this case, CV facilitated the identification of the source of contamination as long as it showed a unique yet distinctive fingerprint. Recently Ursol *et al.* tracked the MOH contamination in extra-virgin olive oil production by characterizing the MOSH/MOAH fractions fingerprint on process intermediates [158].

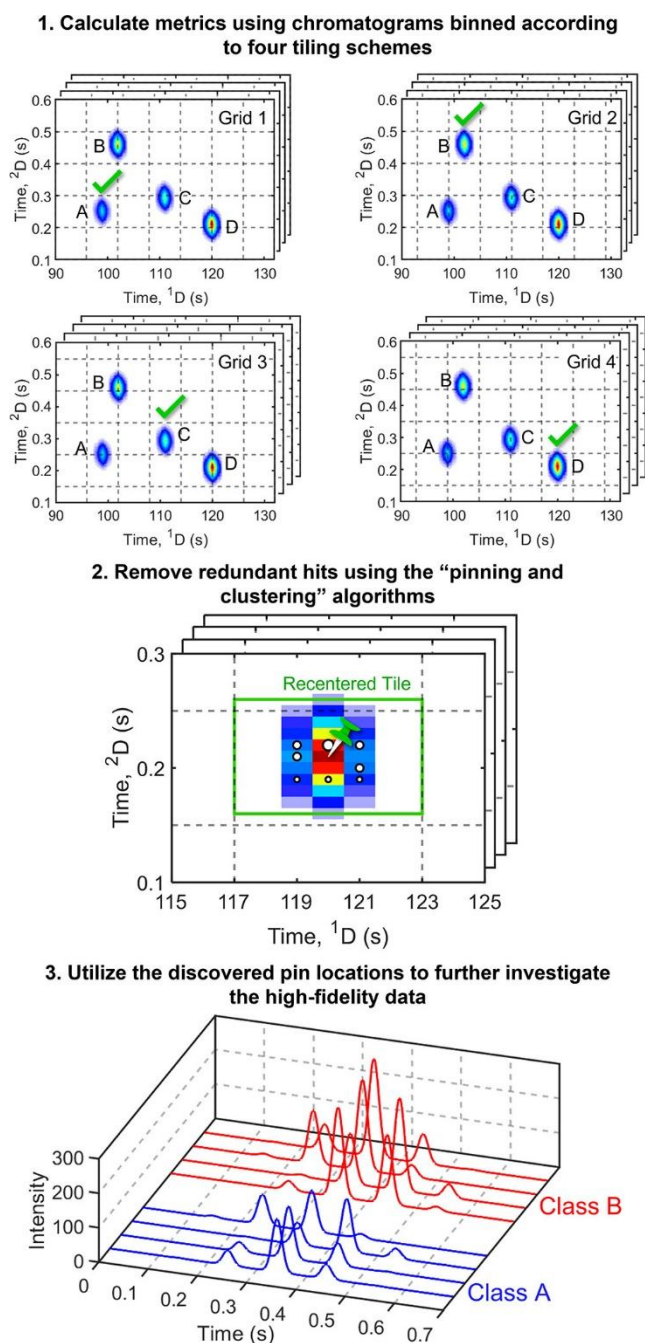


Figure 4.5: Computer Vision by tile-based analysis. Details are discussed in the text. From Ref [136].

Although effective from a computational viewpoint, region features are challenged by retention times inconsistencies and co-elutions. Moreover, work-flows occasionally do not provide visual evidence of pattern differences, possibly preventing prompt diagnosis of sample similarities and differences, as for most of the datapoint features methods.

Computer vision by peak-region features

Peak-region features based on template matching have been introduced in image PR to overcome some malfunctioning of peak feature approaches in scenarios characterized by pattern misalignment and detector response fluctuations [159]. Likewise, peak features

approaches, they achieve the one-feature-to-one-analyte selectivity, but also implicitly match the same peak region across chromatograms, a task that peak features do not necessarily accomplish in all patterns. Peak-region features provide a hybrid concept between regions and peaks by delineating one region in the 2D retention-times plane for each analyte peak detected in the data set. By precisely defining a 2D contour (*i.e.*, region) around 2D peaks footprint on the retention time plane, peak-regions are more specific than rectangular tiles and their re-alignment across chromatograms takes the benefit of the smart templates matching procedure [140].

Peak-region approaches were developed by Reichenbach *et al.* [159–161] and Schmarr *et al.* [162,163] with similar procedures. Raw data signal pre-processing (*i.e.*, rasterization, background noise subtraction, deconvolution, peak detection and integration) is the first step of an automated process applied to the source chromatograms of a samples set. After pre-processing, chromatograms are aligned on the temporal domain before their combination (fusion) to form a single composite chromatogram [164]. Then from 2D peaks detected in the composite chromatogram, a contour region is recorded and stored in a *feature template*. The feature template is then matched to all source chromatograms for geometrical remapping and transformation in the temporal domain. Once the process of feature template matching is completed for all chromatograms of a set, peak region features are tracked across all samples for a truly comprehensive mapping of the chemical diversity.

Since peak regions implicitly match all peaks in all chromatograms, they could be challenged by co-elutions and retention times variability. Whenever MS detection is available, co-eluting peaks falling within the same regions can be recognized by selecting quantitative/discriminant ions or resolved by incorporating deconvolution [153]. On the other hand, random or systematic retention-times variations are effectively addressed by geometric transformations in the retention-times plane [148,165,166]. Peak regions are implemented in a commercial software called “GC Image GC×GC and LC×LC” (GC Image, LLC, Lincoln NE, USA) by Smart Templates™.

Recently Caratti *et al.* proposed a novel approach referred to as *augmented visualization* to highlight pattern differences in butter samples analyzed along the production process [5]. The dataset consisted of 84 chromatograms (7 butter processing steps × 2 inocula × 2 biological replicates × 3 analytical replicates) plus additional quality control samples (QCs). Samples were grouped in seven classes as a function of the food processing stage: from raw sweet cream to cultured butter, then analyzed at four ripening stages (after 0-8-20-40 days). By automated UT fingerprinting available in GC Image Investigator™ (GC Image™, GC Image LLC), the chromatograms were re-aligned by feature template objects. This alignment was at first achieved within chromatograms belonging to the seven classes, whose workflow is illustrated in a schematic diagram depicted in **Figure 4.6**. The six chromatograms (2 biological replicates × 3 analytical replicates) were pre-processed before comprehensive pairwise peak matching in each class. Reliable registration peaks ($n \approx 1,000$ per class), peaks that were matched in at least 50% +1 of the chromatograms, were used as anchor points for template matching, image transformation, and registration. A composite chromatogram (Class Image) was then generated (last operation of Step 1 – **Figure 4.6**) by combining all registered images (*i.e.*, all analytical runs for samples belonging to the same class) in a single

image. The Class Images were then matched over the reliable template and graphic objects (*i.e.*, peak region features) generated and included in a *feature template* used later to match, transform, and align all Class Images chromatograms before their fusion into a cumulative Reference Class Image (right side of **Figure 4.6**) chromatogram. The process generates a feature template collecting all reliable peaks from Class Images and delineating the contour of all detected peak regions. A comprehensive alignment and registration of Class Images was possible; unique identifiers for peak and peak regions features were assigned enabling effective cross-comparison of features responses across many chromatograms (single chromatograms or Class Image chromatograms). By solving the temporal misalignment issue, the comparative visualization is facilitated and augmented visualization (AV) is achieved. AV refers to “*computational techniques for visualizing what cannot be seen with raw image input*” [139]. With the CV based on UT fingerprinting process, comparative visualization was directly linked to feature metadata and resulted *augmented* by molecular information that gives access to higher-level information about the phenomenon under study.

CV was used by Stilo *et al.* to highlight patterns of volatile organic compounds diagnostic of some spoilage phenomena in raw hazelnuts (*Corylus avellana* L.) [9]. Researchers analyzed samples selected among different cultivars, geographical origins, shelf-life and storage conditions and characterized by perceivable sensory defects (*mouldy, rancid, solvent-like, stale, and general unpleasant notes*). The workflow generated composite Class Images from samples grouped by sensory qualification of spoilage with an operative procedure similar to that proposed by Caratti *et al.* [5]. The results were validated against a chromatographic UT fingerprinting workflow based on image PR, providing proof of evidence of CV effectiveness and robustness in complex patterns analysis.

Classical CV techniques can be used to realign chromatograms with significant retention time shifts. Geschwender *et al.* demonstrated an automated two-step approach to alignment of chromatograms from samples of cocoa acquired on two different instruments, *i.e.*, a differential-flow modulated GC×GC-qMS (**Figure 4.7A** left chromatogram) and a cryogenic-modulated GC×GC-TOF MS (**Figure 4.7A** right chromatogram) without translation of the chromatographic conditions [41,167]. The first step, coarse alignment, aims to establish an initial match in scale and location as illustrated in **Figure 4.7B**. The enhanced correlation coefficient (ECC) is used, which is an image similarity metric that is robust to photometric distortions [168]. Multiscale image pyramids [169] are used in conjunction with the ECC to extend the range of ECC optimization search. The second step (**Figure 4.7C**), fine alignment, focuses on achieving peak-to-peak retention time correspondence. It initializes peak matching algorithms from the coarse alignment, and uses high-order global or local transform functions derived from matched peaks.

Examples provided here demonstrate how strategic is the adoption of AI tools at the data processing level and suggest that hybrid approaches which combine different AI concepts (*e.g.*, image PR and CV) could result also very intuitive and easily accessible for prompt identification of compositional differences that originate samples discrimination and classification.

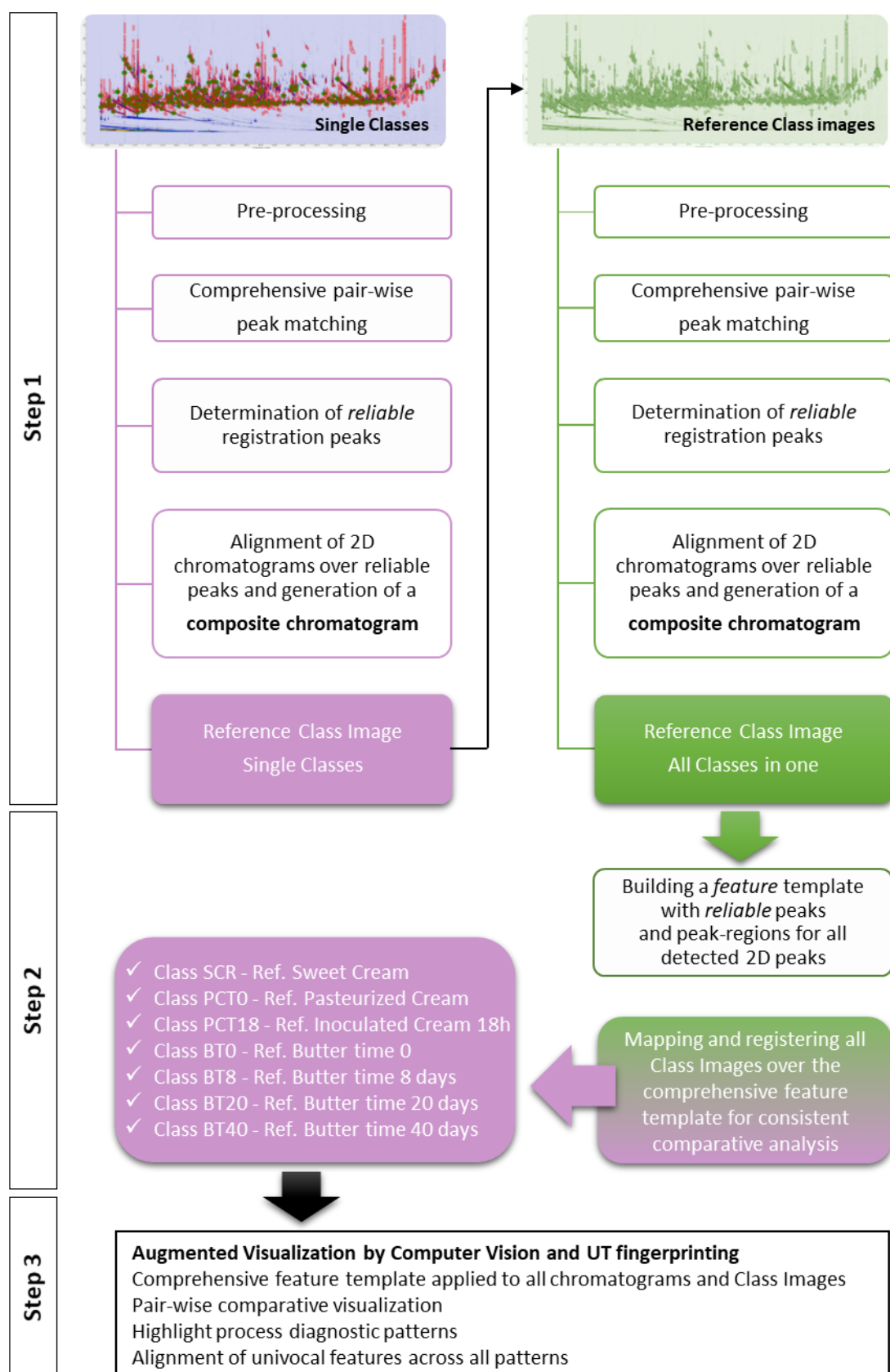


Figure 4-6: Schematic diagram of the computer vision approach based on image pattern recognition and images fusion. Details are discussed in the text. From Ref [5].

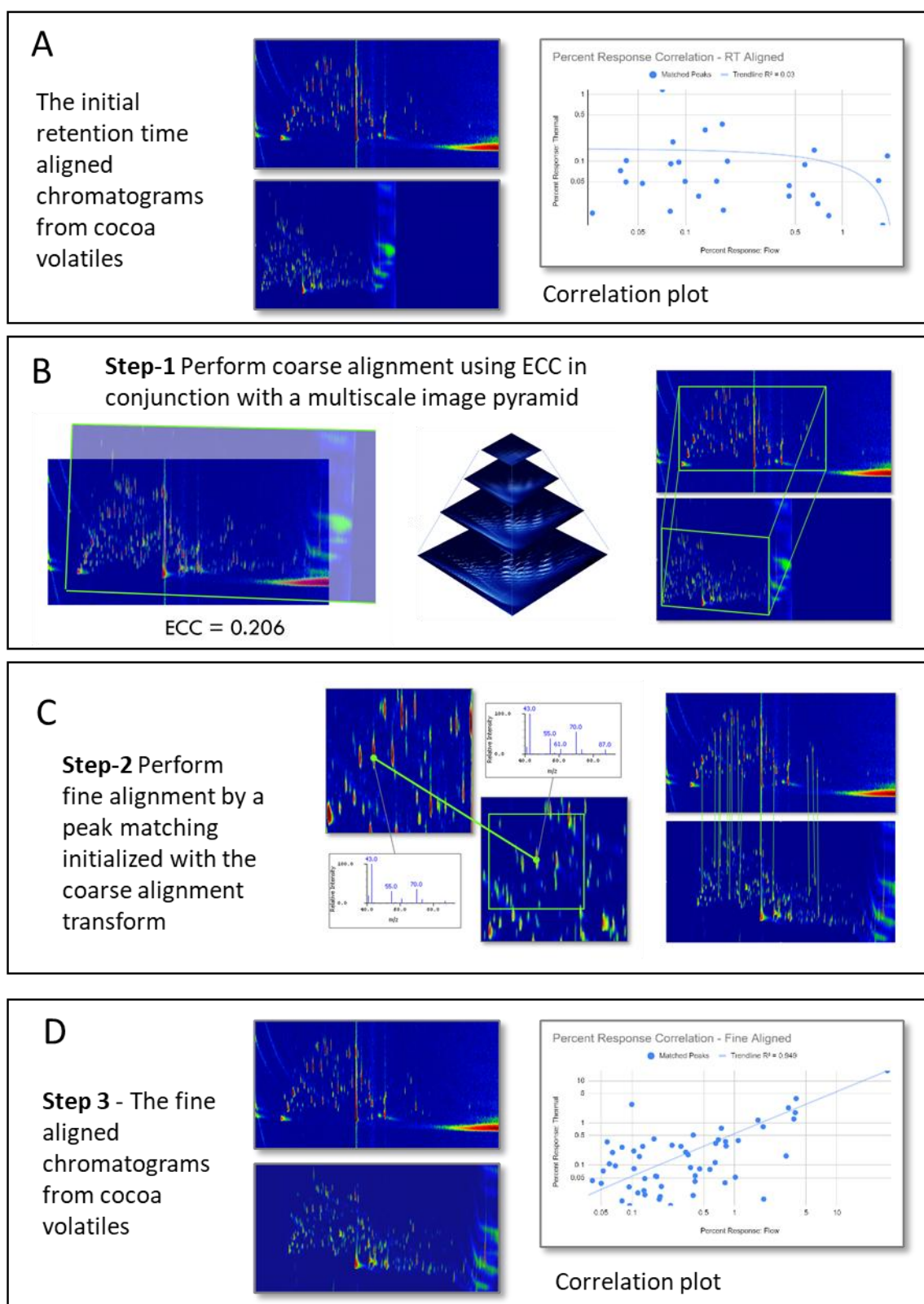


Figure 4.7: Schematic diagram illustrating the workflow for chromatogram images re-alignment in presence of severe retention-time shifts. Cocoa volatiles patterns were analyzed with a differential-flow modulated GCxGC-qMS (upper chromatogram 7A) and with a cryogenic-modulated GCxGC-TOF MS (lower chromatogram 7A).

Re-alignment is obtained by enhanced correlation coefficient (ECC) and multiscale image pyramids. Details are reported in the text. Details on cocoa volatiles profiling are reported in reference literature [41,148].

4.4 Artificial Intelligence Smelling based on sensomics: concept and tools

In this last section, the concept of *AI smelling* based on sensomics is discussed as it was conceptualized in an integrated workflow thanks to the unique features of GC×GC [71]. The food aroma alone contributes about 80% of the whole flavor [69]; it is generated by the interaction of odorants (*i.e.*, odor active volatiles) with odorant receptors (ORs) located in the olfactory epithelium. Potent odorants, even when present at trace and sub-trace level (mg/kg or µg/kg), activate a complex pattern of signals (*i.e.*, the Receptor Code) which integration by the nervous system produces olfactory perceptions [69]. The odor code, also referred to as *aroma blueprint* [69], evokes the unique identity traits of a food enabling its recognition and discrimination among others. To identify the molecular code of olfaction, sensomics has established a protocol that combines multiple, discrete steps aimed at isolating, extracting, and concentrating odor-active compounds before their qualification (odor characterization and description), accurate quantification and validation in aroma recombinates [170].

GC×GC became a central analytical tool for sensomics due to its separation power and enhanced sensitivity compared to 1D-GC or heart-cut 2D-GC. Marriott and co-workers designed and realized many *multi-multidimensional* platforms [171–173] implementing efficient separation of volatiles with olfactory screening by GC-olfactometry (GC-O). Various systems enable to switch between GC×GC and targeted MDGC [174] to operate in critical regions of the separation space with the maximum chromatographic resolution. Although attractive, the direct coupling of GC×GC with olfactometry (GC×GC-O) is quite challenging due to the very narrow bands produced by the 2D separation. D'Acampora *et al.* [175] explored the possibility of odor fingerprint acquisition by GC×GC-O/MS in the field of fragrance analysis. Their attempts remained pioneering as the complexity of execution is limiting. The sense of smell and neuronal integration of signals in humans requires a physiological recovery period that is not compatible with the speed and resolution typical of GC×GC.

On the other hand, off-line GC-O screenings open to some opportunities. They can drive or focus fingerprinting toward odor-active elution regions of the 2D chromatogram supporting ML to effectively discriminate samples based on features likely correlated to peculiar olfactory qualities. An example was provided by Gabetti *et al.* who investigated the unique aroma traits of Piemonte peppermint essential oils by combining GC-O, performed as aroma extract dilution analysis (AEDA), with GC×GC-TOF MS and variable ionization energy (Tandem Ionization™) [176]. By targeting specific regions of the chromatographic space with intense *creamy* and *sweet* notes, still perceivable at higher dilution factors, the classification performances of PLS-DA models improved. Authors argued that an AI-smelling tool based on GC×GC would be capable of making decisions driven by samples' sensory features without the use of a human sensory panel. Nicolotti *et al.* [71] developed a sensomics-based expert system – SEBES capable of predicting food aroma features by accurately quantifying key-aroma

compounds in a simplified methodological workflow. As a proof-of-concept supporting the SEBES effectiveness, the accurate quantification of about 100 compounds out of the 226 key-food odorants (KFOs) listed by Dunkel *et al.* [69] was achieved and their conversion in odor activity values (OAV; ratio of concentration to odor threshold) was demonstrated. GC×GC was fundamental to achieve an extended dynamic range of the method covering the actual concentration range of KFOs in real-world samples. SEBES validation was compared with reference sensomic protocol [50]; in particular, accuracy was very satisfactory with a maximum quantification error of $\pm 20\%$. As a conclusive remark, the authors stated that “*it was successfully shown that it is possible to characterize key food odorants with one single analytical platform and without using the human olfactory system, that is, by artificial intelligence smelling*”.

The AI smelling concept was further expanded to enable aroma blueprint prediction in different food products. For extra-virgin olive (EVO) oil, blueprinting was possible by exploiting the quantification features of the flame ionization detector (FID) with the application of predicted relative response factors (RRFs) based on combustion enthalpies [15,38,165,177]. Methods designed for accurate quantification of volatiles with multiple headspace extraction (MHE) combined with SPME could replace traditional extraction/distillation procedures speeding up the sample preparation step and making it fully automated [170,178]. The system realized with a differential-flow modulator with reverse-inject dynamics, and set for parallel detection by MS and FID, extends the quantification capabilities to all detectable volatiles showing HS linearity (i.e., without saturation of the HS) with benefits for high-throughput screenings as those supporting industrial research.

Recently, the AI smelling machine concept was integrated into a decision-making tool for confectionery industry [37]. The analytical method, completed by automated MHS-SPME combined with GC×GC-MS/FID, was applied to accurately quantify key-aroma compounds, rancidity and spoilage markers in premium quality hazelnuts (*Corylus avellana* L.) over 12 months of method application. By accurate quantification, chemical quality traits for incoming materials were confidently assessed with results transferability over time and across laboratories supporting objective decisions and reliable prediction of sensory features (i.e., AI smelling based on sensomics), geographical origin of samples, and storage time.

4.5 Concluding remarks

Artificial Intelligence concepts based on image PR and CV have been successfully applied to advance multidimensional data array interpretation to derive information and predict samples' properties. The compositional complexity of food challenges conventional data processing approaches, but the multidimensional nature of C2DC data offers unique possibilities for CV and augmented visualization. Robust AI tools are capable of (i) compensating for retention-time shifts that cause image distortion, thereby facilitating comparative visualization and related strategies; (ii) specifically targeting *unknown-knowns*, actively exploiting mass spectral signatures in the presence of co-elutions and misalignments; (iii) identifying the source(s) of contamination in complex mixtures by analysing

chromatographic fingerprints; and (iv) predicting aroma signatures by systematically identifying key food-odorants in presence of interferents.

The most powerful systems are those integrating multiple concepts and facilitating analyst access to embedded information using intuitive tools available in commercial software platforms. In-depth exploitation of the information content of every single analysis from C2DC benefits from AI and soon we expect further advancements through the adoption of generative learning which could explore the full information content from different sources and link it to additional, not necessarily correlated, properties [179].

Supplementary information

Machine Learning techniques: tools description

Principal Components Analysis (PCA): is used for transforming a set of correlated variables into a new set of uncorrelated variables known as principal components. These components are ordered by the amount of variance they capture, allowing for the identification of the most significant patterns in the data. PCA is a powerful tool for dimensionality reduction, noise elimination, and understanding the underlying structure of multivariate datasets. It achieves this by providing a concise representation of the data while retaining the essential information embedded in the original variables. The application of PCA extends beyond dimensionality reduction and encompasses feature extraction, signal processing, and data compression [180].

Cluster Analysis: is a family of techniques designed to group similar instances together based on their inherent similarities and distances. The objective is to uncover patterns and structures within the dataset, facilitating the identification of natural clusters or groups. Various algorithms, such as hierarchical clustering and k-means, are employed to partition the data based on distance metrics, similarity measures, and linkage functions. Cluster analysis is a valuable exploratory data analysis tool, aiding in pattern recognition, segmentation, and the identification of meaningful relationships within complex datasets [181].

t-Distributed Stochastic Neighbor Embedding (t-SNE): is a nonlinear dimensionality reduction technique designed to visualize high-dimensional data in a lower-dimensional space. Unlike linear methods such as PCA, t-SNE focuses on preserving pairwise similarities between instances, making it adept at revealing intricate local structures. By emphasizing relationships between neighboring points, t-SNE provides a powerful tool for capturing complex patterns that may be obscured in higher-dimensional spaces. t-SNE is particularly effective for data visualization, aiding researchers and analysts in gaining insights into the inherent structure of complex datasets [32].

Uniform Manifold Approximation and Projection (UMAP): is a dimensionality reduction technique that shares similarities with t-SNE but introduces several enhancements. UMAP aims to preserve both local and global structures in the data, offering a more consistent representation of distances in the lower-dimensional space. Known for its computational efficiency, UMAP is suitable for large datasets and provides a versatile approach to capturing complex relationships [182].

Fisher Ratio: is traditionally used to compare the variability between groups to the variability within groups. Its formula is calculated by dividing the mean square for between-group variability by the mean square for within-group variability. The Fisher ratio is particularly useful in feature selection and dimensionality reduction, where it helps identify the most relevant variables for distinguishing between groups. It is widely used in various fields such as pattern recognition, machine learning, and statistics for classification problems [183].

k-Nearest Neighbors (kNN): is a versatile and intuitive algorithm belonging to the family of instance-based learning methods. In kNN, the classification of a sample/instance is determined by the majority class among its k nearest neighbors in the feature space. The algorithm is non-parametric, meaning it does not make assumptions about the underlying data distribution. However, the choice of k influences the model's sensitivity to local patterns, and kNN can be sensitive to irrelevant or noisy features [184].

Logistic Regression: it is a powerful and widely used classification algorithm. It models the probability of an instance belonging to a specific class by applying the logistic function to a linear combination of input features. Logistic Regression is a linear model, making it computationally efficient and interpretable. It is well-suited for scenarios where the relationship between features and the outcome is approximately linear, and it is often employed in binary and multiclass classification tasks [185].

Naïve Bayes (NB): is a probabilistic classification algorithm based on Bayes' theorem. It assumes that features are conditionally independent given the class, a "naïve" but often effective assumption. Naïve Bayes calculates the probability of an instance belonging to each class and selects the class with the highest probability. This algorithm is particularly popular due to its efficiency, simplicity, and robust performance in high-dimensional datasets [186].

Linear Discriminant Analysis (LDA): is a method that combines dimensionality reduction and classification. Unlike PCA, LDA considers class labels during the dimensionality reduction process. It aims to find linear combinations of features that maximize the separation between classes while minimizing the variance within each class. LDA is especially powerful when there is a need to uncover discriminatory features in the context of classification tasks [186].

Quadratic Discriminant Analysis (QDA): is an extension of LDA that relaxes the assumption of equal covariance matrices across classes. QDA estimates separate covariance matrices for each class, allowing for greater flexibility in capturing the shape of class distributions. This added flexibility makes QDA particularly useful when dealing with datasets

where the within-class covariances vary significantly. QDA can provide more accurate results in situations where LDA's assumption of equal covariance may not hold [187].

Partial Least Squares – Discriminant Analysis (PLS-DA): is a multivariate statistical technique that combines partial least squares regression with discriminant analysis. PLS-DA is employed in situations where there are more predictors than observations, as well as when the features under exam show a severe collinearity, and the goal is to model the relationship between predictor variables and class labels. The method finds latent variables that maximize the covariance between predictor variables and class labels, effectively capturing relevant information for discrimination. PLS-DA is widely used in several fields of chemistry, especially where high-dimensional data necessitates effective dimensionality reduction and discrimination [188].

Support Vector Machines (SVMs): are powerful supervised learning models used for classification (and regression tasks, too). SVMs operate by finding the hyperplanes that best separates data points of different classes while maximizing the margin between them. The "support vectors," which are the instances closest to the decision boundary, determine the optimal hyperplane. SVMs are effective in high-dimensional spaces and are particularly useful when there is a need for clear separation between classes. They can also handle non-linear relationships through the use of kernel functions, enabling them to capture complex decision boundaries [189].

Random Forest (RF): is an ensemble learning method that builds multiple decision trees and combines their outputs to improve predictive accuracy and control overfitting, and it can be employed for both classification and regression tasks. Each decision tree is constructed by selecting a random subset of features and instances. The randomness introduced in the construction process helps diversify the individual trees, making the ensemble robust and less prone to overfitting. Random Forests are widely used for classification and regression tasks, and they excel in capturing complex relationships and handling large datasets [87].

eXtreme Gradient Boosting (XGBoost): is a boosting algorithm that has gained widespread popularity for its efficiency and predictive power, for both classification and regression tasks. It sequentially builds a series of weak learners (typically decision trees) and combines their outputs. XGBoost introduces regularization terms to control overfitting and uses gradient boosting techniques to minimize the loss function. Its strengths lie in handling missing data, feature selection, and providing high predictive accuracy [87].

Artificial Neural Networks (ANN): are computational models inspired by the structure and functioning of the human brain. Consisting of interconnected nodes organized into layers, ANNs can model complex relationships within data. The input layer receives features, hidden layers process information, and the output layer provides the network's final predictions. During training, the network adjusts the weights associated with connections through a process known as backpropagation, optimizing its ability to make accurate predictions. Deep Learning, a subset of machine learning involving deep neural networks with multiple hidden layers, has been particularly successful in tasks such as image and speech recognition, natural language processing, and autonomous decision-making systems. ANNs' ability to learn intricate hierarchical features makes them applicable to a wide range of domains [87].

Soft-Independent Modelling of Class Analogy (SIMCA): is a multivariate technique that turns particularly suited for scenarios where distinct classes in a dataset have similar overall characteristics but may vary in terms of variability within each class. The method constructs individual statistical models for each class based on its inherent variability. During the classification of an unknown sample, SIMCA evaluates the similarity of the sample to each class model. What sets SIMCA apart is its ability to handle overlapping class structures, making it useful in situations where classes may share common features, but their internal variations differ. SIMCA plays a crucial role in quality control (e.g., food and pharmaceutical analyses), and the classification of spectroscopic or chromatographic data where classifying samples into predefined groups is essential [190].

Unequal Dispersed Class (UNEQ): is another multivariate data analysis methodology specifically designed to address situations where the dispersion, or variability, within different classes in a dataset is unequal. In analytical chemistry, especially in spectroscopy, UNEQ becomes valuable when handling datasets characterized by imbalances in the spread of data points across various classes. Traditional machine learning methods may struggle when faced with imbalanced class structures, but UNEQ takes into account these differences in dispersion during the modeling process. By acknowledging and accommodating the unequal variability, UNEQ ensures that the resulting classification models are robust and accurate even in scenarios where certain classes exhibit higher variability than others. This is particularly crucial when dealing with datasets from diverse sources or experimental conditions, providing a reliable tool for classifying samples into distinct groups despite variations in class dispersions. UNEQ finds applications in various analytical fields, including spectroscopy, chromatography, and other high-dimensional data analysis tasks [191].

Multiple Linear Regression (MLR): is a statistical technique used to model the relationship between multiple independent variables and a dependent variable. In MLR, the model assumes a linear relationship, expressing the dependent variable as a weighted sum of the independent variables, with each weight representing the variable's contribution to the outcome. The coefficients are estimated through the method of least squares, minimizing the

sum of squared differences between predicted and actual values. MLR is a versatile tool that can handle complex relationships; it provides insights into how changes in multiple variables relate to changes in the response, allowing for a comprehensive understanding of the system under investigation.

Principal Component Regression (PCR): is a multivariate statistical technique that combines PCA with (multiple) linear regression. In PCR, the first step involves reducing the dimensionality of the independent variables by transforming them into a set of principal components. These components are linear combinations of the original variables, capturing the maximum variance in the data. Subsequently, these principal components are used as predictors in a (multiple) linear regression model. PCR is particularly useful when faced with multicollinearity issues, where independent variables are highly correlated. By using a reduced set of uncorrelated principal components, PCR provides a more stable and interpretable model while mitigating the impact of multicollinearity[191].

Partial Least Squares Regression (PLS-R): is a powerful regression technique that addresses challenges associated with multicollinearity and high-dimensional datasets. PLS-R, similar to PCR, decomposes both the independent and dependent variables into latent factors. However, unlike PCR, PLS-R aims to maximize the covariance between the predictors and the response variable, placing a greater emphasis on capturing the shared information that contributes to prediction. This makes PLS-R especially effective in situations where the number of predictors is larger than the number of observations. PLS-R is widely applied in several chemical fields, particularly when it is crucial to deeply understand complex relationships between variables is essential. It excels at handling noisy data, providing stable predictions, and offering insights into the underlying structure of the data[192].

Parallel Factor Analysis (PARAFAC): is a powerful technique for analyzing multi-dimensional arrays, commonly referred to as tensors. Unlike traditional factor analysis that focuses on matrices, PARAFAC extends the factorization concept to higher-order tensors. The primary objective of PARAFAC is to decompose a tensor into a sum of rank-one tensors, each corresponding to a unique mode or dimension. The decomposition involves extracting factor matrices for each mode, capturing the interactions and patterns within the data. PARAFAC is particularly advantageous when dealing with complex datasets exhibiting simultaneous variations along multiple dimensions. It is widely applied in fields such as chemometrics, where it enables the extraction of meaningful information from multi-dimensional spectroscopic or chromatographic data. The uniqueness of PARAFAC lies in its ability to unveil underlying structures in multi-dimensional datasets, providing a comprehensive representation of complex relationships[193].

Tucker3: or Tucker decomposition, it is a versatile technique for decomposing multi-dimensional arrays or tensors. Similar to PARAFAC, Tucker3 extends the principles of factorization to higher-order tensors. In the context of Tucker3, a tensor is decomposed into a core tensor and factor matrices for each mode or dimension. The core tensor captures the shared information among all modes, while the factor matrices represent the unique variability specific to each mode. One of the distinguishing features of Tucker3 is its flexibility in allowing different ranks for each mode, offering a more adaptable decomposition. This flexibility makes Tucker3 well-suited for scenarios where modes exhibit varying levels of complexity or importance. Applications of Tucker3 extend to diverse fields, including signal processing, image analysis, and chemometrics, where understanding multi-dimensional relationships is crucial. By retaining the essential features of multi-dimensional datasets, Tucker3 facilitates dimensionality reduction and aids in the extraction of meaningful patterns from intricate data structures[194].

References

- [1] M. Mousavizadegan, A. Firoozbakhtian, M. Hosseini, H. Ju, Machine learning in analytical chemistry: From synthesis of nanostructures to their applications in luminescence sensing, *TrAC - Trends Anal. Chem.* 167 (2023) 117216. doi:10.1016/j.trac.2023.117216.
- [2] L.B. Ayres, F.J.V. Gomez, J.R. Linton, M.F. Silva, C.D. Garcia, Taking the leap between analytical chemistry and artificial intelligence: A tutorial review, *Anal. Chim. Acta.* 1161 (2021) 338403. doi:10.1016/j.aca.2021.338403.
- [3] L.M. Petrick, N. Shomron, AI/ML-driven advances in untargeted metabolomics and exposomics for biomedical applications, *Cell Reports Phys. Sci.* 3 (2022) 100978. doi:10.1016/j.xcrp.2022.100978.
- [4] A.M. Jimenez-Carvelo, Data mining/machine learning methods in foodomics, *Curr. Opin. Food Sci.* 37 (2021) 76–82. doi:10.1016/j.cofs.2020.09.008.
- [5] A. Caratti, S. Squara, C. Bicchi, Q. Tao, D. Geschwender, S.E. Reichenbach, F. Ferrero, G. Borreani, C. Cordero, Augmented visualization by computer vision and chromatographic fingerprinting on comprehensive two-dimensional gas chromatographic patterns: Unraveling diagnostic signatures in food volatilome, *J. Chromatogr. A.* 1699 (2023) 464010. doi:10.1016/j.chroma.2023.464010.
- [6] R.C. Martins, S. Magalhães, P. Jorge, T. Barroso, F. Santos, Metbots: Metabolomics Robots for Precision Viticulture, in: 2019; pp. 156–166. doi:10.1007/978-3-030-30241-2_14.
- [7] M.J. Wilde, B. Zhao, R.L. Cordell, W. Ibrahim, A. Singapuri, N.J. Greening, C.E. Brightling, S. Siddiqui, P.S. Monks, R.C. Free, Automating and Extending Comprehensive Two-Dimensional Gas Chromatography Data Processing by Interfacing Open-Source and Commercial Software, *Anal. Chem.* 92 (2020) 13953–13960. doi:10.1021/acs.analchem.0c02844.
- [8] V.B. Mathema, K. Duangkumpha, K. Wanichthanarak, N. Jariyasopit, E. Dhakal, N. Sathirapongsasuti, C. Kitiyakara, Y. Sirivatanauksorn, S. Khoomrung, CRISP: A deep learning architecture for GC × GC-TOFMS contour ROI identification, simulation and analysis in imaging metabolomics, *Brief. Bioinform.* 23 (2022) 1–17. doi:10.1093/bib/bbab550.
- [9] F. Stilo, E. Liberto, N. Spigolon, G. Genova, G. Rosso, M. Fontana, S.E. Reichenbach, C. Bicchi, C. Cordero, An effective chromatographic fingerprinting workflow based on comprehensive two-dimensional gas chromatography – Mass spectrometry to establish volatiles patterns discriminative of spoiled hazelnuts (*Corylus avellana* L.), *Food Chem.* 340 (2021) 128135. doi:10.1016/j.foodchem.2020.128135.
- [10] V.H.C. Ferreira, L.W. Hantao, R.J. Poppi, Use of color based chromatographic images obtained from comprehensive two-dimensional gas chromatography in authentication analyses, *Talanta.* 234 (2021) 122616. doi:10.1016/j.talanta.2021.122616.
- [11] J. Gros, C.M. Reddy, C. Aeppli, R.K. Nelson, C.A. Carmichael, J.S. Arey, Resolving biodegradation patterns of persistent saturated hydrocarbons in weathered oil samples

from the Deepwater Horizon disaster, *Environ. Sci. Technol.* 48 (2014) 1628–1637. doi:10.1021/es4042836.

- [12] C.S. Blois, O. Fiehn, Using untargeted metabolomics for detecting exposome compounds, *Curr. Opin. Toxicol.* 8 (2018) 87–92. doi:10.1016/j.cotox.2018.03.002.
- [13] A. Zaid, N.H. Hassan, P.J. Marriott, Y.F. Wong, Comprehensive Two-Dimensional Gas Chromatography as a Bioanalytical Platform for Drug Discovery and Analysis, *Pharmaceutics* 15 (2023). doi:10.3390/pharmaceutics15041121.
- [14] S. Squara, F. Stilo, M. Cialie Rosso, E. Liberto, C. Bicchi, C.E.I. Cordero, Exploring food volatilome by advanced chromatographic fingerprinting based on comprehensive two-dimensional gas chromatographic patterns, in: *Compr. Anal. Chem.*, 2022. doi:10.1016/bs.coac.2021.11.008.
- [15] F. Stilo, E. Alladio, S. Squara, C. Bicchi, M. Vincenti, S.E. Reichenbach, C. Cordero, H.R. Bizzo, Delineating unique and discriminant chemical traits in Brazilian and Italian extra-virgin olive oils by quantitative 2D-fingerprinting and pattern recognition algorithms, *J. Food Compos. Anal.* 115 (2023) 104899. doi:10.1016/j.jfca.2022.104899.
- [16] F. Stilo, C. Bicchi, A. Robbat, S.E. Reichenbach, C. Cordero, Untargeted approaches in food-omics: The potential of comprehensive two-dimensional gas chromatography/mass spectrometry, *TrAC Trends Anal. Chem.* 135 (2021) 116162. doi:10.1016/j.trac.2020.116162.
- [17] N.B.L. Milani, E. van Gilst, B.W.J. Pirok, P.J. Schoenmakers, Comprehensive two-dimensional gas chromatography— A discussion on recent innovations, *J. Sep. Sci.* (2023). doi:10.1002/jssc.202300304.
- [18] S.E. Reichenbach, GC×GC data visualization, processing, and analysis, in: *Compr. Anal. Chem.*, 1st ed., Elsevier B.V., 2022: pp. 185–229. doi:10.1016/bs.coac.2021.10.006.
- [19] P.E. Sudol, G.S. Ochoa, C.N. Cain, R.E. Synovec, Tile-based variance rank initiated-unsupervised sample indexing for comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry, *Anal. Chim. Acta.* 1209 (2022) 339847. doi:10.1016/j.aca.2022.339847.
- [20] C. Quiroz-Moreno, M.F. Furlan, J.R. Belinato, F. Augusto, G.L. Alexandrino, N.G.S. Mogollón, RGCxGC toolbox: An R-package for data processing in comprehensive two-dimensional gas chromatography-mass spectrometry, *Microchem. J.* 156 (2020). doi:10.1016/j.microc.2020.104830.
- [21] T.S. Bos, W.C. Knol, S.R.A. Molenaar, L.E. Niezen, P.J. Schoenmakers, G.W. Somsen, B.W.J. Pirok, Recent applications of chemometrics in one- and two-dimensional chromatography, *J. Sep. Sci.* 43 (2020) 1678–1727. doi:10.1002/jssc.202000011.
- [22] J.S. Arey, A. Martin Aparicio, E. Vaiopoulou, S. Forbes, D. Lyon, Modeling the GCxGC Elution Patterns of a Hydrocarbon Structure Library To Innovate Environmental Risk Assessments of Petroleum Substances, *Environ. Sci. Technol.* 56 (2022) 17913–17923. doi:10.1021/acs.est.2c06922.
- [23] N.G. Mahieu, J.L. Genenbacher, G.J. Patti, A roadmap for the XCMS family of software solutions in metabolomics, *Curr. Opin. Chem. Biol.* 30 (2016) 87–93.

doi:10.1016/j.cbpa.2015.11.009.

- [24] P. Zhang, M. Piergiovanni, P. Franceschi, F. Mattivi, U. Vrhovsek, S. Carlin, Application of Comprehensive 2D Gas Chromatography Coupled with Mass Spectrometry in Beer and Wine VOC Analysis, *Analytica*. 4 (2023) 347–373. doi:10.3390/analytica4030026.
- [25] B.J. Pollo, C.A. Teixeira, J.R. Belinato, M.F. Furlan, I.C. de M. Cunha, C.R. Vaz, G.V. Volpato, F. Augusto, Chemometrics, Comprehensive Two-Dimensional gas chromatography and “omics” sciences: Basic tools and recent applications, *TrAC Trends Anal. Chem.* 134 (2021) 116111. doi:https://doi.org/10.1016/j.trac.2020.116111.
- [26] L. Montero, M. Herrero, Two-dimensional liquid chromatography approaches in Foodomics – A review, *Anal. Chim. Acta.* 1083 (2019) 1–18. doi:10.1016/j.aca.2019.07.036.
- [27] C. Ibáñez, A. Valdés, V. García-Cañas, C. Simó, M. Celebier, L. Rocamora-Reverte, Á. Gómez-Martínez, M. Herrero, M. Castro-Puyana, A. Segura-Carretero, E. Ibáñez, J.A. Ferragut, A. Cifuentes, Global Foodomics strategy to investigate the health benefits of dietary constituents, *J. Chromatogr. A.* 1248 (2012) 139–153. doi:10.1016/j.chroma.2012.06.008.
- [28] M. Cialiè Rosso, F. Stilo, C. Bicchi, M. Charron, G. Rosso, R. Menta, S.E. Reichenbach, C.H. Weinert, C.I. Mack, S.E. Kulling, C. Cordero, Combined Untargeted and Targeted Fingerprinting by Comprehensive Two-Dimensional Gas Chromatography to Track Compositional Changes on Hazelnut Primary Metabolome during Roasting, *Appl. Sci.* 11 (2021) 525. doi:10.3390/app11020525.
- [29] C. Pérez-López, B. Oró-Nolla, S. Lacorte, R. Tauler, Regions of Interest Multivariate Curve Resolution Liquid Chromatography with Data-Independent Acquisition Tandem Mass Spectrometry, *Anal. Chem.* 95 (2023) 7519–7527. doi:10.1021/acs.analchem.2c05704.
- [30] R. Costa, C. Fanali, G. Pennazza, L. Tedone, L. Dugo, M. Santonico, D. Sciarrone, F. Cacciola, L. Cucchiari, M. Dachà, L. Mondello, Screening of volatile compounds composition of white truffle during storage by GCxGC-(FID/MS) and gas sensor array analyses, *Lwt.* 60 (2015) 905–913. doi:10.1016/j.lwt.2014.09.054.
- [31] K.M. Pierce, J.L. Hope, J.C. Hoggard, R.E. Synovec, A principal component analysis based method to discover chemical differences in comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry (GC × GC-TOFMS) separations of metabolites in plant samples, *Talanta*. 70 (2006) 797–804. doi:10.1016/j.talanta.2006.01.038.
- [32] L. Van Der Maaten, G. Hinton, Visualizing data using t-SNE, *J. Mach. Learn. Res.* 9 (2008) 2579–2625.
- [33] R.T. Peterson, Chemical biology and the limits of reductionism, *Nat. Chem. Biol.* 4 (2008) 635–638. doi:10.1038/nchembio1108-635.
- [34] F. Stilo, C. Bicchi, A.M. Jimenez-Carvelo, L. Cuadros-Rodriguez, S.E. Reichenbach, C. Cordero, Chromatographic fingerprinting by comprehensive two-dimensional chromatography: Fundamentals and tools, *TrAC Trends Anal. Chem.* 134 (2021) 116133.

doi:10.1016/j.trac.2020.116133.

- [35] R.H. Stadler, L.A. Tran, C. Cavin, P. Zbinden, E.J.M. Konings, Analytical approaches to verify food integrity: Needs and challenges, in: J. AOAC Int., 2016. doi:10.5740/jaoacint.16-0231.
- [36] K.R. Schneider, J. Fanzo, L. Haddad, M. Herrero, J.R. Moncayo, A. Herforth, R. Remans, A. Guarin, D. Resnick, N. Covic, C. Béné, A. Cattaneo, N. Aburto, R. Ambikapathi, D. Aytekin, S. Barquera, J. Battersby, T. Beal, P.B. Molina, C. Cafiero, C. Campeau, P. Caron, P. Conforti, K. Damerau, M. Di Girolamo, F. DeClerck, D. Dewi, I. Elouafi, C. Fabi, P. Foley, T.J. Frazier, J. Gephart, C. Golden, C.G. Fischer, S. Hendriks, M. Honorati, J. Huang, G. Kennedy, A. Laar, R. Lal, P. Lidder, B. Loken, Q. Marshall, Y.J. Masuda, R. McLaren, L. Miachon, H. Muñoz, S. Nordhagen, N. Qayyum, M. Saisana, D. Suhardiman, U.R. Sumaila, M.T. Cullen, F.N. Tubiello, J.-L. Vivero-Pol, P. Webb, K. Wiebe, The state of food systems worldwide in the countdown to 2030, *Nat. Food*. 4 (2023) 1090–1110. doi:10.1038/s43016-023-00885-9.
- [37] S. Squara, A. Caratti, A. Fina, E. Liberto, N. Spigolon, G. Genova, G. Castello, I. Cincera, C. Bicchi, C. Cordero, Artificial Intelligence decision-making tools based on comprehensive two-dimensional gas chromatography data: the challenge of quantitative volatilomics in food quality assessment, *J. Chromatogr. A*. 1700 (2023) 464041. doi:10.1016/j.chroma.2023.464041.
- [38] F. STILO, M. del P. SEGURA BORREGO, C. BICCHI, S. BATTAGLINO, R.M. CALLEJÓN FERNADEZ, M.L. MORALES, S.E. REICHENBACH, J. MCCURRY, D. PERONI, C. CORDERO, Delineating the extra-virgin olive oil aroma blueprint by multiple headspace solid phase microextraction and differential-flow modulated comprehensive two-dimensional gas chromatography, *J. Chromatogr. A*. 1650 (2021) 462232. doi:10.1016/j.chroma.2021.462232.
- [39] R.R. Wilson, Analytical Chemistry of Biological Systems, *Anal. Chem.* 63 (1991) 379 A. doi:10.1021/ac00007a600.
- [40] Z. Liu, J.B. Phillips, Comprehensive Two-Dimensional Gas Chromatography using an On-Column Thermal Modulator Interface, *J. Chromatogr. Sci.* 29 (1991) 227–231. doi:10.1093/chromsci/29.6.227.
- [41] F. Magagna, E. Liberto, S.E. Reichenbach, Q. Tao, A. Carretta, L. Cobelli, M. Giardina, C. Bicchi, C. Cordero, Advanced fingerprinting of high-quality cocoa: Challenges in transferring methods from thermal to differential-flow modulated comprehensive two dimensional gas chromatography, *J. Chromatogr. A*. 1536 (2018) 122–136. doi:10.1016/j.chroma.2017.07.014.
- [42] M. Adahchour, J. Beens, R.J.J. Vreuls, U.A.T. Brinkman, Recent developments in comprehensive two-dimensional gas chromatography (GC x GC). III. Applications for petrochemicals and organohalogenes, *TrAC - Trends Anal. Chem.* 25 (2006) 726–741. doi:10.1016/j.trac.2006.03.005.
- [43] P.Q. Tranchida, P. Donato, F. Cacciola, M. Beccaria, P. Dugo, L. Mondello, Potential of comprehensive chromatography in food analysis, *TrAC - Trends Anal. Chem.* 52 (2013) 186–205. doi:10.1016/j.trac.2013.07.008.

- [44] J. V. Seeley, S.K. Seeley, Multidimensional gas chromatography: Fundamental advances and new applications, *Anal. Chem.* 85 (2013) 557–578. doi:10.1021/ac303195u.
- [45] P. Marriott, R. Shellie, Principles and applications of comprehensive two-dimensional gas chromatography, *TrAC - Trends Anal. Chem.* 21 (2002) 573–583. doi:10.1016/S0165-9936(02)00814-2.
- [46] M. Adahchour, J. Beens, R.J.J. Vreuls, U.A.T. Brinkman, Recent developments in comprehensive two-dimensional gas chromatography (GC x GC). II. Modulation and detection, *TrAC - Trends Anal. Chem.* 25 (2006) 540–553. doi:10.1016/j.trac.2006.04.004.
- [47] M. Adahchour, J. Beens, R.J.J. Vreuls, U.A.T. Brinkman, Recent developments in comprehensive two-dimensional gas chromatography (GC x GC). IV. Further applications, conclusions and perspectives, *TrAC - Trends Anal. Chem.* 25 (2006) 821–840. doi:10.1016/j.trac.2006.03.003.
- [48] M. Adahchour, J. Beens, R.J.J. Vreuls, U.A.T. Brinkman, Recent developments in comprehensive two-dimensional gas chromatography (GC x GC). I. Introduction and instrumental set-up, *TrAC - Trends Anal. Chem.* 25 (2006) 438–454. doi:10.1016/j.trac.2006.03.002.
- [49] C. Cordero, J. Kiefl, P. Schieberle, S.E. Reichenbach, C. Bicchi, Comprehensive two-dimensional gas chromatography and food sensory properties: potential and challenges, *Anal. Bioanal. Chem.* 407 (2015) 169–191. doi:10.1007/s00216-014-8248-z.
- [50] C. Cordero, J. Kiefl, S.E. Reichenbach, C. Bicchi, Characterization of odorant patterns by comprehensive two-dimensional gas chromatography: A challenge in omic studies, *TrAC Trends Anal. Chem.* 113 (2019) 364–378. doi:10.1016/j.trac.2018.06.005.
- [51] B. Gruber, B.A. Weggler, R. Jaramillo, K.A. Murrell, P.K. Piotrowski, F.L. Dorman, Comprehensive two-dimensional gas chromatography in forensic science: A critical review of recent trends, *TrAC - Trends Anal. Chem.* 105 (2018) 292–301. doi:10.1016/j.trac.2018.05.017.
- [52] M.S. Klee, J. Cochran, M. Merrick, L.M. Blumberg, Evaluation of conditions of comprehensive two-dimensional gas chromatography that yield a near-theoretical maximum in peak capacity gain, *J. Chromatogr. A.* 1383 (2015) 151–159. doi:10.1016/j.chroma.2015.01.031.
- [53] J.C. Giddings, Sample dimensionality: A predictor of order-disorder in component peak distribution in multidimensional separation, *J. Chromatogr. A.* 703 (1995) 3–15. doi:10.1016/0021-9673(95)00249-M.
- [54] F. Stilo, E. Gabetti, C. Bicchi, A. Carretta, D. Peroni, S.E. Reichenbach, C. Cordero, J. McCurry, A step forward in the equivalence between thermal and differential-flow modulated comprehensive two-dimensional gas chromatography methods, *J. Chromatogr. A.* 1627 (2020) 461396. doi:10.1016/j.chroma.2020.461396.
- [55] T. Kind, O. Fiehn, Advances in structure elucidation of small molecules using mass spectrometry, *Bioanal. Rev.* (2010). doi:10.1007/s12566-010-0015-9.

- [56] E. Ibáñez, A. Cifuentes, Foodomics: Food science and nutrition in the postgenomic era, 2014. doi:10.1016/B978-0-444-62650-9.00015-4.
- [57] F. Stilo, C. Bicchi, S.E. Reichenbach, C. Cordero, Comprehensive two-dimensional gas chromatography as a boosting technology in food-omic investigations, *J. Sep. Sci.* 44 (2021) 1592–1611. doi:10.1002/jssc.202100017.
- [58] M. Cialiè Rosso, M. Mazzucotelli, C. Bicchi, M. Charron, F. Manini, R. Menta, M. Fontana, S.E. Reichenbach, C. Cordero, Adding extra-dimensions to hazelnuts primary metabolome fingerprinting by comprehensive two-dimensional gas chromatography combined with time-of-flight mass spectrometry featuring tandem ionization: Insights on the aroma potential, *J. Chromatogr. A.* 1614 (2020) 1–11. doi:10.1016/j.chroma.2019.460739.
- [59] S. Squara, F. Stilo, M. Cialiè Rosso, E. Liberto, N. Spigolon, G. Genova, G. Castello, C. Bicchi, C. Cordero, *Corylus avellana* L. Aroma Blueprint: Potent Odorants Signatures in the Volatilome of High Quality Hazelnuts, *Front. Plant Sci.* 13 (2022) 1–25. doi:10.3389/fpls.2022.840028.
- [60] C. Mack, D. Wefers, P. Schuster, C.H. Weinert, B. Egert, S. Bliedung, B. Trierweiler, C. Muhle-Goll, M. Bunzel, B. Luy, S.E. Kulling, Untargeted multi-platform analysis of the metabolome and the non-starch polysaccharides of kiwifruit during postharvest ripening, *Postharvest Biol. Technol.* 125 (2017) 65–76. doi:10.1016/j.postharvbio.2016.10.011.
- [61] F. Stilo, G. Tredici, C. Bicchi, A. Robbat, J. Morimoto, C. Cordero, Climate and Processing Effects on Tea (*Camellia sinensis* L. Kuntze) Metabolome: Accurate Profiling and Fingerprinting by Comprehensive Two-Dimensional Gas Chromatography/Time-of-Flight Mass Spectrometry, *Molecules.* 25 (2020) 2447. doi:10.3390/molecules25102447.
- [62] Y.Y. Broza, P. Mochalski, V. Ruzsanyi, A. Amann, H. Haick, Hybrid volatilomics and disease detection., *Angew. Chem. Int. Ed. Engl.* 54 (2015) 11036–11048. doi:10.1002/anie.201500153.
- [63] M. Phillips, R.N. Cataneo, A. Chaturvedi, P.D. Kaplan, M. Libardoni, M. Mundada, U. Patel, X. Zhang, Detection of an extended human volatome with comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry., *PLoS One.* 8 (2013) e75274. doi:10.1371/journal.pone.0075274.
- [64] A. Amann, B. de L. Costello, W. Miekisch, J. Schubert, B. Buszewski, J. Pleil, N. Ratcliffe, T. Risby, The human volatilome: volatile organic compounds (VOCs) in exhaled breath, skin emanations, urine, feces and saliva., *J. Breath Res.* 8 (2014) 34001. doi:10.1088/1752-7155/8/3/034001.
- [65] J.F. Meadow, A.E. Altrichter, A.C. Bateman, J. Stenson, G.Z. Brown, J.L. Green, B.J.M. Bohannon, Humans differ in their personal microbial cloud., *PeerJ.* 3 (2015) e1258. doi:10.7717/peerj.1258.
- [66] M. Noshad, B.A. Behbahani, I.K. Karabagias, Volatilomic with chemometrics: a toward authentication approach for food authenticity control, *Eur. Food Res. Technol.* (2023). doi:10.1007/s00217-023-04291-0.

- [67] A.E. Lytou, E.Z. Panagou, G.-J.E. Nychas, Volatilomics for food quality and authentication, *Curr. Opin. Food Sci.* 28 (2019) 88–95. doi:10.1016/j.cofs.2019.10.003.
- [68] M.M. Ulaszewska, C.H. Weinert, A. Trimigno, R. Portmann, C. Andres Lacueva, R. Badertscher, L. Brennan, C. Brunius, A. Bub, F. Capozzi, M. Cialiè Rosso, C.E. Cordero, H. Daniel, S. Durand, B. Egert, P.G. Ferrario, E.J.M. Feskens, P. Franceschi, M. Garcia-Aloy, F. Giacomoni, P. Giesbertz, R. González-Domínguez, K. Hanhineva, L.Y. Hemeryck, J. Kopka, S.E. Kulling, R. Llorach, C. Manach, F. Mattivi, C. Migné, L.H. Münger, B. Ott, G. Picone, G. Pimentel, E. Pujos-Guillot, S. Riccadonna, M.J. Rist, C. Rombouts, J. Rubert, T. Skurk, P.S.C. Sri Harsha, L. Van Meulebroek, L. Vanhaecke, R. Vázquez-Fresno, D. Wishart, G. Vergères, Nutrimetabolomics: An Integrative Action for Metabolomic Analyses in Human Nutritional Studies, *Mol. Nutr. Food Res.* 63 (2019) 1800384. doi:10.1002/mnfr.201800384.
- [69] A. Dunkel, M. Steinhaus, M. Kotthoff, B. Nowak, D. Krautwurst, P. Schieberle, T. Hofmann, Nature's chemical signatures in human olfaction: A foodborne perspective for future biotechnology, *Angew. Chemie - Int. Ed.* 53 (2014) 7124–7143. doi:10.1002/anie.201309508.
- [70] J. Charve, C. Chen, A.D. Hegeman, G.A. Reineccius, Evaluation of instrumental methods for the untargeted analysis of chemical stimuli of orange juice flavour, *Flavour Fragr. J.* 26 (2011) 429–440. doi:10.1002/ffj.2078.
- [71] L. Nicolotti, V. Mall, P. Schieberle, Characterization of Key Aroma Compounds in a Commercial Rum and an Australian Red Wine by Means of a New Sensomics-Based Expert System (SEBES) - An Approach to Use Artificial Intelligence in Determining Food Odor Codes, *J. Agric. Food Chem.* 67 (2019) 4011–4022. doi:10.1021/acs.jafc.9b00708.
- [72] S. Pantò, M. Collard, G. Purcaro, Comprehensive Gas Chromatography Coupled to Simultaneous Dual Detection (TOF-MS/FID) as a Confirmatory Method for MOSH and MOAH Determination in Food, *Curr. Trends Mass Spectrom.* (2020) 15–20.
- [73] M. Biedermann, K. Grob, Advantages of comprehensive two-dimensional gas chromatography for comprehensive analysis of potential migrants from food contact materials, *Anal. Chim. Acta.* 1057 (2019) 11–17. doi:10.1016/j.aca.2018.10.046.
- [74] L. Cuadros-Rodríguez, C. Ruiz-Samblás, L. Valverde-Som, E. Pérez-Castaño, A. González-Casado, Chromatographic fingerprinting: An innovative approach for food “identification” and food authentication – A tutorial, *Anal. Chim. Acta.* 909 (2016) 9–23. doi:10.1016/j.aca.2015.12.042.
- [75] E. Want, P. Masson, Processing and Analysis of GC/LC-MS-Based Metabolomics Data, in: *Methods Mol. Biol.*, 2011: pp. 277–298. doi:10.1007/978-1-61737-985-7_17.
- [76] M.M. Koek, R.H. Jellema, J. van der Greef, A.C. Tas, T. Hankemeier, Quantitative metabolomics based on gas chromatography mass spectrometry: Status and perspectives, *Metabolomics.* 7 (2011) 307–328. doi:10.1007/s11306-010-0254-3.
- [77] M.C. Pedrosa, L. Lima, S. Heleno, M. Carrocho, I.C.F.R. Ferreira, L. Barros, Food metabolites as tools for authentication, processing, and nutritive value assessment,

Foods. 10 (2021) 2213. doi:10.3390/foods10092213.

- [78] I. Subramanian, S. Verma, S. Kumar, A. Jere, K. Anamika, Multi-omics Data Integration, Interpretation, and Its Application, *Bioinform. Biol. Insights*. 14 (2020) 117793221989905. doi:10.1177/1177932219899051.
- [79] T.J. Trinklein, R.E. Synovec, Simulating comprehensive two-dimensional gas chromatography mass spectrometry data with realistic run-to-run shifting to evaluate the robustness of tile-based Fisher ratio analysis, *J. Chromatogr. A*. 1677 (2022) 463321. doi:10.1016/j.chroma.2022.463321.
- [80] L.C. Class, G. Kuhnen, S. Rohn, J. Kuballa, Diving deep into the data: A review of deep learning approaches and potential applications in foodomics, *Foods*. 10 (2021) 1803. doi:10.3390/foods10081803.
- [81] V. García-Cañas, C. Simó, M. Herrero, E. Ibáñez, A. Cifuentes, Present and future challenges in food analysis: Foodomics, *Anal. Chem.* 84 (2012) 10150–10159. doi:10.1021/ac301680q.
- [82] M. Mongia, H. Mohimani, Repository scale classification and decomposition of tandem mass spectral data, *Sci. Rep.* 11 (2021) 8314. doi:10.1038/s41598-021-87796-6.
- [83] A. Valdés, G. Álvarez-Rivera, B. Socas-Rodríguez, M. Herrero, E. Ibáñez, A. Cifuentes, Foodomics: Analytical Opportunities and Challenges, *Anal. Chem.* 94 (2022) 366–381. doi:10.1021/acs.analchem.1c04678.
- [84] S. Yamamoto, T. Bamba, A. Sano, Y. Kodama, M. Imamura, A. Obata, E. Fukusaki, Metabolite profiling of soy sauce using gas chromatography with time-of-flight mass spectrometry and analysis of correlation with quantitative descriptive analysis, *J. Biosci. Bioeng.* 114 (2012) 170–175. doi:10.1016/j.jbiosc.2012.03.023.
- [85] A. Valdés, A. Cifuentes, C. León, Foodomics evaluation of bioactive compounds in foods, *TrAC - Trends Anal. Chem.* 96 (2017) 2–13. doi:10.1016/j.trac.2017.06.004.
- [86] P. Balkir, K. Kemahlioglu, U. Yucel, Foodomics: A new approach in food quality and safety, *Trends Food Sci. Technol.* 108 (2021) 49–57. doi:10.1016/j.tifs.2020.11.028.
- [87] M. Schonlau, R.Y. Zou, The random forest algorithm for statistical learning, *Stata J.* 20 (2020) 3–29. doi:10.1177/1536867X20909688.
- [88] E. Górská-Horczyczak, M. Horczyczak, D. Guzek, I. Wojtasik-Kalinowska, A. Wierzbicka, Chromatographic fingerprints supported by artificial neural network for differentiation of fresh and frozen pork, *Food Control*. 73 (2017) 237–244. doi:10.1016/j.foodcont.2016.08.010.
- [89] H. Tsugawa, Y. Tsujimoto, M. Arita, T. Bamba, E. Fukusaki, GC/MS based metabolomics: Development of a data mining system for metabolite identification by using soft independent modeling of class analogy (SIMCA), *BMC Bioinformatics*. 12 (2011) 131. doi:10.1186/1471-2105-12-131.
- [90] B. Khakimov, G. Gürdeniz, S.B. Engelse, Trends in the application of chemometrics to foodomics studies, *Acta Aliment.* 44 (2015) 4–31. doi:10.1556/AAlim.44.2015.1.1.
- [91] G. Hanrahan, F. Udeh, D.G. Patil, Multivariate Calibration Techniques, in: *Encycl. Anal.*

Sci. Second Ed., Elsevier, 2004: pp. 27–32. doi:10.1016/Bo-12-369397-7/00077-7.

- [92] V.K. Truong, M. Dupont, A. Elbourne, S. Gangadoo, P.R. Pathirannahalage, S. Cheeseman, J. Chapman, D. Cozzolino, From academia to reality check: A theoretical framework on the use of chemometric in food sciences, *Foods*. 8 (2019) 164. doi:10.3390/foods8050164.
- [93] T. Skov, A.H. Honoré, H.M. Jensen, T. Næs, S.B. Engelsen, Chemometrics in foodomics: Handling data structures from multiple analytical platforms, *TrAC - Trends Anal. Chem.* 60 (2014) 71–79. doi:10.1016/j.trac.2014.05.004.
- [94] J.C. Hoggard, R.E. Synovec, Parallel factor analysis (PARAFAC) of target analytes in GC × GC-TOFMS data: Automated selection of a model with an appropriate number of factors, *Anal. Chem.* 79 (2007) 1611–1619. doi:10.1021/aco61710b.
- [95] V. D'Atri, T. Causon, O. Hernandez-Alba, A. Mutabazi, J.L. Veuthey, S. Cianferani, D. Guillarme, Adding a new separation dimension to MS and LC-MS: What is the utility of ion mobility spectrometry?, *J. Sep. Sci.* 41 (2018) 20–67. doi:10.1002/jssc.201700919.
- [96] S.P. Putri, M.M.M. Ikram, A. Sato, H.A. Dahlan, D. Rahmawati, Y. Ohto, E. Fukusaki, Application of gas chromatography-mass spectrometry-based metabolomics in food science and technology, *J. Biosci. Bioeng.* 133 (2022) 425–435. doi:10.1016/j.jbiosc.2022.01.011.
- [97] E. Gorrochategui, J. Jaumot, R. Tauler, ROIMCR: A powerful analysis strategy for LC-MS metabolomic datasets, *BMC Bioinformatics*. 20 (2019) 1–17. doi:10.1186/s12859-019-2848-8.
- [98] M. Pérez-Cova, J. Jaumot, R. Tauler, Untangling comprehensive two-dimensional liquid chromatography data sets using regions of interest and multivariate curve resolution approaches, *TrAC - Trends Anal. Chem.* 137 (2021). doi:10.1016/j.trac.2021.116207.
- [99] J. de Sousa, O. Vencálek, K. Hron, J. Václavík, D. Friedecký, T. Adam, Bayesian multiple hypotheses testing in compositional analysis of untargeted metabolomic data, *Anal. Chim. Acta*. 1097 (2020) 49–61. doi:10.1016/j.aca.2019.11.006.
- [100] P.H. Stefanuto, A. Smolinska, J.F. Focant, Advanced chemometric and data handling tools for GC×GC-TOF-MS: Application of chemometrics and related advanced data handling in chemical separations, *TrAC - Trends Anal. Chem.* 139 (2021) 116251. doi:10.1016/j.trac.2021.116251.
- [101] J. Selamat, N.A.A. Rozani, S. Murugesu, Application of the metabolomics approach in food authentication, *Molecules*. 26 (2021) 1–19. doi:10.3390/molecules26247565.
- [102] D. Cozzolino, Foodomics and infrared spectroscopy: from compounds to functionality, *Curr. Opin. Food Sci.* 4 (2015) 39–43. doi:10.1016/j.cofs.2015.05.003.
- [103] M. Petersen, Z. Yu, X. Lu, Application of Raman Spectroscopic Methods in Food Safety: A Review, *Biosensors*. 11 (2021) 187. doi:10.3390/bios11060187.
- [104] Y. Sun, H. Tang, X. Zou, G. Meng, N. Wu, Raman spectroscopy for food quality assurance and safety monitoring: a review, *Curr. Opin. Food Sci.* 47 (2022) 100910.

doi:10.1016/j.cofs.2022.100910.

- [105] A. Trimigno, F.C. Marincola, N. Dellarosa, G. Picone, L. Laghi, Definition of food quality by NMR-based foodomics, *Curr. Opin. Food Sci.* 4 (2015) 99–104. doi:10.1016/j.cofs.2015.06.008.
- [106] K. Fan, M. Zhang, Recent developments in the food quality detected by non-invasive nuclear magnetic resonance technology, *Crit. Rev. Food Sci. Nutr.* 59 (2019) 2202–2213. doi:10.1080/10408398.2018.1441124.
- [107] C.M. Andre, C. Soukoulis, Food quality assessed by chemometrics, *Foods*. 9 (2020) 2–5. doi:10.3390/foods9070897.
- [108] J.L. Aleixandre-Tudo, L. Castello-Cogollos, J.L. Aleixandre, R. Aleixandre-Benavent, Chemometrics in food science and technology: A bibliometric study, *Chemom. Intell. Lab. Syst.* 222 (2022) 104514. doi:10.1016/j.chemolab.2022.104514.
- [109] D. Song, Q. Wu, M. Kamruzzaman, Appropriate use of chemometrics for feasibility study for developing low-cost filter-based multi-parameter detection spectroscopic device for meat proximate analysis, *Chemom. Intell. Lab. Syst.* 238 (2023) 104844. doi:10.1016/j.chemolab.2023.104844.
- [110] K. Wang, Z. Li, J. Li, H. Lin, Raman spectroscopic techniques for nondestructive analysis of agri-foods: A state-of-the-art review, *Trends Food Sci. Technol.* 118 (2021) 490–504. doi:10.1016/j.tifs.2021.10.010.
- [111] A.C. da C. Fulgêncio, G.A.P. Resende, M.C.F. Teixeira, B.G. Botelho, M.M. Sena, Combining portable NIR spectroscopy and multivariate calibration for the determination of ethanol in fermented alcoholic beverages by a multi-product model, *Talanta Open*. 7 (2023) 100180. doi:10.1016/j.talo.2023.100180.
- [112] Z. Kang, Y. Zhao, L. Chen, Y. Guo, Q. Mu, S. Wang, Advances in Machine Learning and Hyperspectral Imaging in the Food Supply Chain, *Food Eng. Rev.* 14 (2022) 596–616. doi:10.1007/s12393-022-09322-2.
- [113] D. Saha, A. Manickavasagan, Machine learning techniques for analysis of hyperspectral images to determine quality of food products: A review, *Curr. Res. Food Sci.* 4 (2021) 28–44. doi:10.1016/j.cofs.2021.01.002.
- [114] C.-H. Feng, Y. Makino, S. Oshita, J.F. García Martín, Hyperspectral imaging and multispectral imaging as the novel techniques for detecting defects in raw and processed meat products: Current state-of-the-art research advances, *Food Control*. 84 (2018) 165–176. doi:10.1016/j.foodcont.2017.07.013.
- [115] A. Soni, Y. Dixit, M.M. Reis, G. Brightwell, Hyperspectral imaging and machine learning in food microbiology: Developments and challenges in detection of bacterial, fungal, and viral contaminants, *Compr. Rev. Food Sci. Food Saf.* 21 (2022) 3717–3745. doi:10.1111/1541-4337.12983.
- [116] Z. Gao, Y. Shao, G. Xuan, Y. Wang, Y. Liu, X. Han, Real-time hyperspectral imaging for the in-field estimation of strawberry ripeness with deep learning, *Artif. Intell. Agric.* 4 (2020) 31–38. doi:10.1016/j.aiia.2020.04.003.
- [117] N.N. Wang, D.W. Sun, Y.C. Yang, H. Pu, Z. Zhu, Recent Advances in the Application of

Hyperspectral Imaging for Evaluating Fruit Quality, *Food Anal. Methods.* 9 (2016) 178–191. doi:10.1007/s12161-015-0153-3.

- [118] H. Seki, T. Ma, H. Murakami, S. Tsuchikawa, T. Inagaki, Visualization of Sugar Content Distribution of White Strawberry by Near-Infrared Hyperspectral Imaging, *Foods.* 12 (2023) 931. doi:10.3390/foods12050931.
- [119] B. Park, S.-C. Yoon, Real-Time Hyperspectral Imaging for Food Safety, in: 2015: pp. 305–329. doi:10.1007/978-1-4939-2836-1_13.
- [120] L.L.P. van Stee, U.A.T. Brinkman, Peak detection methods for GC × GC: An overview, *TrAC - Trends Anal. Chem.* 83 (2016) 1–13. doi:10.1016/j.trac.2016.07.009.
- [121] J.T.V. Matos, R.M.B.O. Duarte, A.C. Duarte, Trends in data processing of comprehensive two-dimensional chromatography: State of the art, *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 910 (2012) 31–45. doi:10.1016/j.jchromb.2012.06.039.
- [122] X. Ning, I.W. Selesnick, L. Duval, Chromatogram baseline estimation and denoising using sparsity (BEADS), *Chemom. Intell. Lab. Syst.* 139 (2014) 156–167. doi:10.1016/j.chemolab.2014.09.014.
- [123] T.F. Tian, S.Y. Wang, T.C. Kuo, C.E. Tan, G.Y. Chen, C.H. Kuo, C.H.S. Chen, C.C. Chan, O.A. Lin, Y.J. Tseng, Web Server for Peak Detection, Baseline Correction, and Alignment in Two-Dimensional Gas Chromatography Mass Spectrometry-Based Metabolomics Data, *Anal. Chem.* 88 (2016) 10395–10403. doi:10.1021/acs.analchem.6b00755.
- [124] B.Y. Using, T.H.E. Codes, T.H.E. User, A. To, T.H.E. License, T. Statedthe, F. License, Documentation for the Gros-Eilers-Arey code to perform baseline correction of GC × GC chromatograms as implemented in Matlab, (2017) 1–4.
- [125] C. Couprie, L. Duval, M. Moreaud, S. Hénon, M. Tebib, V. Souchon, BARCHAN: Blob Alignment for Robust CHromatographic ANalysis, *J. Chromatogr. A.* 1484 (2017) 65–72. doi:10.1016/j.chroma.2017.01.003.
- [126] S. Kim, M. Ouyang, J. Jeong, C. Shen, X. Zhang, A new method of peak detection for analysis of comprehensive two-dimensional gas chromatography mass spectrometry data, *Ann. Appl. Stat.* 8 (2014) 1209–1231. doi:10.1214/14-AOAS731.
- [127] P. Zhang, S. Carlin, P. Franceschi, F. Mattivi, U. Vrhovsek, Application of a Target-Guided Data Processing Approach in Saturated Peak Correction of GC×GC Analysis, *Anal. Chem.* 94 (2022) 1941–1948. doi:10.1021/acs.analchem.1c02719.
- [128] C.T. Dimensional, B. Correction, Under typical controlled conditions , the, Baseline. (2010) 1–4.
- [129] Y. Zheng, R. Fan, C. Qiu, Z. Liu, D. Tian, An improved algorithm for peak detection in mass spectra based on continuous wavelet transform, *Int. J. Mass Spectrom.* 409 (2016) 53–58. doi:10.1016/j.ijms.2016.09.020.
- [130] S.E. Reichenbach, Chapter 4 Data Acquisition, Visualization, and Analysis, *Compr. Anal. Chem.* 55 (2009) 77–106. doi:10.1016/S0166-526X(09)05504-4.

- [131] S.E. Reichenbach, X. Tian, C. Cordero, Q. Tao, Features for non-targeted cross-sample analysis with comprehensive two-dimensional chromatography, *J. Chromatogr. A*. 1226 (2012) 140–148. doi:10.1016/j.chroma.2011.07.046.
- [132] K.L. Berrier, S.E. Prebihalo, R.E. Synovec, Advanced data handling in comprehensive two-dimensional gas chromatography, in: *Sep. Sci. Technol.* (New York), Elsevier, 2020: pp. 229–268. doi:10.1016/B978-0-12-813745-1.00007-6.
- [133] H.D. Bahaghighat, C.E. Freye, R.E. Synovec, Recent advances in modulator technology for comprehensive two dimensional gas chromatography, *TrAC - Trends Anal. Chem.* 113 (2019) 379–391. doi:10.1016/j.trac.2018.04.016.
- [134] M. Ni, S.E. Reichenbach, Pattern matching by sequential subdivision of transformation space, in: *Proc. - Int. Conf. Pattern Recognit.*, 2004: pp. 145–148. doi:10.1109/ICPR.2004.1334082.
- [135] M. Bordiga, M. Rinaldi, M. Locatelli, G. Piana, F. Travaglia, J.D. Coisson, M. Arlorio, Characterization of Muscat wines aroma evolution using comprehensive gas chromatography followed by a post-analytic approach to 2D contour plots comparison, *Food Chem.* 140 (2013) 57–67. doi:10.1016/j.foodchem.2013.02.051.
- [136] T.J. Trinklein, C.N. Cain, G.S. Ochoa, S. Schöneich, L. Mikaliunaite, R.E. Synovec, Recent Advances in GC×GC and Chemometrics to Address Emerging Challenges in Nontargeted Analysis, *Anal. Chem.* 95 (2023) 264–286. doi:10.1021/acs.analchem.2c04235.
- [137] What is Computer Vision?, (n.d.). <https://www.ibm.com/topics/computer-vision> (accessed November 25, 2022).
- [138] G. Danuser, Computer vision in cell biology, *Cell.* (2011). doi:10.1016/j.cell.2011.11.001.
- [139] S. Mori, H. Saito, An overview of augmented visualization: observing the real world as desired, *APSIPA Trans. Signal Inf. Process.* 7 (2018). doi:10.1017/ATSIP.2018.13.
- [140] S.E. Reichenbach, P.W. Carr, D.R. Stoll, Q. Tao, Smart Templates for peak pattern matching with comprehensive two-dimensional liquid chromatography, *J. Chromatogr. A*. 1216 (2009) 3458–3466. doi:10.1016/j.chroma.2008.09.058.
- [141] GC ImageTM, GC Image GC×GC Edition Users' Guide, 2017.
- [142] D. Maltoni, R. Cappelli, D. Meuwly, Challenges for Fingerprint Recognition Spoofing, Skin Diseases, and Environmental Effects, *Handb. Biometrics Forensic Sci.* (2017) 63–83. doi:10.1007/978-3-319-50673-9.
- [143] B. V. Hollingsworth, S.E. Reichenbach, Q. Tao, A. Visvanathan, Comparative visualization for comprehensive two-dimensional gas chromatography, in: *J. Chromatogr. A*, 2006: pp. 51–58. doi:10.1016/j.chroma.2005.11.074.
- [144] K.J. Johnson, R.E. Synovec, Pattern recognition of jet fuels: Comprehensive GC × GC with ANOVA-based feature selection and principal component analysis, in: *Chemom. Intell. Lab. Syst.*, 2002: pp. 225–237. doi:10.1016/S0169-7439(01)00198-8.
- [145] R.E. Mohler, K.M. Dombek, J.C. Hoggard, E.T. Young, R.E. Synovec, Comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry analysis of

metabolites in fermenting and respiring yeast cells, *Anal. Chem.* 78 (2006) 2700–2709. doi:10.1021/ac0521060.

- [146] K.M. Pierce, J.C. Hoggard, J.L. Hope, P.M. Rainey, A.N. Hoofnagle, R.M. Jack, B.W. Wright, R.E. Synovec, Fisher ratio method applied to third-order separation data to identify significant chemical components of metabolite extracts, *Anal. Chem.* 78 (2006) 5068–5075. doi:10.1021/ac0602625.
- [147] J. Vial, B. Pezous, D. Thiébaud, P. Sassi, B. Teillet, X. Cahours, I. Rivals, The discriminant pixel approach: A new tool for the rational interpretation of GCxGC-MS chromatograms, *Talanta*. 83 (2011) 1295–1301. doi:10.1016/j.talanta.2010.07.059.
- [148] S.E. Reichenbach, D.W. Rempe, Q. Tao, D. Bressanello, E. Liberto, C. Bicchi, S. Balducci, C. Cordero, Alignment for Comprehensive Two-Dimensional Gas Chromatography with Dual Secondary Columns and Detectors, *Anal. Chem.* 87 (2015) 10056–10063. doi:10.1021/acs.analchem.5b02718.
- [149] B.A. Parsons, L.C. Marney, W.C. Siegler, J.C. Hoggard, B.W. Wright, R.E. Synovec, Tile-Based Fisher Ratio Analysis of Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry (GC × GC–TOFMS) Data Using a Null Distribution Approach, *Anal. Chem.* 87 (2015) 3812–3819. doi:10.1021/ac504472s.
- [150] L.C. Marney, W. Christopher Siegler, B.A. Parsons, J.C. Hoggard, B.W. Wright, R.E. Synovec, Tile-based Fisher-ratio software for improved feature selection analysis of comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry data, *Talanta*. (2013). doi:10.1016/j.talanta.2013.06.038.
- [151] P.E. Sudol, G.S. Ochoa, R.E. Synovec, Investigation of the limit of discovery using tile-based Fisher ratio analysis with comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry, *J. Chromatogr. A.* (2021). doi:10.1016/j.chroma.2021.462092.
- [152] F. Magagna, L. Valverde-Som, C. Ruiz-Samblás, L. Cuadros-Rodríguez, S.E. Reichenbach, C. Bicchi, C. Cordero, Combined untargeted and targeted fingerprinting with comprehensive two-dimensional chromatography for volatiles and ripening indicators in olive oil, *Anal. Chim. Acta.* 936 (2016) 245–258. doi:10.1016/j.aca.2016.07.005.
- [153] S.E. Reichenbach, C.A. Zini, K.P. Nicolli, J.E. Welke, C. Cordero, Q. Tao, Benchmarking machine learning methods for comprehensive chemical fingerprinting and pattern recognition, *J. Chromatogr. A.* 1595 (2019) 158–167. doi:10.1016/j.chroma.2019.02.027.
- [154] J. Kiefl, C. Cordero, L. Nicolotti, P. Schieberle, S.E. Reichenbach, C. Bicchi, Performance evaluation of non-targeted peak-based cross-sample analysis for comprehensive two-dimensional gas chromatography-mass spectrometry data and application to processed hazelnut profiling, *J. Chromatogr. A.* 1243 (2012) 81–90. doi:10.1016/j.chroma.2012.04.048.
- [155] P.E. Sudol, M. Galletta, P.Q. Tranchida, M. Zoccali, L. Mondello, R.E. Synovec, Untargeted profiling and differentiation of geographical variants of wine samples using headspace solid-phase microextraction flow-modulated comprehensive two-

dimensional gas chromatography with the support of tile-based Fisher ratio analysis, *J. Chromatogr. A.* 1662 (2022) 462735. doi:10.1016/j.chroma.2021.462735.

- [156] M. Biedermann, K. Grob, On-line coupled high performance liquid chromatography-gas chromatography for the analysis of contamination by mineral oil. Part 2: Migration from paperboard into dry foods: Interpretation of chromatograms, *J. Chromatogr. A.* 1255 (2012) 76–99. doi:10.1016/j.chroma.2012.05.096.
- [157] M. Biedermann, K. Grob, Comprehensive two-dimensional gas chromatography for characterizing mineral oils in foods and distinguishing them from synthetic hydrocarbons, *J. Chromatogr. A.* 1375 (2015) 146–153. doi:10.1016/j.chroma.2014.11.064.
- [158] L. Menegoz Ursol, C. Conchione, D. Peroni, A. Carretta, S. Moret, A study on the impact of harvesting operations on the mineral oil contamination of olive oils, *Food Chem.* 406 (2023) 135032. doi:10.1016/j.foodchem.2022.135032.
- [159] C. Cordero, E. Liberto, C. Bicchi, P. Rubiolo, S.E. Reichenbach, X. Tian, Q. Tao, Targeted and non-targeted approaches for complex natural sample profiling by GC×GC-qMS, *J. Chromatogr. Sci.* 48 (2010) 251–261. doi:10.1093/chromsci/48.4.251.
- [160] S.E. Reichenbach, X. Tian, Q. Tao, D.R. Stoll, P.W. Carr, Comprehensive feature analysis for sample classification with comprehensive two-dimensional LC, *J. Sep. Sci.* 33 (2010) 1365–1374. doi:10.1002/jssc.200900859.
- [161] S.E. Reichenbach, X. Tian, Q. Tao, E.B. Ledford, Z. Wu, O. Fiehn, Informatics for cross-sample analysis with comprehensive two-dimensional gas chromatography and high-resolution mass spectrometry (GC×GC-HRMS), *Talanta.* 83 (2011) 1279–1288. doi:10.1016/j.talanta.2010.09.057.
- [162] H.G. Schmarr, J. Bernhardt, U. Fischer, A. Stephan, P. Müller, D. Durner, Two-dimensional gas chromatographic profiling as a tool for a rapid screening of the changes in volatile composition occurring due to microoxygenation of red wines, *Anal. Chim. Acta.* 672 (2010) 114–123. doi:10.1016/j.aca.2010.05.002.
- [163] H.-G. Schmarr, J. Bernhardt, Profiling analysis of volatile compounds from fruits using comprehensive two-dimensional gas chromatography and image processing techniques, *J. Chromatogr. A.* 1217 (2010) 565–574. doi:10.1016/j.chroma.2009.11.063.
- [164] S. Luhn, M. Berth, M. Hecker, J. Bernhard, Using standard positions and image fusion to create proteome maps from collections of two-dimensional gel electrophoresis images, in: *Proteomics*, 2003: pp. 1117–1127. doi:10.1002/pmic.200300433.
- [165] F. Stilo, E. Liberto, S.E. Reichenbach, Q. Tao, C. Bicchi, C. Cordero, Untargeted and Targeted Fingerprinting of Extra Virgin Olive Oil Volatiles by Comprehensive Two-Dimensional Gas Chromatography with Mass Spectrometry: Challenges in Long-Term Studies, *J. Agric. Food Chem.* 67 (2019) 5289–5302. doi:10.1021/acs.jafc.9b01661.
- [166] D.W. Rempe, S.E. Reichenbach, Q. Tao, C. Cordero, W.E. Rathbun, C.A. Zini, Effectiveness of Global, Low-Degree Polynomial Transformations for GC×GC Data Alignment, *Anal. Chem.* 88 (2016) 10028–10035. doi:10.1021/acs.analchem.6b02254.
- [167] D. Geschwender, T. Qingping, C. Heble, S.E. Reichenbach, Automated Alignment for Quantitative Pairwise Differencing of Two-Dimensional Chromatography Data, in: 14th

Multidimens. Chromatogr. Work. Liege, Liege, Belgium, 2023.

- [168] E.Z. Psarakis, G.D. Evangelidis, An enhanced correlation-based method for stereo correspondence with subpixel accuracy, in: Tenth IEEE Int. Conf. Comput. Vis. Vol. 1, IEEE, 2005: p. 907–912 Vol. 1. doi:10.1109/ICCV.2005.33.
- [169] P.J. Burt, Fast filter transform for image processing, Comput. Graph. Image Process. (1981). doi:10.1016/0146-664X(81)90092-7.
- [170] M. Granvogl, P. Schieberle, The sensomics approach: A useful tool to unravel the genuine aroma blueprint of foods and aroma changes during food processing, in: Compr. Anal. Chem., 1st ed., Elsevier B.V., 2022: pp. 41–68. doi:10.1016/bs.coac.2021.10.002.
- [171] C. Kulsing, Y. Nolvachai, P.J. Marriott, Concepts, selectivity options and experimental design approaches in multidimensional and comprehensive two-dimensional gas chromatography, TrAC - Trends Anal. Chem. 130 (2020) 115995. doi:10.1016/j.trac.2020.115995.
- [172] S.T. Chin, G.T. Eyres, P.J. Marriott, System design for integrated comprehensive and multidimensional gas chromatography with mass spectrometry and olfactometry, Anal. Chem. 84 (2012) 9154–9162. doi:10.1021/ac301847y.
- [173] S.T. Chin, G.T. Eyres, P.J. Marriott, Identification of potent odourants in wine and brewed coffee using gas chromatography-olfactometry and comprehensive two-dimensional gas chromatography, J. Chromatogr. A. 1218 (2011) 7487–7498. doi:10.1016/j.chroma.2011.06.039.
- [174] B. Maikhunthod, P.D. Morrison, D.M. Small, P.J. Marriott, Development of a switchable multidimensional/comprehensive two-dimensional gas chromatographic analytical system, J. Chromatogr. A. 1217 (2010) 1522–1529. doi:10.1016/j.chroma.2009.12.078.
- [175] B. d'Acampora Zellner, A. Casilli, P. Dugo, G. Dugo, L. Mondello, Odour fingerprint acquisition by means of comprehensive two-dimensional gas chromatography-olfactometry and comprehensive two-dimensional gas chromatography/mass spectrometry, J. Chromatogr. A. 1141 (2007) 279–286. doi:https://doi.org/10.1016/j.chroma.2006.12.035.
- [176] E. Gabetti, B. Sgorbini, F. Stilo, C. Bicchi, P. Rubiolo, F. Chialva, S.E. Reichenbach, V. Bongiovanni, C. Cordero, A. Cavallero, Chemical fingerprinting strategies based on comprehensive two-dimensional gas chromatography combined with gas chromatography-olfactometry to capture the unique signature of Piemonte peppermint essential oil (*Mentha x piperita* var *Italo-Mitcham*), J. Chromatogr. A. 1645 (2021) 462101. doi:10.1016/j.chroma.2021.462101.
- [177] G. Purcaro, C. Cordero, E. Liberto, C. Bicchi, L.S. Conte, Toward a definition of blueprint of virgin olive oil by comprehensive two-dimensional gas chromatography, J. Chromatogr. A. 1334 (2014) 101–111. doi:10.1016/j.chroma.2014.01.067.
- [178] A. Chaintreau, Simultaneous distillation – extraction : from birth to maturity — review, Flavour Fragr. J. 16 (2001) 136–148.

- [179] K. Hameed, D. Chai, A. Rassau, Texture-based latent space disentanglement for enhancement of a training dataset for ANN-based classification of fruit and vegetables, *Inf. Process. Agric.* (2023). doi:10.1016/j.inpa.2021.09.003.
- [180] H. Hotelling, Analysis of a complex of statistical variables into principal components., *J. Educ. Psychol.* 24 (1933) 417–441. doi:10.1037/h0071325.
- [181] N. Bratchell, Cluster analysis, *Chemom. Intell. Lab. Syst.* 6 (1989) 105–125. doi:10.1016/0169-7439(87)80054-0.
- [182] L. McInnes, J. Healy, J. Melville, UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction, (2018). <http://arxiv.org/abs/1802.03426>.
- [183] T.H. Dat, C. Guan, Feature selection based on Fisher Ratio and Mutual Information analyses for robust Brain Computer Interface, in: *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc., IEEE, 2007: p. 1-337-1-340*. doi:10.1109/ICASSP.2007.366685.
- [184] G. Guo, H. Wang, D. Bell, Y. Bi, K. Greer, KNN model-based approach in classification, *Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*. 2888 (2003) 986–996. doi:10.1007/978-3-540-39964-3_62.
- [185] J.C. Stoltzfus, Logistic regression: A brief primer, *Acad. Emerg. Med.* 18 (2011) 1099–1104. doi:10.1111/j.1553-2712.2011.01185.x.
- [186] G.I. Webb, Naïve Bayes, in: *Encycl. Mach. Learn.*, Springer US, Boston, MA, 2011: pp. 713–714. doi:10.1007/978-0-387-30164-8_576.
- [187] Y. Qin, A review of quadratic discriminant analysis for high-dimensional data, *Wiley Interdiscip. Rev. Comput. Stat.* 10 (2018). doi:10.1002/wics.1434.
- [188] D. Ruiz-Perez, H. Guan, P. Madhivanan, K. Mathee, G. Narasimhan, So you think you can PLS-DA?, *BMC Bioinformatics*. 21 (2020) 2. doi:10.1186/s12859-019-3310-7.
- [189] Z. Jun, The Development and Application of Support Vector Machine, in: *J. Phys. Conf. Ser.*, 2021: p. 052006. doi:10.1088/1742-6596/1748/5/052006.
- [190] R. Vitale, M. Cocchi, A. Biancolillo, C. Ruckebusch, F. Marini, Class modelling by Soft Independent Modelling of Class Analogy: why, when, how? A tutorial, *Anal. Chim. Acta.* 1270 (2023) 341304. doi:10.1016/j.aca.2023.341304.
- [191] M. Casale, C. Malegori, P. Oliveri, E. Liberto, P. Rubiolo, C. Bicchi, C. Cordero, Chapter 10: Chemometrics: Basic principles and applications, *Food Chem. Funct. Anal.* 2020–Janua (2020) 403–451. doi:10.1039/9781788015752-00403.
- [192] S. Wold, M. Sjöström, L. Eriksson, PLS-regression: A basic tool of chemometrics, in: *Chemom. Intell. Lab. Syst.*, 2001: pp. 109–130. doi:10.1016/S0169-7439(01)00155-1.
- [193] R. Bro, PARAFAC. Tutorial and applications, in: *Chemom. Intell. Lab. Syst.*, 1997: pp. 149–171. doi:10.1016/S0169-7439(97)00032-4.
- [194] C.A. Andersson, R. Bro, Improving the speed of multi-way algorithms: Part I. Tucker3, *Chemom. Intell. Lab. Syst.* 42 (1998) 93–103. doi:10.1016/S0169-7439(98)00010-0.

Chapter 5

Comprehensive Two-Dimensional Gas Chromatography as a Key-Technology in Food-omics

This chapter explores the role of comprehensive two dimensional gas chromatography (GC×GC) in the investigation workflows of food-omics and its related fields, such as food metabolomics, nutrimentalomics, sensomics, and food volatilomics. Some key applications dealing with food safety are also mentioned to highlight the actual potential of the technique in providing information on both known and unknown sample components. It begins with a brief overview of the broader context of systems biology and integrative investigation methods. Following this introduction, the chapter delves into the concepts of analytical dimensions and the distinctive features of GC×GC. By presenting a range of relevant examples, it illustrates how GC×GC has enhanced food-omics research, demonstrating how comprehensive multidimensional separation platforms have brought new concepts and tools to the analysis of complex biological samples.

This chapter was submitted for publication in the book entitled “Gas Chromatography-Mass Spectrometry for Omics Applications” Edited by Flavio Antonio Franchina for John Wiley and Sons, Inc., 111 River St, Hoboken, NJ 07030, USA.

5.1 Introduction

Comprehensive two-dimensional gas chromatography (Comprehensive 2DGC or GC×GC) effectively exploits and realizes the concept of a multidimensional (MD) technique as defined by Giddings in 1987 [1]: *"a separation system is considered multidimensional when (i) all components of a sample undergo two or more separation steps, where their movements depend on different variables, and (ii) if two or more components are substantially separated in any single step, they should remain separated until the completion of the separation process."* This definition clearly established the principle of multidimensionality, at that time primarily applicable to two-dimensional thin-layer chromatography (TLC) and paper chromatography, which were widely utilized by biochemists. Within just four years, John Phillips, together with Zi Liu, translated this theoretical framework into practice for gas chromatography (GC) and developed a specialized instrument for this purpose [2]. The combined contributions of the concepts introduced by Giddings and Phillips have laid the groundwork for further MD chromatographic techniques, extending beyond GC to include liquid chromatography (LC) and supercritical fluid chromatography (SFC), such as LC×LC, SFC×SFC, and SFC×LC.

Like all innovative techniques, GC×GC had to progress through three distinct phases to gain full acceptance. The first phase involved the introduction and optimization of its theoretical foundations and operational parameters, spanning approximately 10 years (1991-2000). During this period, the petrochemical industry was the primary driver in the development of GC×GC due to its ongoing need to characterize highly complex samples [3], [4]. The subsequent phase, lasting about another decade (2001-2012), was marked by the popularization of the technique across a diverse range of scientific disciplines [5]–[7]. This era also saw the routine implementation of fingerprinting methods in biological research, where a comprehensive analysis of complex mixtures in a single step became crucial, especially in omics fields such as metabolomics (as discussed below) [8]. Concurrently, there was significant advancement in hardware, standardization of instrumentation, optimization of GC×GC coupling with mass spectrometry (MS), and the development of dedicated software [5], [6], [9]–[13]. Finally, in the current phase, GC×GC has transitioned from academic research settings into routine applications, where ongoing analytical innovation continues to shape new analytical strategies, including the adoption of "dilute-and-shoot" methods.

In this context, research studies utilizing GC×GC as the primary analytical tool should: (i) configure the analytical platform appropriately by integrating all relevant dimensions of analysis; (ii) establish a data processing strategy that optimizes the extraction of meaningful information from raw data; and (iii) employ effective data mining techniques to interpret the results and gain deeper insights into the studied phenomenon. These components of the systematic approach are illustrated in **Figure 5.1**, showing their connection to the investigative goals in food omics. While a detailed discussion of each key element is beyond the scope of this chapter, readers interested in recent developments in data processing and mining for "omics" applications are encouraged to consult recent reviews [9], [11]–[17].

GC×GC has evolved into a fully independent yet complementary technique to conventional one-dimensional GC (1D-GC), allowing for the analysis of highly complex mixtures that

exceed the capabilities of 1D-GC, which typically struggles with samples containing more than 250 components.

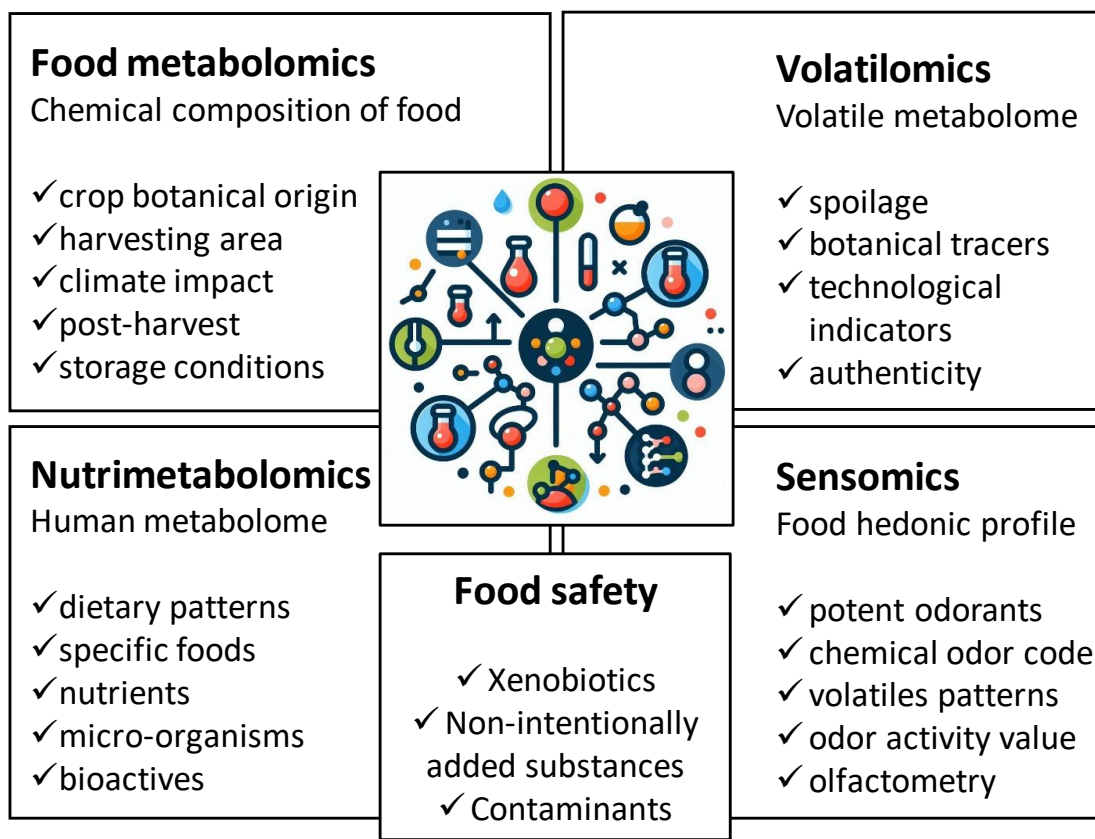


Figure 5.1: Main investigation goals of food omics (i.e., food metabolomics, food volatilomics, sensomics, nutrimetabolomics, and food safety).

Following a brief introductory overview of the principles of systems biology and the concept of analytical dimensions, this chapter will examine the role of GC×GC in various disciplines and subdisciplines related to the food sector (food-omics), including food metabolomics, nutrimetabolomics, sensomics, volatilomics, and food safety.

5.2 Analytical chemistry toolbox to advance knowledge at the boundary between chemistry and biology

The complex nature of biological systems and the numerous levels of interactions underlying biological processes have driven the growth of systems biology, a field focused on modeling this complexity by revealing higher-order network structures and relationships between different conditions, while considering the impact of molecular patterns on multiple targets [18]. In contrast, chemical biology, which has developed alongside systems biology, employs well-established strategies aimed at linking individual compounds to specific targets, following a “one compound/one target” paradigm, often referred to as a reductionist approach in biology [18].

Food chemical research has transitioned from a reductionist approach to the integrative mindset of systems biology, driven by the growing demand for high-quality food rich in nutritional value and bioactive components, as well as advancements in analytical chemistry methods and tools. This shift was anticipated in 1991 by R. Wilson in an editorial for *Analytical Chemistry* [19], where he noted: *"a persistent research frontier is the analytical chemistry of the mixtures of chemical substances generated by living organisms. The problems of separation, molecular identification, and quantification of these mixtures are enormous. They are the ultimate molecular mishmash. The challenges they offer have demanded, and produced, many new concepts in chemical measurement science."*

For chromatographers, 1991 was a pivotal year with the introduction of comprehensive two-dimensional separation in the gas phase [2]. Liu and Phillips demonstrated this by separating a standard mixture using two capillary columns in series: a first-dimension (^1D) polyethylene glycol (PEG) column [21 m $\times$ 0.25 mm ID $\times$ 0.25 μm d_f] and a second-dimension (^2D) apolar dimethyl polysiloxane column [1 m $\times$ 0.10 mm ID $\times$ 0.10 μm d_f], coupled with a thermal modulator. The separation was completed in 2.5 minutes with a modulation period (P_M) of 2 seconds, resulting in a well-ordered retention pattern across the separation space, shaped by the relative volatility and polarity of the analytes.

Since then, despite an initial induction period marked by skepticism from some expert chromatographers who were not fully convinced of the benefits and potential of GC $\times$ GC, the technique has gradually been adopted in various application areas, either as an alternative to or a replacement for 1D-GC [3]–[5], [20]–[26]. The enhanced separation capabilities of GC $\times$ GC, which offer greater separation power, efficiency, and resolution, along with band compression in space that boosts method sensitivity, make it an ideal tool for exploring the chemical dimensions of complex samples [27]. This aligns with the integrationist approach, which seeks detailed patterns of bioactive molecules.

To reach a sufficient level of informative detail about sample composition, it is essential to properly combine multiple analytical dimensions. The following section introduces the concept of multidimensional analytical (MDA) platforms and outlines common GC $\times$ GC configurations along with their informative dimensions.

5.2.1 Analytical dimensions

MDA platforms integrate orthogonal measurement principles by combining chromatographic, electromigration, size exclusion, and/or other physicochemical discrimination methods with spectroscopic techniques such as mass spectrometry (MS), nuclear magnetic resonance (NMR), infrared (IR), and ultraviolet (UV) spectroscopy. By gathering multiplex analytical data, these platforms can effectively discriminate, identify, and quantify sample components or homologous classes, even in highly complex samples with potential interferents. The greater the "degree of orthogonality" in the data collected, the more comprehensive the approach to analyzing food samples becomes [11], [18], [21], [28], [29].

While the integrationist approach seeks patterns of chemicals rather than individual components or biomarkers, the analytical process must still provide reliable information about the identities and quantities—whether relative or absolute—of analytes. To achieve these objectives, MS with unit mass resolution or exact mass determination using high-

resolution MS (HRMS) is crucial. Molecular characteristics and functions can be elucidated through fragmentation patterns obtained via electron ionization (EI-MS) at standard 70 eV or lower energies [30], by detecting molecular ions or adducts generated through chemical ionization (CI-MS), or by employing multiple fragmentation steps in tandem MS (MS/MS or MSⁿ). However, due to the high compositional complexity of food samples, which may contain hundreds of isomeric or isobaric components, clear identification and accurate quantification are not feasible without adequate physicochemical separation, which is provided by chromatography (either 1D or GC×GC) [11], [31].

Additionally, when the sensory properties of food are of interest, or are the primary focus, odor quality assessment via olfactometry becomes a crucial dimension of the analytical system. The section on sensomics offers further insights into the role of on-line or off-line olfactometry when combined with GC×GC platforms [29], [32]–[34].

The separation power of GC×GC is a crucial feature that allows for the generation of highly detailed 2D chromatographic fingerprints with systematically organized retention patterns, opening up new possibilities for data analysis, such as group-type classification [15], [35], [36], and improving the confident identification of unknown compounds [37]. For instance, the strategic combination of column dimensions, stationary phase chemistries, and modulation dynamics to clarify the retention behavior of alkyl-substituted pyrazines in roasted coffee has been demonstrated by Ryan et al. [38] and Mondello et al. [39]. These studies were among the first to showcase the benefits of GC×GC in group-type analyses of food. The authors used a polar × apolar column combination (1D PEG × 2D 5% phenyl - 95% dimethylpolysiloxane) and achieved unambiguous identification by integrating linear retention indices (I^T) in the 1D, MS fragmentation patterns via EI, time-of-flight MS (TOF MS), and 2D retention logic.

Figure 5.2 shows the structured retention pattern of homologue series in butter volatilome. Hydrocarbons (light grey), primary alcohols (red), short chain FAs (green), linear saturated aldehydes (light blue), esters and lactones (cyan), terpenes hydrocarbons (light green) follow a retention order guided by the polarity/volatility elution criterion along the 1D separation and a further discrimination based on the degree of alkylation along the 2D.

Many fermented foods or their primary ingredients exhibit distinct profiles of homologous compounds that arise from common biosynthetic pathways, such as short-chain fatty acids (SCFAs) and primary alcohols in cocoa fermentation [40]–[43] or butter [44], or are formed through shared reaction mechanisms, like the Maillard reaction in roasted foods [45]–[47]. Compound classes such as SCFAs, aliphatic alcohols, linear saturated and unsaturated aldehydes, esters, furans, and pyrroles display rational retention patterns, which greatly aid in their identification and differentiation.

Beyond the ordered retention patterns, the use of 1D retention indices (I^T) is essential for distinguishing isomers or isobaric compounds that exhibit similar electron ionization MS (EI-MS) fragmentation patterns. Most common MS databases, such as those provided by NIST [48] and Wiley [49], include both experimental and estimated retention indices on common stationary phases, allowing MS search software to apply retention-index constraints. This approach was first introduced by Mondello co-workers [50], [51], who created one of the earliest interactive databases for flavor and fragrance analysis.

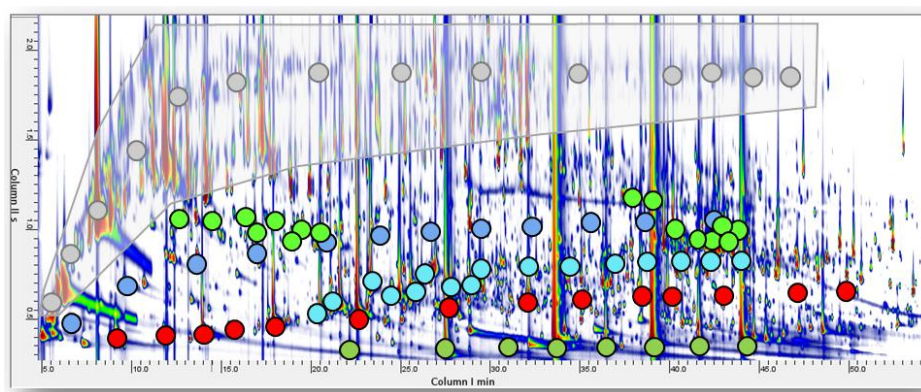


Figure 5.2: Ordered retention patterns for homologue series in butter volatilome. Colored circles highlight: hydrocarbons (light grey), primary alcohols (red), short chain FAs (green), linear saturated aldehydes (light blue), esters and lactones (cyan), terpenes hydrocarbons (light green). The analysis was run with a 1D Heavy-Wax column (30 m $\times$ 0.25 mm ID, 0.25 mm) and a 2D OV1701 column (1.0 $\times$ 0.1 mm ID, 0.1 mm). The system was equipped with a two-stage KT 2004 loop thermal modulator (Zoex Corporation, Houston, TX), cooled with liquid nitrogen and controlled by Optimode™ V.2 (SRA Instruments, Cernusco sul Naviglio, MI, Italy), with PM 2.5 s and hot-jet pulse of 0.25 s. The modulator capillary was 1.0 m $\times$ 0.10 mm ID of the 2D column. The oven program ramped from 50 °C (1 min) to 270 °C (10 min) at 4 °C/min.

More recently, Reichenbach et al. [52] developed an algorithm for calculating I^T that does not require calibration with standard *n*-alkanes or fatty acid methyl esters (FAMES). With this method, putative identifications generated through an automatic and comprehensive, untargeted MS search on the sample components use their retention times as reference points to calculate piecewise linear or logarithmic interpolations of I^T , such as Kovats indices (in isothermal analysis) or Van den Dool indices (*i.e.*, linear retention indices). **Figure 5.3A** shows a plot using a piecewise linear interpolation function of t_R versus library I^T for 21 analytes in a terpene standard mixture. The I^T values are those found in the database entry for the top-ranked analyte based on direct match factor (DMF) similarity. **Figure 5.3B** illustrates a reliable model for preliminary I^T calibration without standardization, which was achieved by removing three entries (IDs #8, #18, and #21) where analyst supervision identified inconsistencies in the incorporated I^T values [52].

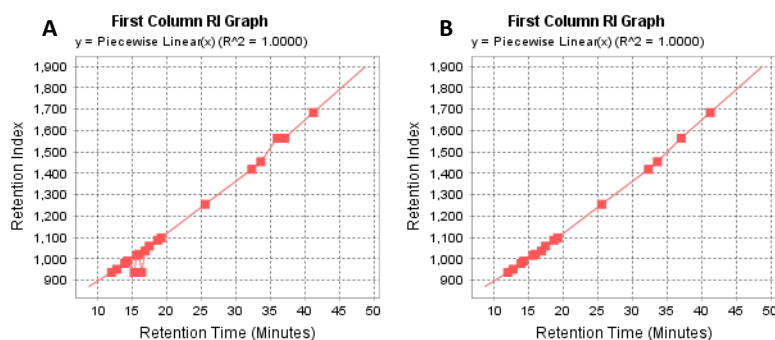


Figure 5.3: Linear retention index (I^T) calibration based on model fitting to the observed relation between t_R and the top ranked analyte resulted by direct match factor (DMF) similarity. (A) 18 of Top 21 analytes (3 #1 hits lacked library I^T); (B) 15 of Top 21 analytes consistent with piecewise linear I^T calibration model. Adapted from Reichenbach *et al.* [52].

When the high compositional complexity of processed foods complicates the clear identification of components based on retention patterns and MS spectra, specialized data processing tools can be beneficial. In cases where homologous patterns overlap, selective extraction of diagnostic MS fragments/signatures from the third analytical dimension using scripting functions allows for the effective isolation of individual components and classes, along with 2D visualization [15], [35], [53], [54]. **Figure 5.4A** presents a contour plot image of volatiles sampled by HS-SPME on dry leaves of a black tea sample in **Figure 5.4B** pink-colored circles highlight 2D peaks where the scripting function: $[\text{Relative}(43) > 0.98] \& [\text{Relative}(57) > 0.78] \& [\text{Relative}(71) > 0.68] \& [\text{Relative}(85) > 0.50]$ was verified - normal and isoalkanes. In **Figure 5.4C** green and cyano colored circles highlight 2D peaks where the scripting function: $[\text{AND}((\text{Relative}(41) > 0.70) \& (\text{Relative}(44) > 0.70) \& (\text{Relative}(57) > 0.50)), ((\text{Relative}(41) > 0.90) \& (\text{Relative}(55) > 0.70) \& (\text{Relative}(69) > 0.40) \& (\text{Relative}(83) > 0.40)))]$ was verified - saturated and unsaturated-aldehydes. CLIC™ Expression (GC-Image, LLC Lincoln NE, USA) enable easy scripting and facilitates homologues visualization in pseudo colorized images.

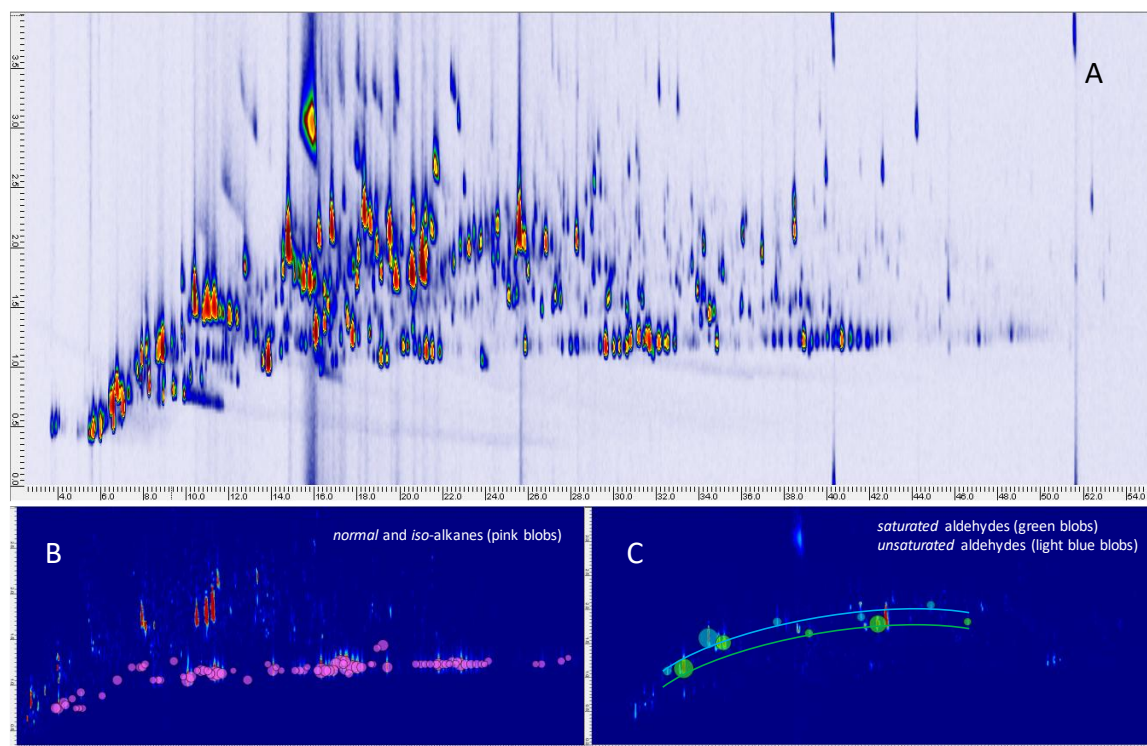


Figure 5.4. Contour plot image of volatiles sampled by HS-SPME on dry leaves of a black tea sample; (4B) pink-colored circles highlight 2D peaks where the scripting function: $[\text{Relative}(43) > 0.98] \& [\text{Relative}(57) > 0.78] \& [\text{Relative}(71) > 0.68] \& [\text{Relative}(85) > 0.50]$ is verified for normal and isoalkanes; (4C) green and cyano colored circles highlight 2D peaks where the scripting function: $[\text{AND}((\text{Relative}(41) > 0.70) \& (\text{Relative}(44) > 0.70) \& (\text{Relative}(57) > 0.50)), ((\text{Relative}(41) > 0.90) \& (\text{Relative}(55) > 0.70) \& (\text{Relative}(69) > 0.40) \& (\text{Relative}(83) > 0.40)))]$ is verified for saturated and unsaturated-aldehydes. CLIC™ Expression (GC-Image, LLC Lincoln NE, USA) enable easy scripting and facilitates homologues visualization in pseudo colorized images. The column set was by a ¹D DB-5 (30 m × 0.25 mm ID, 0.25 mm) and a ²D OV1701 (1 × 0.1 mm ID, 0.1 mm). The system was equipped with a two-stage KT 2004 loop thermal modulator (Zoex Corporation, Houston, TX) cooled with liquid nitrogen controlled by Optimode™ V.2 (SRA Instruments, Cernusco sul Naviglio, MI, Italy). P_M 5s and hot-jet pulse 0.25 s. Modulator capillary: 1 m × 0.10 mm ID of deactivated fused silica capillary. Oven program: 40 °C (1 min) to 170 °C at 2 °C/min and to 250 °C at 20 °C/min (5 min). From Magagna *et al.* [53].

Mass spectrometry (MS) is a crucial component of MDA platforms for food analysis. Beyond its role in confirming analytes identity, MS enhances specificity in the comparative analysis of extensive sample sets and supports reliable untargeted investigations [11]. Notably, MS platforms utilizing variable energy EI (e.g., Tandem Ionization™) [55]–[57], "soft" ionization through laser photo-ionization [58], or chemical ionization (CI) [59] offer valuable capabilities. By switching between two ionization energies [55] and employing a specially designed ion source, these platforms achieve efficient ionization with minimal sensitivity loss, even at low energies (e.g., 10–20 eV) [56], [60]–[64]. Applications of these techniques in food analysis include the study of volatiles and potent odorants in cocoa [56] and extra-virgin olive oil [65], as well as primary metabolite silyl derivatives from hazelnuts [57] and FAMES in biological samples [66].

Time-of-flight (TOF) analyzers are commonly used MS detection platforms in combination with GC×GC. However, only a few of these TOF systems achieve high mass resolution, mass accuracy, and rapid scan speeds. Although examples of high-resolution TOF MS (HR-TOF MS) applications in food are limited, their potential for accurate identification and database construction for complex samples not fully represented in commercial databases is significant. Efforts by Tranchida and colleagues in this area include studies of the unsaponifiable fractions of milk fat from various mammals (cow, buffalo, ewe, and goat) [67], [68] and edible oils [69]. The advantages of HR-TOF MS were particularly evident for the sterol class, where both group-type retention ordering and HR spectra facilitated reliable identification.

Conversely, more conventional single quadrupole instruments are suitable for MS detection in GC×GC applications, offering adequate frequency and sensitivity, though with some limitations in extending the mass range to capture a sufficient number of data points per peak.

5.3 Food-Omics fields and opportunities to advance existing knowledge on food composition and properties

Omics generally refers to the comprehensive characterization and quantification of biological molecules to understand the structure, function, and dynamics of organisms [70]. The first omics discipline, *i.e.* genomics, focused on studying entire genomes and was conceptually distinct from genetics, which examines individual gene variants. Today, various omics disciplines span genetics, biology, and chemistry. In the context of food, many omics fields aim to link chemical patterns and compositions with various properties, including nutritional density, sensory characteristics, biological activity, and the effects of external factors such as climate and processing practices on food. They also explore the impact of bacterial metabolism on food composition, particularly in fermented products.

By establishing a comprehensive and accurate food chemical code, a deeper understanding of several phenomena is possible: (i) identifying key quality markers in raw materials and finished products; (ii) detecting bioactive compounds in food products, extracts, and

biological fluids; (iii) finding biomarkers related to dietary interventions; and (iv) assessing human response indicators related to metabolic changes induced by different food groups and patterns [71].

Within the broader field of foodomics, defined as “*a discipline that examines food and nutrition through advanced omics technologies to enhance consumer health, well-being, and confidence*” [72], numerous specific omics disciplines have emerged. These specialized fields focus on particular interrelationships but employ similar investigation strategies, methodologies and protocols.

Among these disciplines, food metabolomics focuses on elucidating the relationships between the chemical composition of food and external factors such as crop botanical traits and origin, harvesting location, climate effects, post-harvest treatments, storage conditions, and shelf life. This emerging field highlights the pivotal role of GC×GC [73]–[76], which enhances the identification of chemical compounds by providing more comprehensive candidate lists with greater intrinsic information.

Closely related to food metabolomics is nutrimetabolomics, which merges nutritional science with metabolomics. This field explores “*the effects of specific diets, foods, nutrients, microorganisms, or bioactive compounds on the human metabolome*” [77], aiming to deepen our understanding of how dietary patterns impact human health and how the food metabolome (i.e., “*the portion of the human metabolome derived from the digestion and transformation of foods and their components*” [78]) interacts with human physiology. It examines both (i) nutrients that influence metabolic responses in humans and (ii) non-nutrients that may interact with metabolic pathways and receptors, affecting human metabolism.

When investigating food's hedonic properties, sensomics seeks to “*molecularize flavor entities of nature*,” creating a unique odor and taste profile that aids in food recognition and enhances consumer enjoyment [79]. Sensomics and related fields like flavoromics [80] establish a foundation for the objective and rational assessment of food based on chemoreceptive events in the nose and mouth.

When all volatile metabolites including also degradation products and exogenous compounds not strictly derived from metabolic processes, such as environmental contaminants and substances produced by bacteria and molds, are targeted, food volatilomics is delineated. It focuses on the volatilome [81], or volatome [82], and by exploring its multiple chemical dimensions more comprehensive insights into various phenomena compared to the metabolome alone are provided [83]–[85].

Although food safety and authenticity have not yet been fully integrated into an “omics” framework, they remain crucial aspects of food analysis. Food safety is a mandatory requirement for all products intended for human consumption. This review includes food safety due to GC×GC's potential in safety assessments and its ability to suggest measures to enhance food chain safety. Notable examples of food authentication and *identitation* are discussed in the referenced literature [86]–[88].

5.3.1 Approaches of investigation and terminology

In metabolomics, established investigation approaches include *profiling* and *fingerprinting* [89], [90]. Both strategies reveal compositional differences between samples, but they vary in scope. Profiling involves optimizing the analytical process to focus on known components of samples, following a reductionist approach. This method collects detailed data on analyte retention, mass spectra, and detector responses to enable qualitative and/or quantitative comparisons. Targeted profiling refers to this approach when it is directed toward a priori defined analytes (*i.e.*, known components) [91]. Conversely, untargeted profiling (also referred to as non-targeted profiling) occurs when the analytical process generates distinctive features for all components independently by the chance to assign a correct identification, allowing a comprehensive evaluation of the sample's constituents [92], [93]. This approach could be also referred to as untargeted or non-targeted profiling.

Fingerprinting, as described by Fiehn [90] for metabolomics, differs conceptually from profiling. It is a high-throughput approach designed for effective cross-comparative analysis of samples, but it does not necessarily achieve the one-feature/one-component resolution typical of profiling. Instead, fingerprinting aims to capture a comprehensive set of features that ideally represent the chemical constituents of a sample. While profiling's information level—both qualitative and quantitative—is influenced by the analytical dimensions of the system, fingerprinting provides a broader overview.

Certain spectroscopic techniques, such as nuclear magnetic resonance (NMR) and near-infrared spectroscopy (NIR), can deliver high levels of detail, though they may miss signals from trace and ultra-trace components. Recent advancements in MDA platforms have led to more stable and informative systems that enhance the concept of fingerprinting [12], [15].

2D peak patterns generated by techniques like GC×GC or LC×LC can be likened to human fingerprints, where *minutiae features* (*e.g.*, ridge endings and bifurcations) are used for individual identification. In chromatographic fingerprinting, minutiae within 2D peak patterns are detected, aligned, and compared across multiple samples. Depending on the system's dimensionality, these 2D chromatographic features collect multidimensional metadata, including retention times in both chromatographic dimensions (^1t_R , ^2t_R), detector responses, MS spectral signatures, and additional data related to olfactory quality or parallel detection. Chromatographic fingerprinting has the potential to unlock deeper insights within MDA datasets. Recent reviews have discussed its fundamental concepts and tools [12], [13], [15], [17].

The following section highlights selected research in food-omics and related fields that showcases how GC×GC enhances understanding of complex phenomena, provides more accurate and informative results, or offers new investigative perspectives on complex samples.

5.4 Applications of GC×GC in the food-omic domains

5.4.1 Food-metabolomics and volatilomics

GC×GC-based food metabolomics can focus on volatile organic compounds (VOCs), with volatilomics [94]. Through precise profiling and fingerprinting of VOCs, it is possible to gain deeper insights into food quality and spoilage, the impact of climate change on plant phenotypes, and the effects of pedoclimatic variations and geographical origin [34], [42], [43], [46], [76], [85], [95]–[101]. Similarly, the lipid fraction, analyzed through the profiling/fingerprinting of fatty acids methyl esters (FAMES), can provide valuable diagnostic information regarding food quality and origin [102]–[105]. For plant-based foods, primary metabolites—such as amino acids, sugars, low molecular weight acids, and amines—offer insights into the quality of primary materials [57], [75], [106], [107], harvesting practices [73], [75], [108], storage conditions [73], [75], [108], and the effects of processing and environmental stressors on crops [34], [42], [43], [46], [95], [96], [98]–[101].

Mack et al. [106] explored the effects of postharvest ripening on kiwifruit using untargeted metabolomics. Their study tracked metabolic changes across six stages of ripening and two non-marketable stages characterized by significant water loss. Notable variations were observed in sugars, sugar-related substances, and organic acids throughout the ripening stages, with sugars increasing and organic acids decreasing. The study's multiplatform approach—combining GC×GC-qMS, GC-MS, and NMR—allowed for cross-validation of results. GC×GC-qMS proved to be the most sensitive platform, detecting the majority of primary metabolites in kiwifruit. Targeted GC-MS offered better profiling of the most concentrated analytes, such as main sugars and some organic acids, while NMR, despite being less sensitive, demonstrated superior reproducibility for certain analytes. Overall, the metabolome coverage was robust, with approximately 160 analytes reproducibly measured by GC×GC-qMS and GC-MS. Around 45% of these metabolites were identified using an in-house spectral database, and for unknown metabolites, chemical classes were hypothesized based on MS fragmentation patterns and their relative retention in the 2D space.

Using an untargeted multi-platform approach, the authors thoroughly investigated the kiwifruit metabolome's evolution through ripening and over-ripening stages. The combination of three complementary platforms allowed for a broader dynamic range of metabolite concentrations. Significant changes were observed in sugars and related compounds, which increased during ripening, and in organic acids, which decreased as over-ripening progressed. Other metabolites, like methanol and xylose, peaked during the mid-ripening stages, while galacturonic acid was found only in overripe fruits. This study underscores the utility of GC×GC-qMS as a powerful exploratory tool for untargeted food metabolomics.

More recently Squara *et al.* combined different MDA platforms to collect information from volatile and non-volatile metabolites to highlight quality traits and marker patterns in premium quality hazelnuts [76], [85]. The research examined the compositional complexity of the *Corylus avellana* L. metabolome by employing both untargeted and targeted metabolomic profiling methods to predict industrial quality. Using GC×GC and LC combined with high-resolution MS, the study characterized the non-volatile (primary and specialized

metabolites) and volatile components of hazelnuts. To improve classification performance, data fusion techniques, such as low-level (LLDF) and mid-level data fusion (MLDF), were utilized. Through principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA), it demonstrated that geographical origin and postharvest handling significantly influence the specialized metabolome, while storage conditions and duration predominantly affected the volatile metabolome. The findings indicated that MLDF strategies, especially supervised MLDF, offer superior predictive accuracy compared to single-fraction analyses. Notably, the study identified key metabolite patterns, such as aldehydes, alcohols, terpenes, and phenolic compounds, that were closely linked to quality attributes of hazelnuts. The integration of multiple analytical platforms and data fusion approaches holds significant potential for enhancing quality evaluation and optimizing storage and processing methods in the food industry.

In the context of extreme climate events such as drought and temperature fluctuations, GC×GC-TOF MS has proven to be a valuable addition to techniques targeting non-volatile specialized plant metabolites [109], [110]. Stilo *et al.* explored the primary metabolome of tea (*Camellia sinensis* L. Kuntze) to understand how plants adapt to varying climatic conditions, including drought, heavy rainfall, and temperature changes due to elevation [111]. By adapting an extraction and derivatization protocol used in biological fluids metabolomics, they achieved extensive coverage of tea's primary metabolites [112]. Out of 780 detected peak features across over 40 samples, 74 metabolites were tentatively identified based on available metadata and aligned with identification criteria shared within the metabolomics community [113], [114]. Seasonal (spring vs. summer) and altitude-related variations were observed, with notable changes in amino acids, organic acids, polyalcohols, sugars, and some specialized metabolites such as flavan-3-ols and caffeine.

The study highlighted that during spring, the levels of alanine, aspartic acid, glycine, threonine, valine, phenylalanine, phosphoric acid, xylonic acid, and xylitol increased in tea plants. When comparing samples from two harvest years, fructose, glucose, maltose, and arabinose showed consistent up-regulation in spring teas. Altitude-related temperature variations led to increased levels of alanine, isoleucine, tyrosine, catechin, gallic acid, glycolic acid, malic acid, and ribonic acid in low-elevation teas.

The research, built on previous studies of unprocessed tea focusing on volatiles, specialized metabolites, and metals [109], [110], [115], provided deeper insights into the complex interactions between tea plants and climatic factors. The authors aimed to identify key primary metabolites and biomarkers to better understand the feedback loops between human activities and natural systems, particularly in addressing climate impacts on tea plants. The comprehensive data obtained through GC×GC-TOF MS, including known and unknown features, offers new avenues for food research. The extensive metadata—such as retention times in two dimensions, MS fragmentation patterns, and both absolute and relative responses—enables further exploration for specific discrimination and compound identification through EI-MS spectra, ¹D retention indices, and retention pattern logic [11].

Accurate profiling and fingerprinting of fatty acids (FAs) represents another area where GC×GC-MS has provided a new dimension to research [116], [117]. Although the chemical

diversity within lipid fractions is not exceedingly high, the presence of numerous FA isomers with closely similar retention characteristics necessitates high separation power and selectivity for thorough analysis. GC×GC offers both high separation capability and differential selectivity across two chromatographic dimensions, creating unique 2D patterns that assist in identifying and differentiating components lacking reference standards (such as odd-carbon-number or branched FAs). Tranchida and Mondello have focused their research on profiling and fingerprinting FAs from various sources, including plants, algae, and animals [102], [103].

Homologous FAs series within the C_4 - C_{24} range can be converted into fatty acid methyl esters (FAMES) and analyzed using different column combinations. **Figure 5.5A** illustrates the 2D pattern of human plasma FAMES analyzed with an apolar × polar column combination. In this pattern, analytes align along diagonal bands based on carbon number (CN) within the C_{14} to C_{24} range (indicated by black arrows) and are grouped into 2D sub-patterns according to the number of double bonds (DB). Seven DB bands are marked and aligned across the 2D space in **Figure 5.5A**. Additional retention characteristics, such as the relative position of DBs, form diagonal retention bands. The ω -derivatives of $C_{20:5\omega3}$, $C_{20:4\omega3}$, and $C_{20:3\omega3}$ are shown in the magnified section of the chromatogram in **Figure 5.5B**. In **Figures 5.5A** and **5.5B**, 36 FAMES were identified using reference standards, while the remaining 29 components, for which no standards were available, were characterized based on the highly ordered nature of the 2D pattern [118].

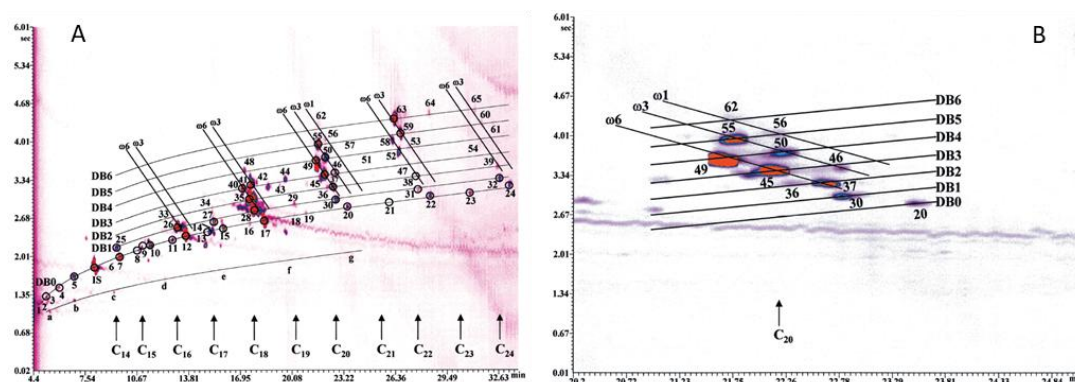


Figure 5.5: (A) 2D pattern of human plasma FAMES within the C_4 - C_{24} range. Analytes are aligned along seven diagonal bands, based on carbon number (CN) within C_{14} and C_{24} (zones indicated by black arrows), and grouped while forming 2D sub-patterns ordered based on double bonds (DB) number. (B) Enlarged area showing the C_{20} FAMES group. The analyses were performed by a column set consisting of a ^1D Equity-1 (30 m × 0.25 mm ID, 0.25 mm) and a ^2D Supelcowax-10 (0.95 × 0.1 mm ID, 0.1 mm). The GC system was equipped with an LMCS Everest longitudinally modulated cryogenic system (LMCS; Chromatography Concepts, Doncaster, Australia), with a mechanical stepper motor drive for movement of the cryotrap operating with liquid CO_2 as cryogenics. Oven program: 180°C to 280°C (10 min), at 3°C/min, P_M 6s. Adapted from Tranchida *et al.* [118].

5.4.2 Nutrimetabolomics

When integrating knowledge of food composition with nutritional science, the field of nutrimetabolomics becomes increasingly relevant, highlighting a promising path toward advanced nutritional care, targeted dietary treatments, and personalized nutrition [77]. This area of research faces significant challenges, particularly due to the high chemical complexity of biological fluids (such as urine, plasma, and saliva) and the difficulty in generating highly ordered patterns in presence of many overlapping chemical dimensions. Furthermore, the vast concentration differences between various analytes and analyte classes can lead to system contamination, reduced method precision, and lower long-term reproducibility [119]–[121] .

Consequently, much of the effort in GC×GC-MS-based nutrimetabolomics focuses on achieving optimal separation power in both chromatographic dimensions, ensuring sufficient column loadability and appropriate injection techniques to fully resolve detectable components across the dynamic concentration range, thereby enabling highly reliable absolute or relative quantitation. Recently, an interdisciplinary team of researchers from the FoodBALL Consortium, under the Joint Programming Initiative "A Healthy Diet for a Healthy Life" (JPI-HDHL) [77], reviewed nutrimetabolomics strategies, suggesting best practices and standardized protocols to achieve consistent, high-quality results.

In this complex research context, GC×GC has been effectively employed by Mack *et al.* [122] to identify robust biomarkers for coffee intake. The authors analyzed the volatile components in human urine collected over 24 hours from participants in the cross-sectional Karlsruhe Metabolomics and Nutrition (KarMeN) study, conducted at the Max Rubner-Institut in Karlsruhe, Germany, from 2011 to 2013. The samples underwent headspace solid-phase microextraction (HS-SPME) followed by GC×GC-qMS and a combined untargeted/targeted fingerprinting approach using pattern recognition algorithms. In total, 138 volatiles were reliably detected, with several showing strong positive correlations to coffee intake. Six key volatiles were identified as the most informative: 2-methyl-furan, guaiacol, 2-/3-methyl-butanoic acid, 2-vinylfuran, and 3,4-dimethyl-2,5-furandione, the latter emerging as the most promising robust biomarker.

A crucial aspect of this study was evaluating the most effective normalization method to account for variations in urine concentration, which provided valuable insights specific to the urine volatilome. The combined untargeted/targeted investigation workflow is depicted in **Figure 5.6**, showcasing the 2D raw data pre-processing steps, including batch correction using quality standards, normalization based on parameters such as creatinine content, osmolarity, urine volume, mass spectral total useful signal (MSTUS), and probabilistic quotient normalization (PQN), followed by data analysis to identify potential biomarkers through correlation and statistical significance assessments. Further details on the normalization methods and their impact on variation coefficients (CVs%) are discussed in the reference [122].

Another noteworthy study explored the metabolic transformation of dietary polyphenols by the colonic microbiota [123]. The researchers conducted a comparative analysis of *in vitro* microbial phenolic metabolite profiles from selected foods and beverages against those found in the urine of healthy individuals who participated in an 8-week dietary intervention

involving the same food items. The study employed GC×GC-TOF MS in a targeted mode to identify known polyphenolic compounds. Green tea, which was the richest in polyphenols among the tested foods, exhibited significant derivatives of its primary metabolites (phenolic acids), characterized by modifications to the carbon backbone and extensive oxidation (primarily hydroxylation). Among these metabolites, 3,4-dihydroxybenzoic acid; 3,4-dihydroxyphenylacetic acid; 3,4-dihydroxyphenylpropionic acid; 3-hydroxybenzoic acid; 3-hydroxyphenylacetic acid; 3-hydroxyphenylpropionic acid; 4-coumaric acid; 4-methylcatechol; 4-hydroxybenzoic acid; benzoic acid; gallic acid; ferulic acid; sinapic acid; enterodiol; and vanillic acid demonstrated a significant, positive correlation between urinary excretion and in vitro colon model results.

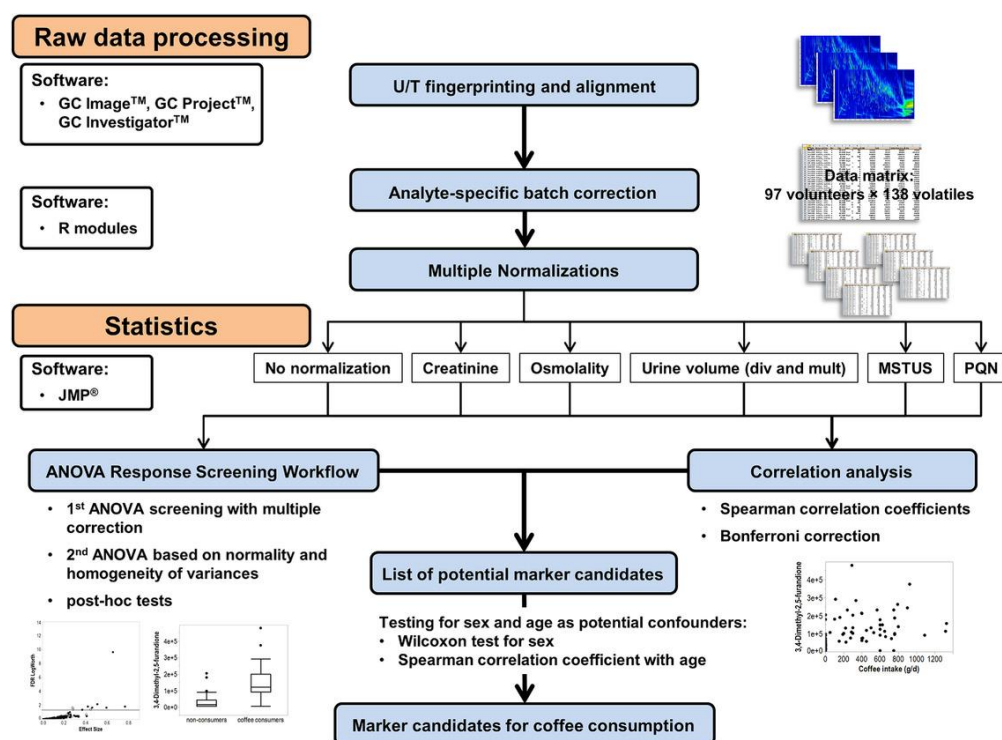


Figure 5.6. Combined untargeted and targeted fingerprinting investigation workflow, including 2D raw data pre-processing step (i.e., batch correction by quality standards and normalization based on different parameters: creatinine content, osmolality, urine volume, mass spectral total useful signal - MSTUS - and probabilistic quotient normalization - PQN) followed by data mining to reveal potential biomarkers through combined evaluation of correlation and statistical meaningfulness. From Mack *et al.* [122].

Additionally, the study found considerably higher levels of hippuric acid, 3-hydroxybenzoic acid, and ferulic acid in the urine of healthy subjects compared to those in the colon model, while the hepatic conversion model accounted for residual phenol-derived metabolites. This research provided valuable insights into the bioavailability of dietary polyphenols and clarified their connections to urinary metabolite patterns.

An untargeted investigation of urinary metabolite profiles provided deeper insights into the effects of solid vs. liquid fructose formulations in animal models [64], [124]. Bressanello *et al.* [64] developed an analytical system using a one-dimensional column with a dual parallel secondary column setup, coupled to both MS and flame ionization detection (FID) (GC×2GC-MS/FID). This approach was designed to detect changes in the metabolic fingerprints of mice urine samples. The study involved mice on either a control diet or a

fructose-enriched diet provided in liquid or solid form, with metabolite profiles tracked over a 12-week period for both individual subjects and sub-groups.

The combined untargeted and targeted fingerprinting (UT fingerprinting) allowed for the coherent clustering of mice according to their dietary regimen [44], [63], [125]. Urinary metabolite profiles from mice given high doses of liquid fructose exhibited significant changes in concentrations of fructose, glucose, citric, pyruvic, malic, malonic, gluconic, cis-aconitic, succinic, and 2-ketoglutaric acids, along with glycine acyl derivatives (N-carboxyglycine, N-butrylglycine, N-isovaleroylglycine, N-phenylacetylglycine) and hippuric acid. Post-targeting analysis revealed additional metabolites not initially identified, including N-acetyl glucosamine, N-acetyl glutamine, malonyl glycine, methyl malonyl glycine, and glutaric acid, with quantitative GC-MS confirming the significance of several potential biomarkers through precise quantification.

Visual feature fingerprinting, using comparative visualization of urine signatures at different time points, demonstrated individual metabolic variations and highlighted the potential for data processing to support personalized strategies. **Figures 5.7A** and **5.7B** display pseudocolor images of mouse #40's urine after 6 (**7A**) and 12 weeks (**7B**) of liquid fructose supplementation, where red graphics indicate features identified by the untargeted/targeted workflow, and green graphics represent column bleeding or interferences, which were excluded from the analysis. **Figure 5.7C** presents a comparative visualization, color-coded to emphasize differences between the urine signatures at 12 weeks (reference pattern) and 6 weeks (analyzed pattern), with yellow areas indicating targeted analytes with notable response differences. Comparative visualization realized by UT fingerprinting based on image pattern recognition can be classified within the Artificial Intelligence tools as an augmented visualization approach [12], [44].

5.4.3 Sensomics and Artificial Intelligence smelling

Sensomics is focused on deciphering the complex combination of aroma and taste-active molecules that interact with human chemosensory receptors and are processed by the brain to produce flavor perceptions [126]. Aroma alone accounts for approximately 80% of the hedonic experience of food [79]. This is primarily influenced by volatile compounds present in trace amounts (mg/kg or µg/kg) that typically have low solubility in water, low polarity, and molecular weights under 300 atomic mass units (amu). Odorants interact with Odorant Receptors (ORs) in the olfactory epithelium, generating complex signal patterns, or the "Receptor Code," which the nervous system integrates to create olfactory perceptions [79], [127]–[129].

A crucial aspect of understanding the "aroma blueprint" or the chemical composition of food odors [79], [130] involves thoroughly characterizing all potential ligands and interfering/modulating compounds that form the Chemical Odor Code. This process follows a structured workflow that includes isolating, extracting, and concentrating volatile compounds; chromatographically separating them while using olfactometric screening to identify odor-active components; accurately quantifying them; and validating their aroma contributions through recombination and omission tests [79], [126], [131].

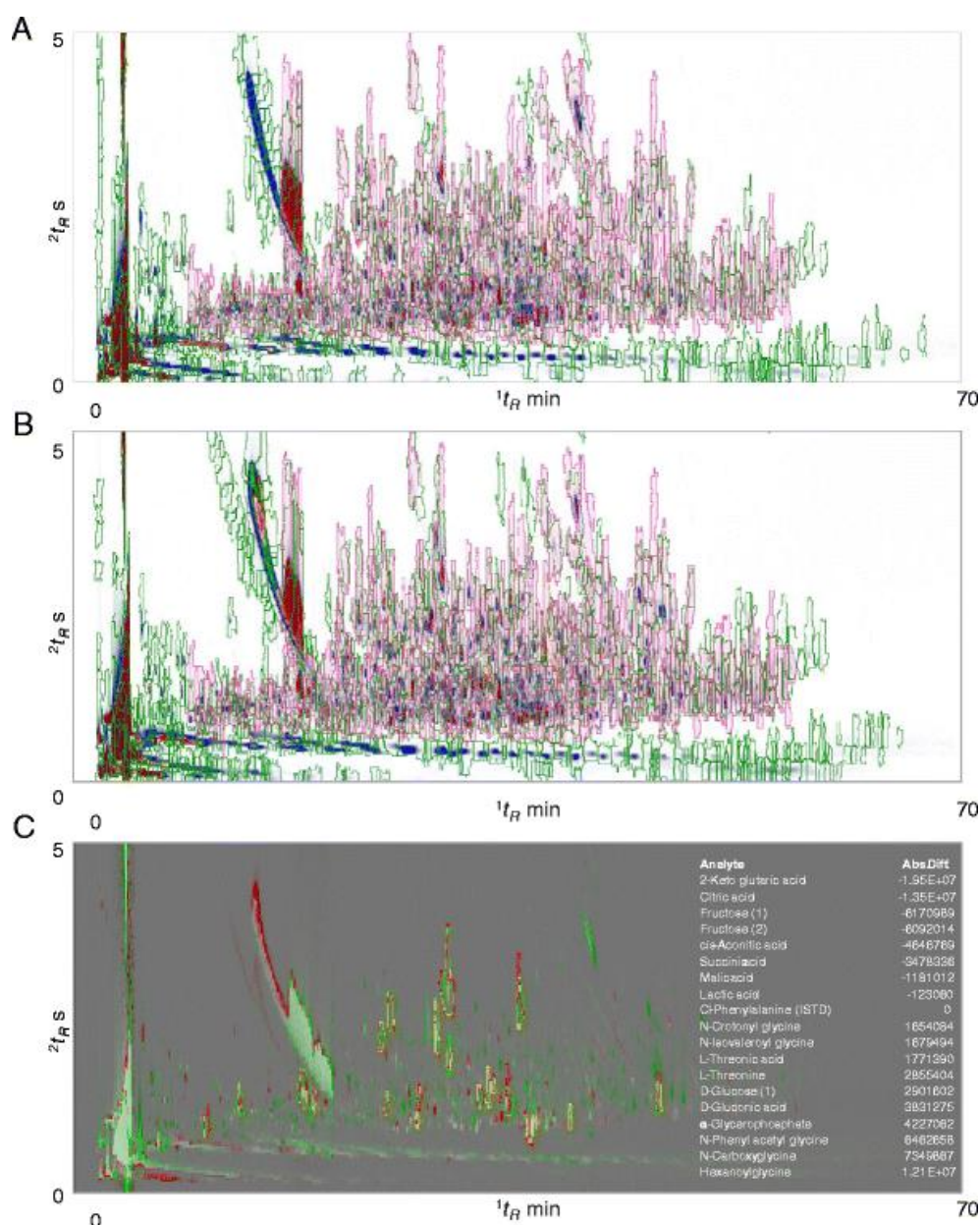


Figure 5.7. Visual features fingerprinting (MS signal full scan acquisition) was performed by comparing urine sample signatures at different time points during the diet intervention. Red graphics indicate untargeted or targeted features, while green graphics represent column bleeding or interferences that were excluded from the analysis. (7A) Pseudocolor image of urine samples from mouse #40 after 6 weeks of diet intervention with liquid fructose supplementation, showing the analyzed pattern. (7B) Pseudocolor image of urine samples from mouse #40 after 12 weeks of diet intervention with liquid fructose supplementation, depicting the reference pattern. (7C) Comparative visualization with colorization to emphasize relative differences between patterns; yellow areas on the 2D space denote targeted analytes, with response differences listed. The analysis was run with a ¹D SE52 column (30 m × 0.25 mm ID, 0.25 mm) and a ²D OV1701 column (1.4 × 0.1 mm ID, 0.1 mm). The system was equipped with a two-stage KT 2004 loop thermal modulator (Zoex Corporation, Houston, TX), cooled with liquid nitrogen and controlled by Optimode™ V.2 (SRA Instruments, Cernusco sul Naviglio, MI, Italy), with P_M 5s and hot-jet pulse of 0.35 s. The modulator capillary was 0.6 m × 0.10 mm ID of the 2D column. The oven program ramped from 60 °C (1 min) to 300 °C (10 min) at 4 °C/min. From Bressanello *et al.* [64].

GC×GC platforms are well-suited for implementing the tools needed in sensomics. Significant progress in developing multi-dimensional chromatographic platforms has been made by Marriott and collaborators [132], [133]. These systems can switch between

comprehensive two-dimensional gas chromatography (GC×GC) and targeted multidimensional gas chromatography (MDGC) to achieve complete chromatographic resolution, particularly in complex regions where key odorants are present [134]. Additionally, Chin *et al.* [135] designed a platform that incorporates an orthogonal dimension of olfactometry to study aroma-active compounds in beverages like coffee and wine. This system combines gas chromatography-olfactometry (GC-O) with GC×GC, and uses a combination of time-of-flight mass spectrometry (TOF MS), flame ionization detection (FID), and flame photometric detection (FPD) in sulfur mode. This setup allows for the screening, separation, and identification of odor-active compounds effectively.

Conversely, off-line GC-O screenings present certain advantages. They can guide or refine fingerprinting toward regions in the 2D chromatogram that contain odor-active compounds, enhancing machine learning (ML) capabilities to discriminate samples based on characteristics likely linked to specific olfactory attributes. A case study by Gabetti *et al.* illustrates this approach, as they examined the distinctive aroma profiles of Piemonte peppermint essential oils using GC-O through aroma extract dilution analysis (AEDA), alongside GC×GC-TOF MS and variable ionization energy (Tandem Ionization™) [29]. By concentrating on chromatographic areas associated with strong creamy and sweet notes, detectable even at high dilution factors, they were able to enhance the performance of PLS-DA classification models. The authors suggested that a GC×GC-based AI-smelling tool could potentially make sensory-driven decisions without relying on a human sensory panel.

Beyond the multi-multidimensional GC×GC platforms, which play a crucial role in flavor research, conventional systems have now advanced enough to serve as essential tools in sensomics and flavoromics [136]. Nicolotti *et al.* [33] developed a rapid, sensomics-based expert system called SEBES, designed to predict key aroma compounds in foods without relying on all the individual steps of the traditional workflow [79], [126], [131]. Their method isolates and quantifies approximately 100 of the 226 known key food odorants (KFOs) identified by Dunkel *et al.* [79]. Using specialized data processing that includes odor thresholds (OTs) information, odor activity values (OAVs) were calculated by comparing concentrations to their respective OTs, thereby encoding the sample's combinatorial odor profile. The authors emphasized that GC×GC was vital for extending the method's dynamic range, enabling precise quantification of analytes present at vastly different concentrations in real samples. The system's efficiency and resolution also allowed for direct injection of crude extracts without additional sample preparation. The SEBES system was validated against traditional sensomics methods, demonstrating a high accuracy with a maximum quantification error of ±20%. The authors concluded that “*it is possible to characterize key food odorants with one single analytical platform and without using the human olfactory system, that is, by artificial intelligence smelling.*”

Numerous other studies have highlighted the key role of GC×GC in flavor analysis. Notably, several have focused on the direct analysis of volatile fractions using headspace (HS) approaches [40], [41], [46], [54], [63], [65], [99], [130], [137], [138]. An emerging area of application involves examining volatile compounds in extra-virgin olive (EVO) oil, where some odorants contribute to desirable attributes or aroma defects [63], [95], [100], [101]. Stilo *et al.* [65] investigated the complex volatile profile of EVO oil from Picual olives harvested at

different ripening stages. They employed HS-SPME combined with GC×GC-TOF MS, utilizing both hard and soft ionization techniques (tandem ionization), identifying around 130 compounds out of over 450 detectable volatiles. For many of the target analytes, which play a significant role in the aroma profile of EVO oils [130], [139], [140], linearity in headspace sampling was achieved [141], [142], allowing for accurate quantification in a fully automated process. The use of tandem ionization TOF MS also enhanced the method's dynamic range and response linearity, thereby increasing its analytical capacity and informational value.

Recently, the concept of an AI-based smelling machine has been incorporated into a decision-making tool for the confectionery industry [85]. This approach, utilizing automated MHS-SPME with GC×GC-MS/FID, was used to precisely quantify key aroma compounds, as well as markers of rancidity and spoilage, in premium hazelnuts (*C. avellana* L.) over a 12-month period. Accurate quantification enabled the assessment of chemical quality traits for incoming materials, with results being consistent over time and across different laboratories. This approach facilitated objective decision-making and reliable predictions of sensory attributes (*i.e.*, AI-driven sensomics), geographical origin, and storage duration of samples.

5.4.4 Food safety

Mineral oil hydrocarbons (MOHs) are complex mixtures composed of various aliphatic hydrocarbons, including linear alkanes, branched isoalkanes, and cyclic compounds like cycloalkanes or naphthenes, collectively known as mineral oil saturated hydrocarbons (MOSHs). They also contain aromatic hydrocarbons (MOAHs), characterized by aromatic structures with one or more benzene rings and extensive alkyl substitutions. Some MOAH components are recognized for their mutagenic, toxic, and tumor-promoting properties.

The presence of MOSH and MOAH in raw food materials, semi-processed, and finished products is often linked to contamination from external sources. Identifying these contamination sources is crucial for implementing safety measures throughout the food supply chain. However, accurately quantifying and characterizing MOH fractions presents significant challenges for analytical chemistry due to numerous interfering substances found in food. This necessitates the use of robust confirmatory methods to ensure reliability [143]–[145].

Currently, the most advanced methods for quantifying MOSH and MOAH in food involve LC either offline or online, combined with GC-FID [146]. However, the existence of hundreds to thousands of isomers results in an unresolved complex mixture (UCM) in GC, and the lack of a detection method with adequate specificity can compromise accurate quantitation even when using LC-GC [147], [148]. High-resolution MS (HRMS) is valuable but not sufficient alone to distinctly separate exogenous MOSHs and MOAHs as mandated by EU Commission Decision 657/2002 [149] and recent updates [150].

In this context, the advantages of comprehensive two-dimensional gas chromatography (GC×GC) became clear. Biedermann and Grob introduced chromatographic fingerprinting of MOHs using a medium polar × apolar column combination to differentiate compound classes based on retention patterns governed by their relative polarity and volatility. This approach also facilitates the identification of contamination sources through pattern recognition techniques [151]. **Figure 5.8** illustrates the comparison between

chromatographic profiles obtained by the standard LC-GC-FID method (left side) for MOSH and MOAH fractions and the corresponding GC×GC-TOF MS pseudocolor images of Asian rice suspected of MOH contamination.

The fingerprinting capabilities of GC×GC are demonstrated by the well-defined pattern of *n*-alkanes, ranging from *n*-C₁₃ (an internal standard) to *n*-C₂₅, in uncontaminated rice samples. Branched paraffins, which show longer retention times in the second dimension, elute along the same line as *n*-alkanes, while oligomeric polyolefins (POSH), particularly C₂₁ oligomers likely derived from polypropylene (PP) packaging, appear as a distinct cluster. Monounsaturated PP oligomers are identified between the *n*-alkanes and the POSH cluster. Additionally, slanted elution bands with shorter retention in the second dimension correspond to naphthenes and cyclic hydrocarbons, such as dodecyl cyclopentanes and undecyl cyclohexanes.

The lower section of **Figure 5.8** displays both the 1D and 2D profiles of the MOAH fraction. While the LC-GC method results in an unresolved hump, GC×GC provides more structured elution patterns, clearly identifying various groups, including diisopropyl naphthalenes (DIPN), which may indicate contamination from recycled paperboard.

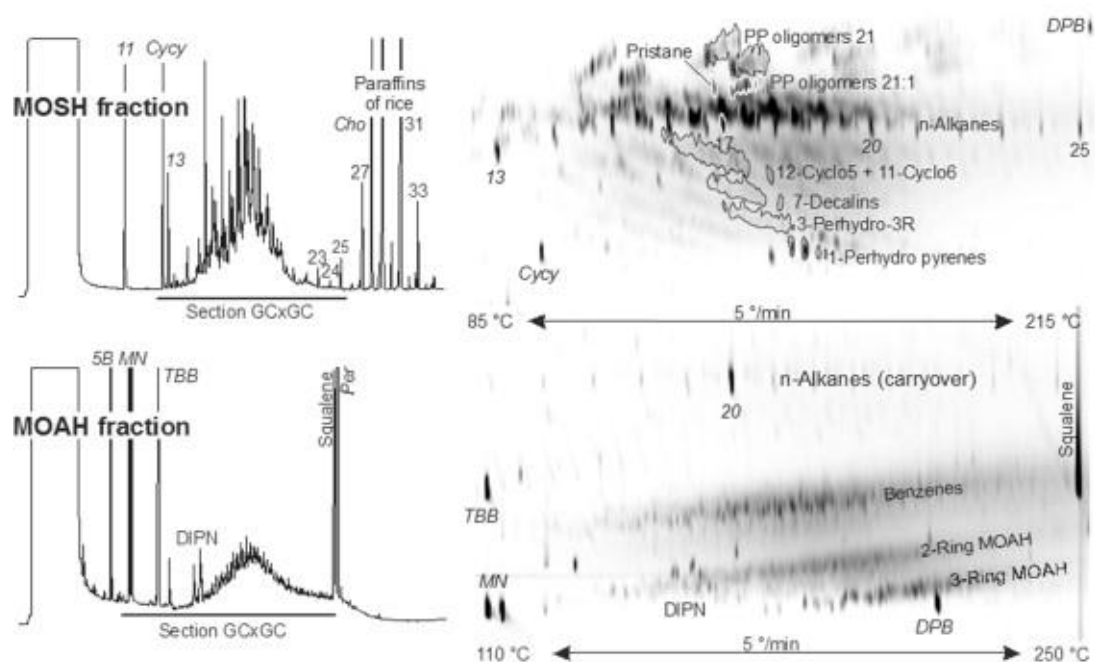


Figure 5.8. LC-GC-FID chromatograms (left side) of MOSH and MOAH fractions and the corresponding GC×GC-TOF MS contour plots (right side) from Asian rice samples. Compounds assignment are: (11) for *n*-undecane, (Cycy) for cyclohexyl cyclohexane, (Cho) for cholestane, (13) for *n*-tridecane, (20) for *n*-eicosane (DPB) for diphenoxy benzene, (5B) for *n*-pentyl benzene, (MN) for 1- and 2-methyl naphthalene, (TBB) for tri tert. butyl benzene, (Per) for perylene. The column set was: ¹D column OV-17 (15 m × 0.25 mm × 0.15 μm) and ²D column PS-255 (2.5 m × 0.15 mm × 0.055 μm). Thermal modulation by loop-type modulator. Oven program: 70 °C (3 min) to 310 °C, at 5.0 °C/min; P_M 6s. Acronyms are explicated in the text. From Biedermann and Grob [151].

GC×GC-TOF MS plays a vital role in the precise quantification of MOSH and MOAH, offering opportunities for effective fingerprinting and contamination source identification. This technique could revolutionize the field by providing multidimensional solutions that integrate online LC pre-fractionation with GC×GC and parallel detection using FID/MS [152]. Such an approach, which also supports untargeted investigations, could simultaneously detect non-intentionally added substances (NIAS) [145].

Looking ahead, GC×GC shows significant promise in the context of food safety for the monitoring of chlorinated paraffins (CPs), which are synthetic chlorinated n-alkanes used extensively in industry as lubricants, paint components, coating agents, and sealants. Depending on the length of their carbon chains, CPs are classified into short-chain chlorinated paraffins (SCCPs; C₁₀-C₁₃), medium-chain chlorinated paraffins (MCCPs; C₁₄-C₁₇), and long-chain chlorinated paraffins (LCCPs; C₁₈-C₃₀) [153]. Due to their potential toxicity, persistence, and bioaccumulation, SCCPs and MCCPs have attracted recent scrutiny from the European Food Safety Authority (EFSA) [154].

GC×GC coupled with low-resolution TOF MS has been effectively employed for the accurate quantification of SCCPs even in the absence of specific calibration standards for individual components. Researchers [154] created a calibration method linking the total response factor (RF) to chlorine (Cl) content based on peak volumes of quantifier ions for 24 congener groups. Results from both CP products and urban air samples showed significant variation among CP materials, with seasonal changes in urban air levels primarily driven by temperature fluctuations. The highest SCCP concentrations were found in summer, with lighter congeners (such as the C₁₀ and C₁₆ groups) more prevalent in the gas phase, and heavier ones (like the C₁₃ and C₁₇-C₁₈ groups) found mainly in the particle phase.

These advances, which combine accurate assessments with the identification of contamination sources through informative fingerprinting using GC×GC-TOF MS, could address the data gaps highlighted in the EFSA report [154]. The report noted limited data on SCCP and MCCP levels in certain fish species and suggested that dietary exposure may be underestimated due to the lack of comprehensive data on CPs in other foods. Systematic mapping of potential environmental contamination sources affecting the food chain could provide the robust data needed for a thorough risk assessment, enhancing food safety.

5.5 The greenness potential of GC×GC when applied to food analysis

GC×GC offers notable greenness potential in food analysis when evaluated through the AGREE (Analytical Greenness) principles and metrics [155]. Its green profile is influenced by solvent and reagent use, energy consumption, and waste generation when analyzed in light of the information potential it provides. While GC×GC requires solvents for sample preparation, miniaturized techniques like (HS)-SPME or direct on-column sample injection can minimize, or even eliminate solvent volumes. The technique itself generates minimal waste, and adopting greener solvents or solvent-free preparation methods further enhances its sustainability. The use of non-hazardous or less toxic materials contributes positively to its overall environmental impact.

Energy consumption is a consideration for GC×GC due to the additional dimension of separation and the sophisticated instrumentation required. Although it might involve higher energy use compared to traditional GC methods, especially when thermal/cryogenic modulators are used, its efficiency in providing enhanced separation power can reduce the need for repeated analyses, lowering overall energy usage. The ability of GC×GC to analyze complex samples with fewer re-analyses offsets the energy demands associated with the technique.

GC×GC's capability to work with smaller sample volumes and its efficiency in providing comprehensive analytical data contribute to its greenness. When combined with advancements in technology, such as green solvents and effective waste management practices, GC×GC stands out as a valuable, sustainable analytical tool in the context of food quality and safety analysis.

5.6 Conclusions

Food-omics research demands a thorough and holistic approach to effectively capture the compositional complexity of samples and establish reliable correlations with external stimuli and complex biological processes. Comprehensive two-dimensional chromatography offers a powerful tool to address these challenges by handling the chemical complexity and varying concentration ranges of samples/fractions, thereby providing a solid foundation for hypothesis development. Enhanced by specialized data processing techniques, this approach further expands research capabilities and aligns with modern individualized and personalized research trends.

GC×GC-MS stands out as a pivotal analytical platform in food-omics research. Its extensive application has the potential to advance the field significantly, offering deeper insights into food quality, nutritional value, and health benefits. The demand for practical solutions in industry does not preclude the use of MDA platforms and omics approaches within industrial research contexts. A broader investigative scope enhances understanding of how processing, storage, fermentation, and biotransformation impact food quality and safety [44], [76], [85], [156], [157].

Initially, the full capabilities of GC×GC may not have been fully recognized, but its widespread use across various fields and the integration of interdisciplinary strategies have underscored its vital role. It has evolved from being an underutilized technique to becoming an indispensable tool in food analysis: *"from a technique that did not exist... to a technique that was missing"* [158].

References

- [1] J. C. Giddings, "Concepts and comparisons in multidimensional separation," *J. High Resolut. Chromatogr.*, vol. 10, no. 5, pp. 319–323, 1987.
- [2] Z. Liu and J. B. Phillips, "Comprehensive Two-Dimensional Gas Chromatography using an On-Column Thermal Modulator Interface," *J. Chromatogr. Sci.*, vol. 29, no. 6, pp. 227–231, Jun. 1991.
- [3] M. Adahchour, J. Beens, R. J. J. Vreuls, and U. A. T. Brinkman, "Recent developments in comprehensive two-dimensional gas chromatography (GC × GC). I. Introduction and instrumental set-up," *TrAC - Trends Anal. Chem.*, vol. 25, no. 5, pp. 438–454, 2006.
- [4] M. Adahchour, J. Beens, R. J. J. Vreuls, and U. A. T. Brinkman, "Recent developments in comprehensive two-dimensional gas chromatography (GC x GC). II. Modulation and detection," *TrAC - Trends Anal. Chem.*, vol. 25, no. 6, pp. 540–553, 2006.
- [5] J. V. Seeley and S. K. Seeley, "Multidimensional gas chromatography: Fundamental advances and new applications," *Anal. Chem.*, vol. 85, no. 2, pp. 557–578, 2013.
- [6] S. E. Prebihalo *et al.*, "Multidimensional Gas Chromatography: Advances in Instrumentation, Chemometrics, and Applications," *Anal. Chem.*, vol. 90, no. 1, pp. 505–532, 2018.
- [7] P. Q. Tranchida, G. Purcaro, P. Dugo, L. Mondello, and G. Purcaro, "Modulators for comprehensive two-dimensional gas chromatography," *TrAC - Trends Anal. Chem.*, vol. 30, no. 9, pp. 1437–1461, 2011.
- [8] O. Fiehn, "Metabolomics - The link between genotypes and phenotypes," *Plant Mol. Biol.*, vol. 48, no. 1–2, pp. 155–171, 2002.
- [9] K. M. Pierce, B. Kehimkar, L. C. Marney, J. C. Hoggard, and R. E. Synovec, "Review of chemometric analysis techniques for comprehensive two dimensional separations data," *J. Chromatogr. A*, vol. 1255, pp. 3–11, Sep. 2012.
- [10] K. M. Pierce, J. C. Hoggard, R. E. Mohler, and R. E. Synovec, "Recent advancements in comprehensive two-dimensional separations with chemometrics," *J. Chromatogr. A*, vol. 1184, no. 1–2, pp. 341–352, 2008.
- [11] F. Stilo, C. Bicchi, A. Robbat, S. E. Reichenbach, and C. Cordero, "Untargeted approaches in food-omics: The potential of comprehensive two-dimensional gas chromatography/mass spectrometry," *TrAC Trends Anal. Chem.*, vol. 135, p. 116162, Feb. 2021.
- [12] A. Caratti *et al.*, "Boosting comprehensive two-dimensional chromatography with artificial intelligence: Application to food-omics," *TrAC Trends Anal. Chem.*, vol. 174, p. 117669, May 2024.
- [13] T. J. Trinklein, C. N. Cain, G. S. Ochoa, S. Schöneich, L. Mikaliunaite, and R. E. Synovec, "Recent Advances in GC×GC and Chemometrics to Address Emerging Challenges in Nontargeted Analysis," *Analytical Chemistry*, vol. 95, no. 1, pp. 264–286, 2023.
- [14] B. J. Pollo *et al.*, "Chemometrics, Comprehensive Two-Dimensional gas chromatography and 'omics' sciences: Basic tools and recent applications," *TrAC Trends*

Anal. Chem., vol. 134, p. 116111, 2021.

- [15] F. Stilo, C. Bicchi, A. M. Jimenez-Carvelo, L. Cuadros-Rodriguez, S. E. Reichenbach, and C. Cordero, "Chromatographic fingerprinting by comprehensive two-dimensional chromatography: Fundamentals and tools," *TrAC Trends Anal. Chem.*, vol. 134, p. 116133, Jan. 2021.
- [16] T. S. Bos *et al.*, "Recent applications of chemometrics in one- and two-dimensional chromatography," *Journal of Separation Science*, vol. 43, no. 9–10, pp. 1678–1727, 2020.
- [17] P.-H. Stefanuto, A. Smolinska, and J.-F. Focant, "Advanced chemometric and data handling tools for GC×GC-TOF-MS," *TrAC Trends Anal. Chem.*, vol. 139, p. 116251, Jun. 2021.
- [18] R. T. Peterson, "Chemical biology and the limits of reductionism," *Nat. Chem. Biol.*, vol. 4, no. 11, pp. 635–638, Nov. 2008.
- [19] R. R. Wilson, "Analytical Chemistry of Biological Systems," *Anal. Chem.*, vol. 63, no. 7, p. 379 A, 1991.
- [20] M. Adahchour, J. Beens, R. J. J. Vreuls, and U. A. T. Brinkman, "Recent developments in comprehensive two-dimensional gas chromatography (GC x GC). III. Applications for petrochemicals and organohalogens," *TrAC - Trends Anal. Chem.*, vol. 25, no. 7, pp. 726–741, 2006.
- [21] P. Q. Tranchida, P. Donato, F. Cacciola, M. Beccaria, P. Dugo, and L. Mondello, "Potential of comprehensive chromatography in food analysis," *TrAC - Trends Anal. Chem.*, vol. 52, pp. 186–205, 2013.
- [22] P. Marriott and R. Shellie, "Principles and applications of comprehensive two-dimensional gas chromatography," *TrAC - Trends in Analytical Chemistry*, vol. 21, no. 9–10, pp. 573–583, 2002.
- [23] M. Adahchour, J. Beens, R. J. J. Vreuls, and U. A. T. Brinkman, "Recent developments in comprehensive two-dimensional gas chromatography (GC x GC). IV. Further applications, conclusions and perspectives," *TrAC - Trends Anal. Chem.*, vol. 25, no. 8, pp. 821–840, 2006.
- [24] C. Cordero, J. Kiefl, P. Schieberle, S. E. Reichenbach, and C. Bicchi, "Comprehensive two-dimensional gas chromatography and food sensory properties: potential and challenges," *Anal. Bioanal. Chem.*, vol. 407, no. 1, pp. 169–191, Jan. 2015.
- [25] C. Cordero, J. Kiefl, S. E. Reichenbach, and C. Bicchi, "Characterization of odorant patterns by comprehensive two-dimensional gas chromatography: A challenge in omic studies," *TrAC Trends Anal. Chem.*, vol. 113, pp. 364–378, Apr. 2019.
- [26] B. Gruber, B. A. Weggler, R. Jaramillo, K. A. Murrell, P. K. Piotrowski, and F. L. Dorman, "Comprehensive two-dimensional gas chromatography in forensic science: A critical review of recent trends," *TrAC - Trends Anal. Chem.*, vol. 105, pp. 292–301, 2018.
- [27] J. C. Giddings, "Sample dimensionality: A predictor of order-disorder in component peak distribution in multidimensional separation," *J. Chromatogr. A*, vol. 703, no. 1–2, pp. 3–15, Jun. 1995.

- [28] P. Q. Tranchida, G. Purcaro, M. Maimone, and L. Mondello, "Impact of comprehensive two-dimensional gas chromatography with mass spectrometry on food analysis," *J. Sep. Sci.*, vol. 39, no. 1, pp. 149–161, 2016.
- [29] E. Gabetti *et al.*, "Chemical fingerprinting strategies based on comprehensive two-dimensional gas chromatography combined with gas chromatography-olfactometry to capture the unique signature of Piemonte peppermint essential oil (*Mentha x piperita* var *Italo-Mitcham*)," *J. Chromatogr. A*, vol. 1645, p. 462101, May 2021.
- [30] C. Cordero, H. G. Schmarr, S. E. Reichenbach, and C. Bicchi, "Current Developments in Analyzing Food Volatiles by Multidimensional Gas Chromatographic Techniques," *Journal of Agricultural and Food Chemistry*, vol. 66, no. 10, pp. 2226–2236, Feb-2018.
- [31] P. J. Marriott and Y. Nolvachai, "Separation Multidimensionality for Improved Sample Characterization—Is It Worth the Effort?," *J. Agric. Food Chem.*, vol. 69, no. 39, pp. 11535–11537, Oct. 2021.
- [32] B. d'Acampora Zellner, A. Casilli, P. Dugo, G. Dugo, and L. Mondello, "Odour fingerprint acquisition by means of comprehensive two-dimensional gas chromatography-olfactometry and comprehensive two-dimensional gas chromatography/mass spectrometry," *J. Chromatogr. A*, vol. 1141, no. 2, pp. 279–286, 2007.
- [33] L. Nicolotti, V. Mall, and P. Schieberle, "Characterization of Key Aroma Compounds in a Commercial Rum and an Australian Red Wine by Means of a New Sensomics-Based Expert System (SEBES) - An Approach to Use Artificial Intelligence in Determining Food Odor Codes," *J. Agric. Food Chem.*, vol. 67, no. 14, pp. 4011–4022, 2019.
- [34] S. T. Chin, G. T. Eyres, and P. J. Marriott, "Application of integrated comprehensive/multidimensional gas chromatography with mass spectrometry and olfactometry for aroma analysis in wine and coffee," *Food Chem.*, vol. 185, pp. 355–361, 2015.
- [35] M. K. Jennerwein, M. Eschner, T. Gröger, T. Wilharm, and R. Zimmermann, "Complete group-type quantification of petroleum middle distillates based on comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry (GCxGC-TOFMS) and visual basic scripting," *Energy and Fuels*, vol. 28, no. 9, pp. 5670–5681, 2014.
- [36] C. Vendeuvre, R. Ruiz-Guerrero, F. Bertoncini, L. Duval, and D. Thiébaud, "Comprehensive two-dimensional gas chromatography for detailed characterisation of petroleum products," *Oil Gas Sci. Technol.*, vol. 62, no. 1, pp. 43–55, 2007.
- [37] D. D. Yan *et al.*, "Chemotyping of new hop (*Humulus lupulus* L.) genotypes using comprehensive two-dimensional gas chromatography with quadrupole accurate mass time-of-flight mass spectrometry," *J. Chromatogr. A*, 2017.
- [38] D. Ryan, R. Shellie, P. Tranchida, A. Casilli, L. Mondello, and P. Marriott, "Analysis of roasted coffee bean volatiles by using comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry," *J. Chromatogr. A*, vol. 1054, no. 1–2, pp. 57–65, 2004.
- [39] L. Mondello *et al.*, "Comprehensive multidimensional GC for the characterization of roasted coffee beans," *J. Sep. Sci.*, vol. 27, no. 5–6, pp. 442–450, 2004.

- [40] F. Magagna *et al.*, “Comprehensive Chemical Fingerprinting of High-Quality Cocoa at Early Stages of Processing: Effectiveness of Combined Untargeted and Targeted Approaches for Classification and Discrimination,” *J. Agric. Food Chem.*, vol. 65, no. 30, pp. 6329–6341, 2017.
- [41] F. Magagna *et al.*, “Advanced fingerprinting of high-quality cocoa: Challenges in transferring methods from thermal to differential-flow modulated comprehensive two dimensional gas chromatography,” *J. Chromatogr. A*, vol. 1536, pp. 122–136, 2018.
- [42] E. M. Humston, Y. Zhang, G. F. Brabeck, A. McShea, and R. E. Synovec, “Development of a GCxGC-TOFMS method using SPME to determine volatile compounds in cacao beans,” *J. Sep. Sci.*, vol. 32, no. 13, pp. 2289–2295, 2009.
- [43] E. M. Humston, J. D. Knowles, A. McShea, and R. E. Synovec, “Quantitative assessment of moisture damage for cacao bean quality using two-dimensional gas chromatography combined with time-of-flight mass spectrometry and chemometrics,” *J. Chromatogr. A*, vol. 1217, no. 12, pp. 1963–1970, 2010.
- [44] A. Caratti *et al.*, “Augmented visualization by computer vision and chromatographic fingerprinting on comprehensive two-dimensional gas chromatographic patterns: Unraveling diagnostic signatures in food volatilome,” *J. Chromatogr. A*, vol. 1699, p. 464010, Jun. 2023.
- [45] J. Kiefl, G. Pollner, and P. Schieberle, “Sensomics analysis of key hazelnut odorants (*Corylus avellana* L. 'Tonda Gentile') using comprehensive two-dimensional gas chromatography in combination with time-of-flight mass spectrometry (GCxGC-TOF-MS),” *J. Agric. Food Chem.*, vol. 61, no. 22, pp. 5226–5235, 2013.
- [46] C. Cordero *et al.*, “Profiling food volatiles by comprehensive two-dimensional gas chromatography coupled with mass spectrometry: Advanced fingerprinting approaches for comparative analysis of the volatile fraction of roasted hazelnuts (*Corylus avellana* L.) from different origins,” *J. Chromatogr. A*, vol. 1217, no. 37, pp. 5848–5858, 2010.
- [47] C. Cordero, C. Bicchi, and P. Rubiolo, “Group-type and fingerprint analysis of roasted food matrices (coffee and hazelnut samples) by comprehensive two-dimensional gas chromatography,” *J. Agric. Food Chem.*, vol. 56, no. 17, pp. 7655–7666, 2008.
- [48] N. I. of S. and Technology, *NIST/EPA/NIH Mass Spectral Library with Search Program: (Data Version: NIST 08, Software Version 2.0f)*, vol. 2010, no. 1/22. Gaithersburg MD: National Institute of Standards and Technology (NIST), 2005.
- [49] F. W. McLafferty, “Wiley Registry of Mass Spectral Data, 12th Edition.” Wiley, 2020.
- [50] E. Liberto *et al.*, “Enantiomer identification in the flavour and fragrance fields by ‘interactive’ combination of linear retention indices from enantioselective gas chromatography and mass spectrometry,” *J. Chromatogr. A*, vol. 1195, no. 1–2, pp. 117–126, 2008.
- [51] B. d’Acampora Zellner, C. Bicchi, P. Dugo, P. Rubiolo, G. Dugo, and L. Mondello, “Linear retention indices in gas chromatographic analysis: A review,” *Flavour and Fragrance Journal*, vol. 23, no. 5, pp. 297–314, Sep-2008.

- [52] S. E. Reichenbach, Q. Tao, C. Cordero, and C. Bicchi, "A data-challenge case study of analyte detection and identification with comprehensive two-dimensional gas chromatography with mass spectrometry (GC×GC-MS)," *Separations*, vol. 6, no. 3, 2019.
- [53] F. Magagna *et al.*, "Black tea volatiles fingerprinting by comprehensive two-dimensional gas chromatography – Mass spectrometry combined with high concentration capacity sample preparation techniques: Toward a fully automated sensomic assessment," *Food Chem.*, vol. 225, pp. 276–287, 2017.
- [54] L. Nicolotti *et al.*, "Quantitative fingerprinting by headspace-Two-dimensional comprehensive gas chromatography-mass spectrometry of solid matrices: Some challenging aspects of the exhaustive assessment of food volatiles," *Anal. Chim. Acta*, vol. 798, pp. 115–125, 2013.
- [55] Markes International, "Select-eV: The next generation of ion source technology," *Technical Note*, vol. Applicatio. 2016.
- [56] C. Cordero *et al.*, "Comprehensive two-dimensional gas chromatography coupled with time of flight mass spectrometry featuring tandem ionization: Challenges and opportunities for accurate fingerprinting studies," *J. Chromatogr. A*, vol. 1597, pp. 132–141, Jul. 2019.
- [57] M. Cialie Rosso *et al.*, "Adding extra-dimensions to hazelnuts primary metabolome fingerprinting by comprehensive two-dimensional gas chromatography combined with time-of-flight mass spectrometry featuring tandem ionization: Insights on the aroma potential," *J. Chromatogr. A*, vol. 1614, no. 460739, pp. 1–11, 2020.
- [58] W. Welthagen, S. Mitschke, F. Muhlberger, and R. Zimmermann, "One-dimensional and comprehensive two-dimensional gas chromatography coupled to soft photo ionization time-of-flight mass spectrometry : A two- and three-dimensional separation approach," *J. Chromatogr. A*, vol. 1150, pp. 54–61, 2007.
- [59] C. R. Warren, "Use of chemical ionization for GC-MS metabolite profiling," *Metabolomics*, vol. 9, no. SUPPL.1, pp. 110–120, 2013.
- [60] M. S. Alam, C. Stark, and R. M. Harrison, "Using Variable Ionization Energy Time-of-Flight Mass Spectrometry with Comprehensive GC×GC to Identify Isomeric Species," *Anal. Chem.*, vol. 88, no. 8, pp. 4211–4220, 2016.
- [61] L. M. Dubois *et al.*, "Thermal desorption comprehensive two-dimensional gas chromatography coupled to variable-energy electron ionization time-of-flight mass spectrometry for monitoring subtle changes in volatile organic compound profiles of human blood," *J. Chromatogr. A*, vol. 1501, pp. 117–127, 2017.
- [62] C. E. Freye, N. R. Moore, and R. E. Synovec, "Enhancing the chemical selectivity in discovery-based analysis with tandem ionization time-of-flight mass spectrometry detection for comprehensive two-dimensional gas chromatography," *J. Chromatogr. A*, vol. 1537, pp. 99–108, 2018.
- [63] F. Magagna *et al.*, "Combined untargeted and targeted fingerprinting with comprehensive two-dimensional chromatography for volatiles and ripening indicators in olive oil," *Anal. Chim. Acta*, vol. 936, pp. 245–258, Sep. 2016.

- [64] D. Bressanello *et al.*, “Combined untargeted and targeted fingerprinting by comprehensive two-dimensional gas chromatography: revealing fructose-induced changes in mice urinary metabolic signatures,” *Anal. Bioanal. Chem.*, vol. 410, no. 11, pp. 2723–2737, 2018.
- [65] F. Stilo, E. Liberto, S. E. Reichenbach, Q. Tao, C. Bicchi, and C. Cordero, “Exploring the Extra-Virgin Olive Oil Volatilome by Adding Extra Dimensions to Comprehensive Two-Dimensional Gas Chromatography and Time-of-Flight Mass Spectrometry Featuring Tandem Ionization: Validation of Ripening Markers in Headspace Linearity Conditio,” *J. AOAC Int.*, vol. 104, no. 2, pp. 274–287, 2021.
- [66] M. Beccaria, F. A. Franchina, M. Nasir, T. Mellors, J. E. Hill, and G. Purcaro, “Investigation of mycobacteria fatty acid profile using different ionization energies in GC-MS,” *Anal. Bioanal. Chem.*, vol. 410, no. 30, pp. 7987–7996, Dec. 2018.
- [67] P. Q. Tranchida, S. Salivo, I. Bonaccorsi, A. Rotondo, P. Dugo, and L. Mondello, “Analysis of the unsaponifiable fraction of lipids belonging to various milk-types by using comprehensive two-dimensional gas chromatography with dual mass spectrometry/flame ionization detection and with the support of high resolution time-of-flight mass ,” *J. Chromatogr. A*, vol. 1313, pp. 194–201, 2013.
- [68] P. Q. Tranchida, S. Salivo, F. A. Franchina, I. Bonaccorsi, P. Dugo, and L. Mondello, “Qualitative and quantitative analysis of the unsaponifiable fraction of vegetable oils by using comprehensive 2D GC with dual MS/FID detection,” *Anal. Bioanal. Chem.*, vol. 405, no. 13, pp. 4655–4663, 2013.
- [69] I. Aloisi, M. Zoccali, P. Dugo, P. Q. Tranchida, and L. Mondello, “Fingerprinting of the Unsaponifiable Fraction of Vegetable Oils by Using Cryogenically-Modulated Comprehensive Two-Dimensional Gas Chromatography-High Resolution Time-of-Flight Mass Spectrometry,” *Food Anal. Methods*, vol. 13, no. 7, pp. 1523–1529, 2020.
- [70] C. Simó, A. Cifuentes, and V. B. T.-C. A. C. García-Cañas, Eds., “Comprehensive Analytical Chemistry,” in *Fundamentals of Advanced Omics Technologies: From Genes to Metabolites*, vol. 63, Elsevier, 2014, p. ii.
- [71] F. Stilo, C. Bicchi, S. E. Reichenbach, and C. Cordero, “Comprehensive two-dimensional gas chromatography as a boosting technology in food-omic investigations,” *J. Sep. Sci.*, vol. 44, no. 8, pp. 1592–1611, Apr. 2021.
- [72] F. Capozzi and A. Bordonì, “Foodomics: A new comprehensive approach to food and nutrition,” *Genes Nutr.*, vol. 8, no. 1, pp. 1–4, 2013.
- [73] C. H. Weinert, B. Egert, and S. E. Kulling, “On the applicability of comprehensive two-dimensional gas chromatography combined with a fast-scanning quadrupole mass spectrometer for untargeted large-scale metabolomics,” *J. Chromatogr. A*, vol. 1405, pp. 156–167, 2015.
- [74] D. M. Jacobs, M. A. van den Berg, and R. D. Hall, “Towards superior plant-based foods using metabolomics,” *Current Opinion in Biotechnology*. 2021.
- [75] M. L. Romo-Pérez *et al.*, “Metabolite profiling of onion landraces and the cold storage effect,” *Plant Physiol. Biochem.*, vol. 146, pp. 428–437, 2020.

- [76] S. Squara *et al.*, “Artificial intelligence decision making tools in food metabolomics: Data fusion unravels synergies within the hazelnut (*Corylus avellana* L.) metabolome and improves quality prediction,” *Food Res. Int.*, vol. 194, p. 114873, Oct. 2024.
- [77] M. M. Ulaszewska *et al.*, “Nutrimetabolomics: An Integrative Action for Metabolomic Analyses in Human Nutritional Studies,” *Mol. Nutr. Food Res.*, vol. 63, no. 1, p. 1800384, Jan. 2019.
- [78] A. Scalbert *et al.*, “The food metabolome: A window over dietary exposure,” *Am. J. Clin. Nutr.*, vol. 99, no. 6, pp. 1286–1308, 2014.
- [79] A. Dunkel *et al.*, “Nature’s chemical signatures in human olfaction: A foodborne perspective for future biotechnology,” *Angew. Chemie - Int. Ed.*, vol. 53, no. 28, pp. 7124–7143, 2014.
- [80] J. Charve, C. Chen, A. D. Hegeman, and G. A. Reineccius, “Evaluation of instrumental methods for the untargeted analysis of chemical stimuli of orange juice flavour,” *Flavour Fragr. J.*, vol. 26, no. 6, pp. 429–440, 2011.
- [81] Y. Y. Broza, P. Mochalski, V. Ruzsanyi, A. Amann, and H. Haick, “Hybrid volatilomics and disease detection,” *Angew. Chem. Int. Ed. Engl.*, vol. 54, no. 38, pp. 11036–11048, Sep. 2015.
- [82] M. Phillips *et al.*, “Detection of an extended human volatome with comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry,” *PLoS One*, vol. 8, no. 9, p. e75274, 2013.
- [83] A. E. Lytou, E. Z. Panagou, and G.-J. E. Nychas, “Volatilomics for food quality and authentication,” *Curr. Opin. Food Sci.*, vol. 28, pp. 88–95, Aug. 2019.
- [84] M. Noshad, B. A. Behbahani, and I. K. Karabagias, “Volatilomic with chemometrics: a toward authentication approach for food authenticity control,” *European Food Research and Technology*. 2023.
- [85] S. Squara *et al.*, “Artificial Intelligence decision-making tools based on comprehensive two-dimensional gas chromatography data: the challenge of quantitative volatilomics in food quality assessment,” *J. Chromatogr. A*, vol. 1700, p. 464041, Jul. 2023.
- [86] G. Purcaro, L. Barp, M. Beccaria, and L. S. Conte, “Fingerprinting of vegetable oil minor components by multidimensional comprehensive gas chromatography with dual detection,” *Anal. Bioanal. Chem.*, vol. 407, no. 1, pp. 309–319, 2015.
- [87] G. Purcaro, L. Barp, M. Beccaria, and L. S. Conte, “Characterisation of minor components in vegetable oil by comprehensive gas chromatography with dual detection,” *Food Chem.*, vol. 212, pp. 730–738, 2016.
- [88] L. Cuadros-Rodríguez, C. Ruiz-Samblás, L. Valverde-Som, E. Pérez-Castaño, and A. González-Casado, “Chromatographic fingerprinting: An innovative approach for food ‘identification’ and food authentication – A tutorial,” *Anal. Chim. Acta*, vol. 909, pp. 9–23, Feb. 2016.
- [89] E. C. Horning and M. G. Horning, “Human Metabolic Profiles Obtained by GC and GC/MS,” *J. Chromatogr. Sci.*, vol. 9, no. 3, pp. 129–140, Mar. 1971.

- [90] O. Fiehn, "Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks," *Comp. Funct. Genomics*, vol. 2, no. 3, pp. 155–168, 2001.
- [91] D. J. Beale *et al.*, "Review of recent developments in GC–MS approaches to metabolomics-based research," *Metabolomics*, vol. 14, no. 11, p. 152, 2018.
- [92] J. L. Wolfender, G. Marti, A. Thomas, and S. Bertrand, "Current approaches and challenges for the metabolite profiling of complex natural extracts," *J. Chromatogr. A*, vol. 1382, pp. 136–164, 2015.
- [93] A. Alonso, S. Marsal, and A. Julià, "Analytical Methods in Untargeted Metabolomics: State of the Art in 2015," *Front. Bioeng. Biotechnol.*, vol. 3, p. 23, 2015.
- [94] C. Bicchi and M. Maffei, *The Plant Volatilome: Methods of Analysis. in High-Throughput Phenotyping in Plants. Methods in Molecular Biology (Methods and Protocols)*. Totowa, NJ: Humana Press, 2012.
- [95] T. Cajka, K. Ridelova, E. Klimankova, M. Cerna, F. Pudil, and J. Hajslova, "Traceability of olive oil based on volatiles pattern and multivariate analysis," *Food Chem.*, vol. 121, no. 1, pp. 282–289, 2010.
- [96] I. Stanimirova *et al.*, "Tracing the geographical origin of honeys based on volatile compounds profiles assessment using pattern recognition techniques," *Food Chem.*, vol. 118, no. 1, pp. 171–176, 2010.
- [97] P. E. Sudol, M. Galletta, P. Q. Tranchida, M. Zoccali, L. Mondello, and R. E. Synovec, "Untargeted profiling and differentiation of geographical variants of wine samples using headspace solid-phase microextraction flow-modulated comprehensive two-dimensional gas chromatography with the support of tile-based Fisher ratio analysis," *J. Chromatogr. A*, vol. 1662, p. 462735, Jan. 2022.
- [98] Y. F. Wong, P. Perlmutter, and P. J. Marriott, "Untargeted metabolic profiling of Eucalyptus spp. leaf oils using comprehensive two-dimensional gas chromatography with high resolution mass spectrometry: Expanding the metabolic coverage," *Metabolomics*, vol. 13, no. 5, pp. 1–17, 2017.
- [99] F. Stilo, E. Liberto, S. E. Reichenbach, Q. Tao, C. Bicchi, and C. Cordero, "Untargeted and Targeted Fingerprinting of Extra Virgin Olive Oil Volatiles by Comprehensive Two-Dimensional Gas Chromatography with Mass Spectrometry: Challenges in Long-Term Studies," *J. Agric. Food Chem.*, vol. 67, no. 18, pp. 5289–5302, May 2019.
- [100] A. Da Ros *et al.*, "Complementary untargeted and targeted metabolomics for differentiation of extra virgin olive oils of different origin of purchase based on volatile and phenolic composition and sensory quality," *Molecules*, vol. 24, no. 16, pp. 1–17, 2019.
- [101] L. T. Vaz-Freire, M. D. R. G. da Silva, and A. M. C. Freitas, "Comprehensive two-dimensional gas chromatography for fingerprint pattern recognition in olive oils produced by two different techniques in Portuguese olive varieties Galega Vulgar, Cobrançosa e Carrasquenha," *Anal. Chim. Acta*, vol. 633, no. 2, pp. 263–270, 2009.
- [102] P. Q. Tranchida *et al.*, "Elucidation of fatty acid profiles in vegetable oils exploiting group-type patterning and enhanced sensitivity of comprehensive two-dimensional gas chromatography," *J. Sep. Sci.*, vol. 31, no. 10, pp. 1797–1802, 2008.

- [103] P. Q. Tranchida, P. Donato, G. Dugo, L. Mondello, and P. Dugo, "Comprehensive chromatographic methods for the analysis of lipids," *TrAC - Trends Anal. Chem.*, vol. 26, no. 3, pp. 191–205, 2007.
- [104] A. Fina *et al.*, "A high throughput method for fatty acid profiling using simultaneous microwave-assisted extraction and derivatization followed by reversed fill/flush flow modulation comprehensive multidimensional gas chromatography," *Adv. Sample Prep.*, vol. 4, p. 100039, Oct. 2022.
- [105] D. Ferrara *et al.*, "Composition and nutritional values of fatty acids in marine organisms by one-step microwave-assisted extraction/derivatization and comprehensive two-dimensional gas chromatography -flame ionization detector," *J. Chromatogr. B*, vol. 1236, p. 124074, Apr. 2024.
- [106] C. Mack *et al.*, "Untargeted multi-platform analysis of the metabolome and the non-starch polysaccharides of kiwifruit during postharvest ripening," *Postharvest Biol. Technol.*, vol. 125, pp. 65–76, 2017.
- [107] A. Kanbar *et al.*, "Cold tolerance of woodland strawberry (*Fragaria vesca*) is linked to Cold Box Factor 4 and the dehydrin Xero2," *J. Exp. Bot.*, p. eiae263, Jul. 2024.
- [108] H.-G. Schmarr and J. Bernhardt, "Profiling analysis of volatile compounds from fruits using comprehensive two-dimensional gas chromatography and image processing techniques," *J. Chromatogr. A*, vol. 1217, no. 4, pp. 565–574, 2010.
- [109] N. Kfoury *et al.*, "Striking changes in tea metabolites due to elevational effects," *Food Chem.*, vol. 264, pp. 334–341, 2018.
- [110] N. Kfoury *et al.*, "Plant-Climate Interaction Effects: Changes in the Relative Distribution and Concentration of the Volatile Tea Leaf Metabolome in 2014–2016," *Front. Plant Sci.*, vol. 10, no. November, pp. 1–10, Nov. 2019.
- [111] F. Stilo, G. Tredici, C. Bicchi, A. Robbat, J. Morimoto, and C. Cordero, "Climate and Processing Effects on Tea (*Camellia sinensis* L. Kuntze) Metabolome: Accurate Profiling and Fingerprinting by Comprehensive Two-Dimensional Gas Chromatography/Time-of-Flight Mass Spectrometry," *Molecules*, vol. 25, no. 10, p. 2447, May 2020.
- [112] E. C. Y. Chan, K. K. Pasikanti, and J. K. Nicholson, "Global urinary metabolic profiling procedures using gas chromatography-mass spectrometry," *Nat. Protoc.*, vol. 6, no. 10, pp. 1483–1499, 2011.
- [113] A. Scalbert *et al.*, "Mass-spectrometry-based metabolomics: Limitations and recommendations for future progress with particular focus on nutrition research," *Metabolomics*, vol. 5, no. 4, pp. 435–458, 2009.
- [114] N. Alygizakis *et al.*, "Towards a harmonized identification scoring system in LC-HRMS/MS based non-target screening (NTS) of emerging contaminants," *TrAC Trends Anal. Chem.*, vol. 159, p. 116944, Feb. 2023.
- [115] J. Morimoto *et al.*, "Untargeted/targeted 2D gas chromatography/mass spectrometry detection of the total volatile tea metabolome," *Molecules*, vol. 24, no. 20, pp. 1–14, 2019.
- [116] P. Delmonte, A. R. Fardin-Kia, and J. I. Rader, "Separation of fatty acid methyl esters by GC-online hydrogenation × GC," *Anal. Chem.*, vol. 85, no. 3, pp. 1517–1524, Feb. 2013.

- [117] P. Q. Tranchida, F. A. Franchina, P. Dugo, and L. Mondello, "A flow-modulated comprehensive gas chromatography-mass spectrometry method for the analysis of fatty acid profiles in marine and biological samples," *J. Chromatogr. A*, vol. 1255, pp. 171–176, 2012.
- [118] P. Q. Tranchida *et al.*, "Acquisition of deeper knowledge on the human plasma fatty acid profile exploiting comprehensive 2-D GC," *J. Sep. Sci.*, vol. 31, no. 19, pp. 3347–3351, 2008.
- [119] S. Squara *et al.*, "Extending the breadth of saliva metabolome fingerprinting by smart template strategies and effective pattern realignment on comprehensive two-dimensional gas chromatographic data," *Anal. Bioanal. Chem.*, no. 0123456789, Jan. 2023.
- [120] M. Cialiè Rosso *et al.*, "Exploring extra dimensions to capture saliva metabolite fingerprints from metabolically healthy and unhealthy obese patients by comprehensive two-dimensional gas chromatography featuring Tandem Ionization mass spectrometry," *Anal. Bioanal. Chem.*, vol. 413, no. 2, pp. 403–418, Jan. 2021.
- [121] M. F. Almstetter, P. J. Oefner, and K. Dettmer, "Comprehensive two-dimensional gas chromatography in metabolomics," *Anal. Bioanal. Chem.*, vol. 402, no. 6, pp. 1993–2013, Feb. 2012.
- [122] C. I. Mack *et al.*, "Robust Markers of Coffee Consumption Identified Among the Volatile Organic Compounds in Human Urine," *Mol. Nutr. Food Res.*, vol. 63, no. 10, 2019.
- [123] C. Vetrani *et al.*, "Metabolic transformations of dietary polyphenols: comparison between in vitro colonic and hepatic models and in vivo urinary metabolites," *J. Nutr. Biochem.*, vol. 33, pp. 111–118, 2016.
- [124] R. Mastrocola *et al.*, "Toxicological impact of high fructose intake on gut microbiota and liver/intestine integrity: Any differences between solid and liquid formulations?," *Diabetologia*, vol. 60, pp. S245–S245, 2017.
- [125] M. Cialiè Rosso *et al.*, "Combined Untargeted and Targeted Fingerprinting by Comprehensive Two-Dimensional Gas Chromatography to Track Compositional Changes on Hazelnut Primary Metabolome during Roasting," *Appl. Sci.*, vol. 11, no. 2, p. 525, Jan. 2021.
- [126] P. Schieberle and T. Hofmann, "Mapping the combinatorial code of food flavors by means of molecular sensory science approach," in *Food Flavors: Chemical, Sensory and Technological Properties*, 2011, pp. 413–438.
- [127] S. Firestein, "How the olfactory system makes sense of scent.," *Nature*, vol. 413, no. 6852, pp. 211–218, 2001.
- [128] H. Breer, J. Fleischer, and J. Strotmann, "The sense of smell: Multiple olfactory subsystems," *Cell. Mol. Life Sci.*, vol. 63, no. 13, pp. 1465–1475, 2006.
- [129] K. Audouze *et al.*, "Identification of odorant-receptor interactions by global mapping of the human odorome," *PLoS One*, vol. 9, no. 4, 2014.
- [130] G. Purcaro, C. Cordero, E. Liberto, C. Bicchi, and L. S. Conte, "Toward a definition of blueprint of virgin olive oil by comprehensive two-dimensional gas chromatography," *J. Chromatogr. A*, vol. 1334, pp. 101–111, 2014.

- [131] W. Grosch, "Evaluation of the key odorants of foods by dilution experiments, aroma models and omission," *Chem. Senses*, vol. 26, no. 5, pp. 533–545, 2001.
- [132] C. Kulsing, Y. Nolvachai, and P. J. Marriott, "Concepts, selectivity options and experimental design approaches in multidimensional and comprehensive two-dimensional gas chromatography," *TrAC - Trends Anal. Chem.*, vol. 130, p. 115995, 2020.
- [133] S. T. Chin, G. T. Eyres, and P. J. Marriott, "System design for integrated comprehensive and multidimensional gas chromatography with mass spectrometry and olfactometry," *Anal. Chem.*, vol. 84, no. 21, pp. 9154–9162, 2012.
- [134] B. Maikhunthod, P. D. Morrison, D. M. Small, and P. J. Marriott, "Development of a switchable multidimensional/comprehensive two-dimensional gas chromatographic analytical system," *J. Chromatogr. A*, vol. 1217, no. 9, pp. 1522–1529, 2010.
- [135] S. T. Chin, G. T. Eyres, and P. J. Marriott, "Identification of potent odourants in wine and brewed coffee using gas chromatography-olfactometry and comprehensive two-dimensional gas chromatography," *J. Chromatogr. A*, vol. 1218, no. 42, pp. 7487–7498, 2011.
- [136] A. Gracka, H. H. Jelen, M. Majcher, A. Siger, and A. Kaczmarek, "Flavoromics approach in monitoring changes in volatile compounds of virgin rapeseed oil caused by seed roasting," *J. Chromatogr. A*, vol. 1428, pp. 292–304, 2016.
- [137] G. Dugo *et al.*, "Elucidation of the volatile composition of marsala wines by using comprehensive two-dimensional gas chromatography," *Food Chem.*, vol. 142, pp. 262–268, 2014.
- [138] F. Stilo *et al.*, "An effective chromatographic fingerprinting workflow based on comprehensive two-dimensional gas chromatography – Mass spectrometry to establish volatiles patterns discriminative of spoiled hazelnuts (*Corylus avellana* L.)," *Food Chem.*, vol. 340, p. 128135, Mar. 2021.
- [139] M. T. Morales, G. Luna, and R. Aparicio, "Comparative study of virgin olive oil sensory defects," *Food Chem.*, vol. 91, no. 2, pp. 293–301, 2005.
- [140] C. Oliver-Pozo, D. Trypidis, R. Aparicio, D. L. García-González, and R. Aparicio-Ruiz, "Implementing Dynamic Headspace with SPME Sampling of Virgin Olive Oil Volatiles: Optimization, Quality Analytical Study, and Performance Testing," *J. Agric. Food Chem.*, vol. 67, no. 7, pp. 2086–2097, 2019.
- [141] B. Kolb and L. S. Ettre, *Static Headspace – Gas Chromatography: Theory and Practice Second Edition*, Second. New York: Wiley-VCH, 2006.
- [142] F. Stilo, C. Cordero, B. Sgorbini, C. Bicchi, and E. Liberto, "Highly Informative Fingerprinting of Extra-Virgin Olive Oil Volatiles: The Role of High Concentration-Capacity Sampling in Combination with Comprehensive Two-Dimensional Gas Chromatography," *Separations*, vol. 6, no. 3, p. 34, Jul. 2019.
- [143] M. Biedermann, C. Munoz, and K. Grob, "Epoxidation for the analysis of the mineral oil aromatic hydrocarbons in food. An update," *J. Chromatogr. A*, vol. 1624, 2020.
- [144] M. Zoccali, L. Barp, M. Beccaria, D. Sciarrone, G. Purcaro, and L. Mondello, "Improvement of mineral oil saturated and aromatic hydrocarbons determination in

edible oil by liquid-liquid-gas chromatography with dual detection,” *J. Sep. Sci.*, vol. 39, no. 3, pp. 623–631, 2016.

- [145] M. Biedermann and K. Grob, “Advantages of comprehensive two-dimensional gas chromatography for comprehensive analysis of potential migrants from food contact materials,” *Anal. Chim. Acta*, vol. 1057, pp. 11–17, 2019.
- [146] S. Weber, K. Schrag, G. Mildau, T. Kuballa, S. G. Walch, and D. W. Lachenmeier, “Analytical Methods for the Determination of Mineral Oil Saturated Hydrocarbons (MOSH) and Mineral Oil Aromatic Hydrocarbons (MOAH)—A Short Review,” *Anal. Chem. Insights*, vol. 13, p. 1177390118777757, Jan. 2018.
- [147] M. Biedermann and K. Grob, “On-line coupled high performance liquid chromatography-gas chromatography for the analysis of contamination by mineral oil. Part 2: Migration from paperboard into dry foods: Interpretation of chromatograms,” *Journal of Chromatography A*, vol. 1255. Elsevier, pp. 76–99, 14-Sep-2012.
- [148] M. Biedermann, C. Munoz, and K. Grob, “Update of on-line coupled liquid chromatography – gas chromatography for the analysis of mineral oil hydrocarbons in foods and cosmetics,” *J. Chromatogr. A*, vol. 1521, pp. 140–149, Oct. 2017.
- [149] Commission Decision 657/2002, “Commission Decision 2002/657/EC implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results,” *Off. J. Eur. Union*, vol. L221, no. 23 May 1996, pp. 8–36, 2002.
- [150] E. Commission, “Commission Implementing Regulation (EU) 2021/808 of 22 March 2021 on the performance of analytical methods for residues of pharmacologically active substances used in food-producing animals and on the interpretation of results as well as on the methods to,” *Off. J. Eur. Union*, vol. 180, no. 401, pp. 84–109, 2021.
- [151] M. Biedermann and K. Grob, “Comprehensive two-dimensional gas chromatography for characterizing mineral oils in foods and distinguishing them from synthetic hydrocarbons,” *J. Chromatogr. A*, vol. 1375, pp. 146–153, 2015.
- [152] S. Pantò, M. Collard, and G. Purcaro, “Comprehensive Gas Chromatography Coupled to Simultaneous Dual Detection (TOF-MS/FID) as a Confirmatory Method for MOSH and MOAH Determination in Food,” *Curr. Trends Mass Spectrom.*, no. July 2020, pp. 15–20, 2020.
- [153] Y. Zou *et al.*, “Determination of short-chain chlorinated paraffins using comprehensive two-dimensional gas chromatography coupled with low resolution mass spectrometry,” *J. Chromatogr. A*, vol. 1581–1582, pp. 135–143, 2018.
- [154] D. Schrenk *et al.*, “Risk assessment of chlorinated paraffins in feed and food,” *EFSA J.*, vol. 18, no. 3, pp. 1–220, 2020.
- [155] M. Tobiszewski, “Metrics for green analytical chemistry,” 2016.
- [156] G. Bauwens *et al.*, “Investigation of the effect of refining on the presence of targeted mineral oil aromatic hydrocarbons in coconut oil,” *Food Addit. Contam. - Part A*, 2023.
- [157] G. Bauwens, S. Pantó, and G. Purcaro, “Mineral oil saturated and aromatic hydrocarbons quantification: Mono- and two-dimensional approaches,” *J. Chromatogr. A*, 2021.

- [158] C. Cordero and C. Bicchi, “Third GCxGC-MS School - Sic et simpliciter - Mass Spectrometry Group Italian Chemical Society,” in *Division of Mass Spectrometry - Italian Chemical Society*, 2020.

Experimental Section I

Green(er) Transition of Established Analytical Methods

Chapter 6

EPA Method 8260D Making Greener the Volatile Organic Compounds Detection in Environmental Samples by Gas Chromatography-Mass Spectrometry

This chapter explores the adaptation of EPA Method 8260D for the detection of volatile organic compounds (VOCs) in environmental samples, with a focus on enhancing sustainability in gas chromatography-mass spectrometry (GC-MS) workflows. Procedural aspects, including sample preparation, calibration, and quality assurance, are examined alongside rigorous validation and data analysis protocols to ensure precision and reliability. The effectiveness of purge-and-trap (P&T) techniques, as detailed in EPA Method 5030, is emphasized for VOC enrichment in aqueous samples, with an evaluation of operational conditions, trap materials, and thermal desorption parameters. Experimental investigations address the technical requirements for capillary GC columns and assess the use of hydrogen as a sustainable carrier gas and nitrogen as an alternative purge gas. These greener approaches are evaluated for their analytical efficacy and environmental advantages, offering a comprehensive framework for modernizing VOC detection while aligning with regulatory standards and sustainable analytical practices.

6.1 Introduction to EPA Method 8260D

EPA Method 8260D, "Volatile Organic Compounds by Gas Chromatography-Mass Spectrometry (GC-MS)", is an established analytical protocol for identifying and quantifying volatile organic compounds (VOCs) in a wide range of environmental samples. This method is essential for compliance with environmental regulations, as it detects VOCs that pose health risks when present in groundwater, soil, air, and other matrices. It utilizes GC for compound separation and MS for identification and quantification, providing accurate, sensitive results suitable for regulatory and research applications.

The purpose of Method 8260D is to enable precise quantification and identification of VOCs, including aliphatic and aromatic hydrocarbons, halogenated compounds, alcohols, ketones, ethers, and esters, in environmental matrices. This method is particularly useful for the environmental assessment of contaminated sites, drinking water quality, and compliance monitoring of emissions from industrial processes. Its detection capabilities span from low parts-per-billion (ppb) to parts-per-trillion (ppt) levels, making it suitable for trace-level analysis.

Method 8260D integrates GC with MS for separation and detection. Key principles include:

- **Sample Separation by GC:** The GC portion separates VOCs in a sample using an inert carrier gas (e.g., helium). Samples are introduced into the GC as a vapor, allowing compounds to elute through the column based on their chemical properties, yielding distinct retention times for each analyte.
- **Compound Identification by MS:** Compounds exiting the GC enter the MS, where they are ionized and fragmented. The MS generates a characteristic mass spectrum for each compound, with peaks corresponding to specific m/z values. Identification is achieved by comparing these spectra with those in established spectral libraries.
- **Quantification by Calibration Standards:** By creating a calibration curve based on known standards, the peak area or height response for each compound in the sample is compared, allowing accurate concentration determination. Internal standards are added to correct for potential variations in the instrument or matrix effects, ensuring data accuracy.

6.1.1 Procedural outline

Sample Preparation:

- **Water Samples:** Typically prepared using purge-and-trap (P&T) techniques. VOCs are purged from a water sample using an inert gas (e.g., nitrogen), concentrated on a sorbent trap, and then thermally desorbed into the GC-MS system.
- **Soil Samples:** VOCs are extracted using either methanol extraction or a heated headspace approach to ensure the efficient transfer of VOCs into the gas phase for GC-MS analysis.

Calibration and Instrument Tuning:

- **Calibration Standards:** Method 8260D uses multi-point calibration (usually five to seven levels) to create a robust calibration curve for each VOC. Calibration standards should cover the expected concentration range of the analytes in the samples.
- **Internal and Surrogate Standards:** Internal standards correct for variability in sample introduction and instrumental response, while surrogate standards assess method performance for each sample.
- **Instrument Tuning:** The MS system must be tuned using a tuning compound, such as bromofluorobenzene (BFB), ensuring ionization efficiency, resolution, and mass accuracy before each analytical batch.

Sample Analysis:

- Each sample is introduced into the GC-MS system, where VOCs are separated and detected sequentially. Retention times and mass spectra are compared to those of calibration standards for compound identification.
- Analysis runs include both blanks and quality control samples (e.g., matrix spikes and duplicates) to ensure reliability and validity of the results.

6.1.2 Quality Assurance and Controls

Quality assurance is critical in Method 8260D to maintain consistency, sensitivity, and accuracy. The method specifies rigorous control measures:

Instrument Tuning and Performance Check:

- **MS Tuning:** Daily tuning with BFB confirms that the MS system is correctly calibrated to ensure reproducible mass spectra. Specific ion ratios (m/z 50, 95, 96, 173) are monitored to verify that BFB spectra meet defined intensity criteria within the method.
- **Retention Time Verification:** Regularly checking the retention times of internal standards and target compounds ensures consistent separation and peak identification.

Calibration Curve and Continuing Calibration Verification (CCV):

- A multi-level calibration curve is created and verified with each batch of samples. Calibration checks include a CCV standard analyzed every 10 samples or once per analysis batch, whichever is more frequent, to assess instrument stability. CCV results must fall within $\pm 15\%$ of initial calibration levels to be valid.
- Calibration checks are also performed with a laboratory control sample (LCS), spiked with known analytes to assess accuracy and ensure that matrix effects do not influence results.

Quality Control Samples:

- **Method Blanks:** Analyzing method blanks ensures no contamination is introduced during the sample preparation and analysis process. Ideally, target analytes should be below detection limits in the blank.

- Laboratory Control Samples (LCS) and Spikes: LCS samples containing known VOC concentrations are analyzed to confirm that the method accurately quantifies analytes in standard conditions.
- Matrix Spikes (MS) and Matrix Spike Duplicates (MSD): MS and MSD samples are prepared by spiking a known amount of VOCs into a sample matrix. Recovery rates from MS/MSD samples (typically within 70-130%) indicate whether matrix interferences are present and that the method can accurately quantify VOCs within environmental samples.

Detection Limit Determination:

- The method establishes detection limits by calculating the minimum detectable concentrations for each VOC, considering the instrument's noise level, blank sample results, and calibration response. These limits are crucial for reporting data below regulatory action levels.

6.1.3 Validation and Data Analysis

Validation and data analysis in EPA Method 8260D rely on stringent criteria to confirm data reliability, accuracy, and repeatability.

Validation Criteria:

- Precision and Accuracy: Validation includes precision checks (repeatability) and accuracy checks (recovery) for both calibration standards and quality control samples. Acceptance criteria typically specify a relative standard deviation (RSD) of $\leq 15\%$ for repeated measures of calibration standards and spiked samples.
- Limit of Quantitation (LOQ) Verification: The method requires that each analyte's LOQ is verified by analyzing standard mixtures at concentrations near the LOQ to confirm that the instrument provides reproducible and reliable results at these low levels.
- Surrogate Recovery: Each sample is spiked with surrogate standards before analysis. Surrogate recoveries are monitored for each batch, with acceptable ranges usually between 70-130%. Any out-of-range surrogate recoveries necessitate reanalysis or corrective action.

Data Analysis and Interpretation:

- Compound Identification: Each VOC is identified based on its unique mass spectrum and retention time, with library matching against known standards. Identification criteria include similarity thresholds (e.g., match scores of $\geq 80\%$) to ensure confident identification.
- Quantification and Reporting: Quantification is performed by comparing the analyte response to the internal standard and referencing it to the calibration curve. Method 8260D specifies reporting in $\mu\text{g/L}$ for aqueous samples and $\mu\text{g/kg}$ for solid samples.
- Data Flags and Qualification: Data that does not meet quality criteria (e.g., poor surrogate recovery, high blank levels, or calibration drift) is flagged or qualified to indicate potential uncertainty or data limitations.

Reporting and Documentation:

- Results include the VOC concentration for each analyte along with associated quality control information. This includes detection limits, surrogate recoveries, spike recovery rates, and instrument tuning details, all of which support the validity and reliability of reported values.
- Quality Control Summary: A comprehensive quality control summary accompanies each report, detailing all quality control sample results, calibration verification, and any deviations from standard criteria. This ensures transparency and allows for accurate interpretation of the data in regulatory or research contexts.

6.2 P&T for VOCs in Aqueous Samples (EPA Method 5030)

EPA Method 5030, "Purge-and-Trap for Aqueous Samples", is a sample preparation procedure developed for extracting VOCs from aqueous samples before analysis. As a preparatory step, Method 5030 is often coupled with analytical techniques such as GC-MS (e.g., EPA Method 8260D). This method is crucial for quantifying low-level VOCs in environmental water samples, such as groundwater, surface water, and leachates, and is applied widely in environmental monitoring, compliance testing, and contamination assessment.

The purpose of EPA Method 5030 is to effectively extract and concentrate VOCs from aqueous matrices, allowing for their accurate quantification and identification. Given that VOCs readily volatilize at ambient conditions, this method leverages P&T techniques to achieve efficient analyte transfer from the aqueous phase to a GC-MS compatible form. Method 5030 applies specifically to the extraction of volatile organic pollutants such as benzene, toluene, trichloroethylene, and vinyl chloride, among others, within environmental matrices.

EPA Method 5030 operates on the P&T principle, which is designed to exploit the volatility of VOCs to isolate them from aqueous samples efficiently. This process involves three key stages: purging, trapping, and desorption.

- Purging: The aqueous sample is placed in a purge vessel, typically a 5- or 25-mL vial. An inert gas (such as helium or nitrogen) is bubbled through the sample, causing VOCs to transfer from the liquid phase into the gas phase.
- Trapping: VOCs carried in the inert gas stream are passed through a sorbent trap (often a packed bed containing materials like Tenax, silica gel, or activated carbon). This trap captures and concentrates VOCs, allowing subsequent transfer to the analytical system.
- Desorption: After purging, the sorbent trap is rapidly heated to release trapped VOCs, which are then introduced directly into the GC via a heated transfer line. This thermal desorption step enables the analytes to enter the GC column for further separation and quantification.

6.2.1 Procedural Outline of EPA Method 5030

The procedural steps in Method 5030 are as follows:

Sample Collection and Preservation:

- Water samples are collected in containers designed to minimize VOC loss (e.g., airtight glass vials with minimal headspace). The method recommends specific preservatives, such as hydrochloric acid, to lower sample pH and prevent microbial degradation of VOCs.
- Samples should be cooled to 4°C immediately after collection and processed within holding times specified by regulatory standards (typically 14 days).

Sample Preparation:

A measured volume (typically 5-25 mL) of the sample is transferred to the purge vessel. Internal and surrogate standards are added to the sample to assess matrix effects, method efficiency, and instrument performance.

P&T Process:

Purging: The inert carrier gas (helium or nitrogen) flows through the sample at a rate of 40-50 mL/min for a period of 10-15 minutes, which is typically sufficient to extract VOCs with varying volatilities.

- Trapping: The gas carrying the volatilized analytes passes through a multi-bed sorbent trap, commonly consisting of hydrophobic sorbent materials that exhibit high VOC retention. Trap cooling may be employed to prevent water vapor interference.

Thermal Desorption:

Following the purge, the trap is rapidly heated (usually to 180-220°C), causing the release of the concentrated VOCs. This desorbed gas is then directed into the GC column through a heated transfer line, minimizing analyte loss and preserving sample integrity.

6.2.2 Target VOCs

Analytes targeted by EPA 5030 and of interest for aqueous samples monitoring are listed in **Table 6.1** together with CAS Registry numbers, response levels as reported in the official EPA procedure, and stability.

EPA Method 5030 is suitable for most VOCs with boiling points below 200°C, especially those that are slightly or sparingly water-soluble. Although VOCs that are more water-soluble, such as ketones and alcohols, can be analyzed using this method, they often exhibit reduced purging efficiency and thus higher quantitation limits and variable precision, particularly when measured by GC or GC-MS. Improved purging efficiency for water-soluble compounds can be achieved by purging at elevated temperatures, such as 80°C, compared to standard temperatures (20-40°C).

Challenges arise in analyzing fuel oxygenates, such as methyl t-butyl ether (MTBE) and related compounds (e.g., ETBE, TAME, DIPE, TAEE, TAA, TBA), due to their high solubility in water, which diminishes P&T efficiency and leads to underestimations when standard procedures are used. EPA studies evaluated adjustments to analytical conditions, identifying that low-level recoveries of MTBE and related oxygenates in aqueous samples can be improved

with minor modifications to Method 5030, specifically using an 80°C purge temperature and a DB-WAX capillary column. Results showed that using an RTX-Volatile capillary column with elevated purging significantly enhanced alcohol recovery, with optimal oxygenate recovery achieved under heated purge conditions and a DB-WAX column. Consequently, for accurate quantification of oxygenates like t-butyl alcohol (TBA) and t-amyl alcohol (TAA) at low concentrations, a heated 80°C purge and a DB-WAX column are recommended. This regulation includes all analyzable compounds; however, several methodologies allow the detection and quantification of a more limited subset of target compounds.

Table 6.1. Target VOCs

Compound	CAS No. ^a	Response	Stability
Acetone	67-64-1	pp/ht	hs
Acetonitrile	75-05-8	pp	nd
Acrolein (Propenal)	107-02-8	pp	ms
Acrylonitrile	107-13-1	pp	hs
Allyl alcohol	107-18-6	ht	nd
Allyl chloride	107-05-1	c	ms
t-Amyl alcohol (TAA)	75-85-4	ht	hs
t-Amyl ethyl ether (TAEE)	919-94-8	c/ht	hs
t-Amyl methyl ether (TAME)	994-05-8	c/ht	hs
Benzene	71-43-2	c	hs
Benzyl chloride	100-44-7	c	nd
Bis(2-chloroethyl)sulfide	505-60-2	pp	nd
Bromoacetone	598-31-2	pp	nd
Bromochloromethane	74-97-5	c	hs
Bromodichloromethane	75-27-4	c	ms
Bromoform	75-25-2	c	hs
Bromomethane	74-83-9	c	hvs
n-Butanol	71-36-3	ht	nd
2-Butanone (MEK)	78-93-3	pp	hvs
t-Butyl alcohol (TBA)	75-65-0	ht	nd
Carbon disulfide	75-15-0	pp	hvs
Carbon tetrachloride	56-23-5	c	hvs
Chloral hydrate	302-17-0	pp	nd
Chlorobenzene	108-90-7	c	hvs
Chlorodibromomethane	124-48-1	c	nd
Chloroethane	75-00-3	c	ms
2-Chloroethanol	107-07-3	pp	nd
2-Chloroethyl vinyl ether	110-75-8	c	ls
Chloroform	67-66-3	c	hs
Chloromethane	74-87-3	c	hvs
Chloroprene	126-99-8	c	nd
Crotonaldehyde	4170-30-3	pp	nd
1,2-Dibromo-3-chloropropane	96-12-8	pp	ms
1,1,2-Dibromoethane	106-93-4	c	hs
Dibromomethane	74-95-3	c	hs
1,2-Dichlorobenzene	95-50-1	c	hs
1,3-Dichlorobenzene	541-73-1	c	ms
1,4-Dichlorobenzene	106-46-7	c	ms
cis-1,4-Dichloro-2-butene	1476-11-5	c	nd
trans-1,4-Dichloro-2-butene	110-57-6	pp	ls
Dichlorodifluoromethane	75-71-8	c	hs
1,1-Dichloroethane	75-34-3	c	hs
1,2-Dichloroethane	107-06-2	c	hs
1,1-Dichloroethene	75-35-4	c	hvs
cis-1,2-Dichloroethene	156-59-4	c	hs
trans-1,2-Dichloroethene	156-60-5	c	ms
1,2-Dichloropropane	78-87-5	c	hs
1,3-Dichloro-2-propanol	96-23-1	pp	nd
cis-1,3-Dichloropropene	10061-01-5	c	ls
trans-1,3-Dichloropropene	10061-02-6	c	ls

Table 6.1. Continue

Compound	CAS No. ^a	Response	Stability
1,2,3,4-Diepoxybutane	1464-53-5	c	nd
Diethyl ether	60-29-7	c	nd
Diisopropyl ether (DIPE)	108-20-3	c/ht	hs
1,4-Dioxane	123-91-1	ht/pp	nd
Ethylbenzene	100-41-4	c	hvs
Ethylene oxide	75-21-8	pp	nd
Ethyl methacrylate	97-63-2	c	ms
Ethyl tert-butyl ether (ETBE)	637-92-3	c/ht	hs
Hexachlorobutadiene	87-68-3	c	ms
2-Hexanone	591-78-6	pp	hvs
Iodomethane	74-88-4	c	nd
Isobutyl alcohol	78-83-1	ht/pp	nd
Isopropylbenzene	98-82-8	c	ms
Malononitrile	109-77-3	pp	nd
Methacrylonitrile	126-98-7	pp	hs
Methylene chloride	75-09-2	c	hs
Methyl methacrylate	80-62-6	c	ms
4-Methyl-2-pentanone (MIBK)	108-10-1	pp	ms
Methyl tert-butyl ether (MTBE)	1634-04-4	c/ht	hs
Naphthalene	91-20-3	c	ms
Nitrobenzene	98-95-3	c	nd
2-Nitropropane	79-46-9	c	nd
N-Nitroso-di-n-butylamine	924-16-3	pp	nd
Paraldehyde	123-63-7	pp	nd
2-Pentanone	107-87-9	pp	nd
2-Picoline	109-06-8	pp	nd
1-Propanol	71-23-8	ht/pp	nd
2-Propanol	67-63-0	ht/pp	nd
β-Propiolactone	57-57-8	pp	nd
Propionitrile (ethyl cyanide)	107-12-0	ht	nd
n-Propylamine	107-10-8	c	nd
Styrene	100-42-5	c	hvs
1,1,1,2-Tetrachloroethane	630-20-6	c	hs
1,1,2,2-Tetrachloroethane	79-34-5	c	nd
Tetrachloroethene	127-18-4	c	ms
Toluene	108-88-3	c	hs
o-Toluidine	95-53-4	pp	nd
1,2,4-Trichlorobenzene	120-82-1	c	hs
1,1,1-Trichloroethane	71-55-6	c	ms
1,1,2-Trichloroethane	79-00-5	c	hs
Trichloroethene	79-01-6	c	ms
Trichlorofluoromethane	75-69-4	c	ls
1,2,3-Trichloropropane	96-18-4	c	ls
Vinyl acetate	108-05-4	c	ls
Vinyl chloride	75-01-4	c	hvs
o-Xylene	95-47-6	c	hvs
m-Xylene	108-38-3	c	hvs
p-Xylene	106-42-3	c	hvs

^a Chemical Abstract Service Registry Number

c = Adequate response by this technique

ht= Method analyte only when purged at elevated temperature, e.g., 80°C

pp= Poor purging efficiency resulting in high quantitation limits. Use of an alternative sample preparative method is strongly recommended. May be amenable to purging at elevated temperature.

nd= Not determined

hs= High stability in preserved water samples (> 60 days). Longer holding times may be appropriate, see Method 5035, Appendix A, Table A.1 footnote and ref. 47 for additional information

ms= Medium stability in preserved water samples (15 - 60 days). Longer holding times may be appropriate, see Method 5035, Appendix A.

ls= Low stability in preserved water samples (< 14 days), analyses should be performed as soon as possible. May be degraded if acid preserved.

hvs= Highly variable stability in preserved water samples. Longer holding times may be appropriate, see Method 5035, Appendix A.

6.2.3 P&T recommended conditions

The P&T system is a critical component in preparing aqueous samples for VOCs analysis. This system introduces an inert gas to purge VOCs from the sample and traps them for desorption into the GC. The system can either operate manually or automatically, transferring a precise sample volume (e.g., 5 or 25 mL), and adds surrogates, matrix spikes, and internal standards as necessary. The system specifications and operational guidelines, according to EPA Method 5030 are as follows:

Purging Chamber

The purging chamber is designed to accept 5-mL sample volumes with a water column depth of at least 3 cm. The chamber headspace, situated above the water column and below the trap, must have a volume under 15 mL. The purge gas is introduced at the chamber base, no more than 5 mm from the bottom, forming fine bubbles (<3 mm diameter) that facilitate VOC transfer. The sample purger meets these criteria; alternative devices may be used if validated for comparable performance.

Trap Specifications and Conditioning

The recommended trap is a 25-cm tube with an internal diameter of 0.267 cm, consisting of three adsorbent layers: 2,6-diphenylene oxide polymer, silica gel, and coconut charcoal, each comprising one-third of the trap. To extend the trap's lifespan, a 1.0-cm methyl silicone-coated packing can be added to the inlet. If low-boiling fluorocarbons (e.g., dichlorodifluoromethane) are not target analytes, the charcoal can be omitted, and the polymer filling increased. For VOCs with boiling points above 35°C, both silica gel and charcoal can be omitted, with the polymer filling the trap entirely. Initial conditioning involves backflushing the trap overnight at 180°C with a minimum 20 mL/min inert gas flow, while daily conditioning requires a 10-minute backflush at 180°C.

Desorber

The desorber must heat the trap rapidly to 180°C, with the polymer section of the trap restricted to 180°C and other sections to a maximum of 220°C during bake-out. The desorber unit can be used as a standalone component or directly coupled with a GC.

Trap Packing Materials and Alternatives

The following materials are used for trap packing:

- 2,6-Diphenylene oxide polymer: 60/80 mesh, e.g., Tenax GC.
- Methyl silicone packing: 3% OV-1 on Chromosorb-W, 60/80 mesh.
- Silica gel: 35/60 mesh, grade 15.
- Coconut charcoal: Prepared by crushing through a 26-mesh screen.

Alternative materials, including hydrophobic carbon molecular sieves and graphitized carbon black, can be used provided they maintain comparable retention and desorption properties to the standard Tenax/silica gel/charcoal configuration. Viable alternatives include Carboxpack and Carboxieve-based combinations (e.g., VOCARB 3000 and VOCARB 4000), which require desorption temperatures between 245-270°C. Higher desorption

temperatures may lead to analyte decomposition, especially for compounds like 2-chloroethyl vinyl ether and 2,2-dichloropropane.

Thermal decomposition must be monitored daily using a check standard with chloromethane and bromomethane formation, particularly if levels exceed 0.5 µg/L, as high decomposition levels necessitate trap replacement and recalibration.

Heated Purging Chamber

The system requires a heater or oil bath that maintains the purge chamber temperature within $\pm 1^{\circ}\text{C}$, adjustable from ambient to 100°C , to optimize the purging efficiency of VOCs, particularly those with higher water solubility.

This setup ensures VOCs are efficiently purged, trapped, and delivered for accurate GC analysis, providing reliable results for environmental VOC monitoring.

Table 6.2 Reports the P&T conditions as recommended by EPA 5030 method for the VOCs determination in environmental samples as for EPA 8015, EPA 8021, and EPA 8260.

Table 6.2: P&T operating parameters

	Analysis Method	
	8015	8021/8260
Purge gas	N ₂ or He	He or N ₂
Purge gas flow rate (mL/min)	20	40
Purge time (min)	15.0 $\pm$ 0.1	11.0 $\pm$ 0.1
Purge temperature ($^{\circ}\text{C}$)	Ambient ²	Ambient ²
Desorb temperature ($^{\circ}\text{C}$)	180	180
Backflush inert gas flow (mL/min)	20-60	20-60 ¹
Desorb time (min)	1.5	4

¹ The desorption flow rate for Method 8021 with a wide bore capillary column will optimize at approximately 10 to 15 mL/minute.

² Purge temperature is 80°C for Methods 8015 and 8260 when analyzing for TBA, TAA or other fuel oxygenate alcohols.

6.3 Capillary GC columns requirements

For VOC analysis using GC, any capillary column that meets the performance standards of the specified determinative method can be utilized. The selection of a column, as well as its operational conditions and retention times, should align with the specific requirements of the method.

Wide-Bore Columns

Wide-bore columns can handle standard gas flows from the trap during thermal desorption, allowing chromatography to commence concurrently with desorption. When working with MS that have limited pumping capacity, an additional interface between the column and the MS may be needed. Acceptable interfaces include an open split interface, an all-glass jet separator, or a cryogenic device. The interface type and its settings can significantly influence detection limits, making proper selection and calibration essential. Other interfaces are permissible if they meet the method's performance standards.

Narrow-Bore Columns

Using a narrow-bore column necessitates reduced gas flows, typically around 2-4 mL/min. Due to these low desorption flow rates, early-eluting analytes must be refocused to elute in narrow bands, which can be achieved using a cryogenic interface. This type of interface generally employs liquid nitrogen to condense the desorbed analytes in a narrow band on an uncoated fused silica precolumn. Once desorption from the trap is complete, rapid heating under carrier gas transfers the analytes onto the analytical column, with the column end positioned within a few millimeters of the MS ion source.

One challenge with cryogenic interfaces is the potential for water in the trap to desorb as ice, which can obstruct the interface. This blockage impairs sensitivity and reduces chromatographic resolution. Observing low surrogate recoveries may indicate ice formation in the interface, signaling a need for remedial measures to restore analytical performance.

6.4 *Use of hydrogen as alternative carrier gas*

Hydrogen can serve as an acceptable carrier gas for VOC analysis, but adjustments are required to achieve performance comparable to that obtained with helium.

Hydrogen Purity and Compatibility

For optimal performance, the highest purity hydrogen ($\geq 99.999\%$) should be used, sourced from a hydrogen generator or high-purity cylinders to minimize contamination from hydrocarbons or moisture. Using stainless steel tubing, rather than copper, can improve the durability of the gas lines, as copper tubing may degrade and become brittle over time when exposed to hydrogen. Ensure that the MS ion source is compatible with hydrogen; contacting the MS manufacturer is advisable to confirm this compatibility. The pressure in the MS source may need to be reduced to prevent chemical ionization or other unwanted reactions. This reduction can be achieved by using narrower bore columns (≤ 0.18 mm ID), decreasing the flow rate to the MS, or incorporating internal MS vacuum pumps (e.g., turbo pumps) with higher efficiency. Systems with internal diffusion pumps are typically incompatible with hydrogen.

System Cleaning and Interference Management

Hydrogen's "scrubbing" effect cleans metal surfaces in the analytical system, which may increase hydrocarbon background interferences. A system bake-out with high hydrogen flow can help reduce these interferences to acceptable levels. Installing new filters in the gas lines, or omitting them if gas purity allows, is also recommended to avoid introducing impurities from the filters.

Calibration and Response Factor Adjustments

When using hydrogen, the response of target analytes can decrease by approximately 2-5 times compared to helium. The relative response factors (RFs), originally established with helium, may not apply to hydrogen due to these response reductions. Laboratories should therefore either establish new RF criteria or consult RFs provided

by the instrument manufacturer. The reactivity of analytes with hydrogen can vary based on the specific instrument settings, so a stability assessment for target analytes should be included as part of the demonstration of capability (DOC) process.

Quality Control and Safety Measures

All quality control (QC) steps of the method should be re-validated with hydrogen prior to analyzing samples, including any solvent modifications for calibration standards or extracts. Hydrogen's flammability necessitates additional safety precautions to prevent accumulation of explosive gas concentrations. This includes venting lines from the GC inlet and MS rough pump to exhaust systems, performing leak tests on gas line connections, and turning off hydrogen at the source before opening gas lines or venting the MS for maintenance. Users should consult their instrument manufacturer's safety guidelines to implement additional hydrogen safety protocols as needed.

6.5 Italian legislation framework: Decree 156/06

Italian Decree 152/06, known as the Environmental Code, is a foundational legislative framework aimed at protecting Italy's natural resources, ecosystems, and public health. This comprehensive law encompasses various environmental aspects, including air, soil, and especially water protection, making it an essential tool in Italy's environmental governance. The decree regulates pollutant discharges, sets environmental quality standards, and outlines the responsibilities of entities involved in environmental management. By establishing stringent guidelines for the quality of natural resources and specific pollution controls, Decree 152/06 seeks to preserve ecological integrity and ensure that Italy complies with EU directives, aligning national policies with the European Green Deal and sustainability goals.

In terms of water protection, Decree 152/06 is particularly significant, as it mandates systematic controls on surface water, groundwater, and drinking water to prevent contamination from industrial, agricultural, and urban sources. This legislation has strict standards for permissible contaminant levels, which public authorities must monitor and manage to prevent environmental degradation and health risks. To ensure water quality, the decree defines specific parameters for testing and requires routine sample collection and analysis. Such mandatory controls are essential for early detection of pollutants like nitrates, heavy metals, and organic compounds, which pose serious health risks if left unchecked. Through these standards, the decree facilitates early interventions and mitigates the adverse effects of contamination on both human health and aquatic ecosystems.

The decree also highlights the importance of compliance enforcement, empowering regional and national authorities to conduct inspections and impose penalties on violators. Regular testing of water samples from rivers, lakes, aquifers, and treatment facilities forms the backbone of these enforcement measures. This regulatory oversight is crucial, not only for maintaining public trust in water safety but also for ensuring that Italy's water bodies can meet both national and EU environmental standards. By enforcing systematic water quality controls, Decree 152/06 plays a vital role in supporting sustainable water resource

management, which is essential in addressing Italy's environmental challenges, such as water scarcity, pollution, and the impacts of climate change on water quality.

The Italian Decree 152/06 establishes precise detection limits for VOCs in water, with threshold concentrations set to mitigate human and ecological risks associated with these compounds. These limits, defined by specific maximum allowable concentrations for each VOC (e.g., benzene, toluene, chlorinated solvents like trichloroethylene), reflect the compound's toxicological profile and environmental persistence. Analytical testing protocols must ensure quantification of VOCs at levels typically in the low ppb or ppt range, depending on the compound's volatility and regulatory requirements for drinking water, groundwater, and surface water quality. These threshold levels are strictly in line with the European Union's Drinking Water Directive and Water Framework Directive, ensuring harmonized standards and compliance with EU environmental objectives.

Decree 152/06 mandates that results exceeding VOC detection limits trigger immediate regulatory actions, such as additional sampling, potential source identification, and contamination mitigation efforts. The established limits enforce not only routine monitoring of VOC levels but also the application of risk-based corrective actions in cases of non-compliance. By enforcing these precise detection limits, the decree supports both public health safety and long-term environmental stewardship, as VOCs can bioaccumulate and exert chronic toxic effects in aquatic systems. For accurate reporting and comparability, laboratories must employ validated, decree-compliant methods that meet these detection criteria, ensuring that regulatory agencies can implement swift responses to potential exceedances in monitored water bodies.

6.6 Nitrogen as purge gas

The scarcity of helium, coupled with increasing demand from the medical, scientific, and industrial sectors, is driving up the price of this increasingly rare resource. In the P&T extraction technique, the purge phase involves stripping volatile analytes from the sample matrix using helium, which then concentrates the analytes on an analytical trap. However, the rising costs and challenges in helium procurement have necessitated the investigation of alternative purge gases. EPA Method 5030 and related regulations reference both helium and nitrogen as potential purge gases for the analysis of VOCs using P&T systems. Nevertheless, helium has been predominantly employed in nearly all applications, leveraging its superior inertness and efficacy in analyte recovery. This section explores the transition to nitrogen as a purge gas, emphasizing its advantages as a renewable, more readily available, and environmentally sustainable alternative. The results presented evaluate nitrogen's performance in compliance with EPA guidelines, offering insights into its potential to maintain analytical precision while supporting green chemistry principles.

6.6.1 Evaluation of requested parameters

The analyses were conducted using a Centurion-Evolution 2 EST Analytical purge-and-trap (EV2 P&T) system coupled to an Agilent 8890/5977 GC-MS. The Centurion sampler was equipped with a nitrogen gas restrictor and operated in "water" mode using a syringe. The EV2 P&T, fitted with a Vocab 3000 trap, was calibrated to work with nitrogen purge flows.

The EV₂ P&T transfer line was connected to the GC inlet for the transfer and separation of analytes. A 30 m x 0.25 mm x 1.4 µm DB-624UI column was installed in the GC, and the mass detector was configured to operate in Selected Ion Monitoring (SIM)/SCAN mode.

In **Tables 6.3, 6.4** and **6.5** P&T Concentrator Evolution 2, Autosampler P&T Centurion and GC-MS 8890/5977C parameters are reported, respectively.

Table 6.3: P&T Concentrator Evolution 2 experimental parameters

Parameters	Set values
Trap Type	Vocarb 3000
Valve Oven Temperature	130°C
Transfer Line Temperature	130°C
Trap Temperature	35°C
Moisture Reduction Trap (MoRT) Temperature	39 °C
Purge Time	11 min
Purge Flow	40mL/min
Dry Purge Temperature	Off
Dry Purge Flow	40mL/min
Dry Purge Time	2.0 min
Desorb Pressure Control	On
Desorb Pressure	7psi
Desorb Time	0.5 min
Desorb Preheat Delay	5 sec
Desorb Temperature	250°C
Moisture Reduction Trap (MoRT) Bake Temperature	210°C
Bake Temperature	260°C
Spurge Vessel Bake Temperature	110°C
Bake Time	8 min
Bake Flow	85 mL/min

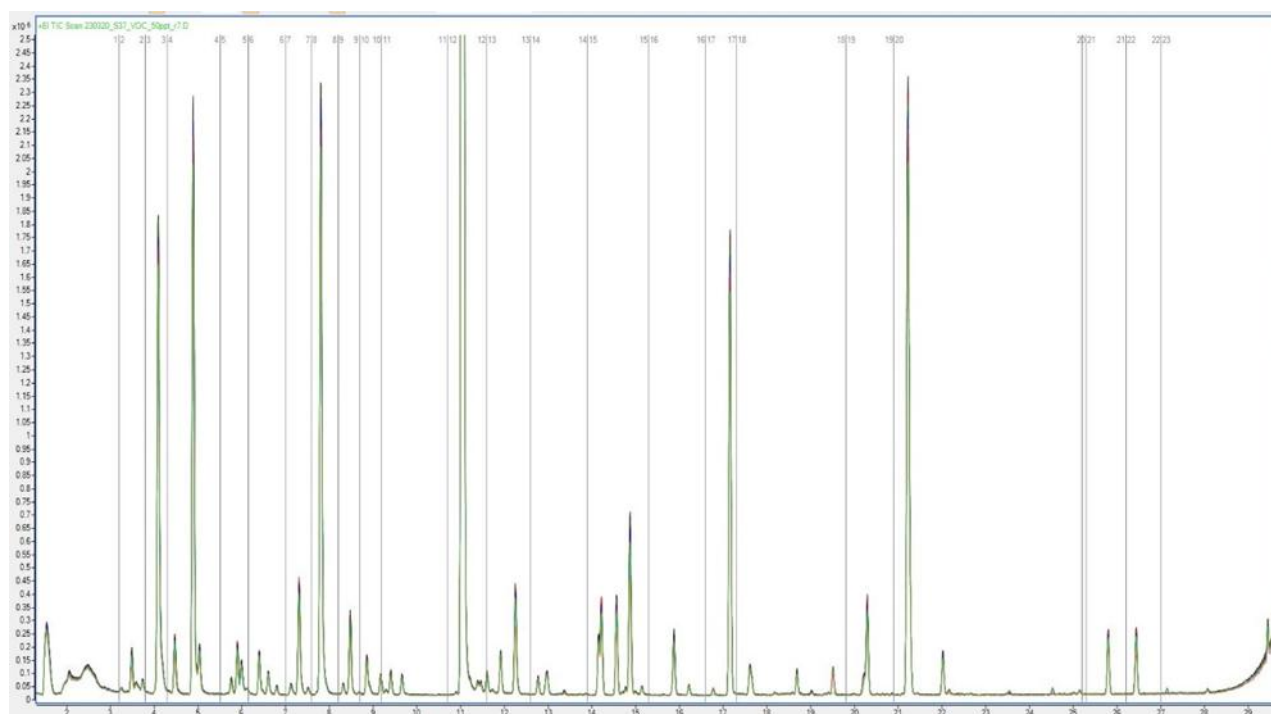
Table 6.4: Autosampler P&T Centurion experimental parameters

Parameters	Set values
Sample Type	Water
Sample Fill Mode:	Syringe
Sample Volume	25 mL
Sample Prime Time:	7 sec
Sample Transfer Time:	35 sec
Syringe Rinse	On/7 mL
Number of Syringe Rinses	2
Sample Loop Rinse	On/10 sec
Sample Loop Sweep	Time 5 sec
Number of Spurge Rinses	Syringe/2
Rinse Volume	25 mL
IS volume	5 µL

Table 6.5: GC-MS 8890/5977C experimental parameters

Parameters	Set values
Inlet	Split/Splitless
Inlet Temperature	220°C
Inlet Head Pressure	10.245 psi
Mode	Split
Split Ratio	10:1
Column	DB-624UI 30m x 0.25mm I.D. 1.4µm film thickness
Oven Temperature Program	35°(1') to 220°C (1.33') @15°C/min
Column Flow Rate	1.3 mL/min
Gas	Helium
Total Flow	17.3mL/min
Source Temperature	230°C
Quad Temperature	150°C
MS Transfer Line Temperature	180°C
Scan Range m/z	35-300 m/z
Scans	5.2 scans/sec
Solvent Delay	0.7 min

In **Figure 6.1** the chromatogram illustrates an example of the separation of the target analytes under the described experimental conditions. Additionally, all SIM acquisition windows are displayed, as required by official methods.

**Figure 6.1:** Chromatogram of target analytes mixture

For the assessment of repeatability, multiple analyses were performed at different analyte concentrations. Specifically, ten analytical replicates were conducted at 50 ppt, 1 ppt and, 0.1 ppt. The last concentration was selected to examine repeatability at a level 10 times lower than the limits prescribed by Legislative Decree 152/06. For this reason, **Table 6.6** lists the results as %RSD at 0.1 ppt for the two compounds regulated at a 1 ppt threshold, and at 1 ppt for all target analytes. At a concentration of 1 ppt, some RSD% values exceed the 15% threshold specified by the regulation. However, these compounds are regulated at higher concentration levels. Conducting repeatability experiments at an increased concentration of

50 ppt (still below the regulatory limits) results in all RSD% values falling within the acceptable 15% range.

Table 6.6: RSD% at different concentration levels for target analytes.

Analyte	$\mu\text{g/L(ppb)}^*$	RSD %		
		10 replicates at 0.1 ppt	10 replicates at 1 ppt	10 replicates at 50 ppt
Chloromethane	1.5	-	63.9	2.2
Vinyl chloride	0.5	-	15.6	1.0
1,1-Dichloroethylene	0.05	-	5.1	1.5
trans-1,2-Dichloroethylene	60	-	5.5	1.1
1,1-Dichloroethane	810	-	3.8	1.1
cis-1,2-Dichloroethylene	60	-	3.6	1.0
Chloroform	0.15	-	8.1	1.2
Benzene	1	-	6.9	1.2
1,2-Dichloroethane	3	-	5.5	1.4
(I.S.) Fluorobenzene	-	-	-	-
Trichloroethylene	1.5	-	4.2	0.6
1,2-Dichloropropane	0.15	-	2.7	1.0
Bromodichloromethane	0.17	-	1.9	0.9
Toluene	15	-	29.6	1.3
1,1,2-Trichloroethane	0.2	-	2.8	1.8
Tetrachloroethylene	1.1	-	7.0	1.8
Dibromochloromethane	0.13	-	1.4	1.7
1,2-Dibromoethane	0.001	2.7	2.0	1.6
Chlorobenzene	40	-	2.5	1.4
Ethylbenzene	50	-	21.8	2.9
(m+p) Xylenes	10	-	25.9	2.9
Styrene	25	-	63.7	2.5
Bromoform	0.3	-	2.9	1.4
(I.S.) p-Bromofluorobenzene	-	-	-	-
1,1,2,2-Tetrachloroethane	0.05	-	2.5	1.0
1,2,3-Trichloropropane	0.001	2.8	2.1	2.1
1,4-Dichlorobenzene	0.5	-	9.8	2.1
(I.S.) 1,2-Dichlorobenzene-D ₄	-	-	-	-
1,2-Dichlorobenzene	270	-	1.8	2.7
1,2,4-Trichlorobenzene	190	-	3.0	2.1
Hexachloro-1,3-butadiene	0.15	-	3.2	2.4
1,2,4,5-Tetrachlorobenzene	1.8	-	5.4	3.3

* The threshold contamination concentrations were extracted from Legislative Decree 152/06.

In addition, **Figures 6.2** shows, by way of example, the overlapped chromatograms of the analytical replicates at 1 ppt (limit required by law) for 1,2-dibromoethane (**A**, ion 107 m/z) and 1,2,3-trichloropropane (**B**, ion 75 m/z) respectively.

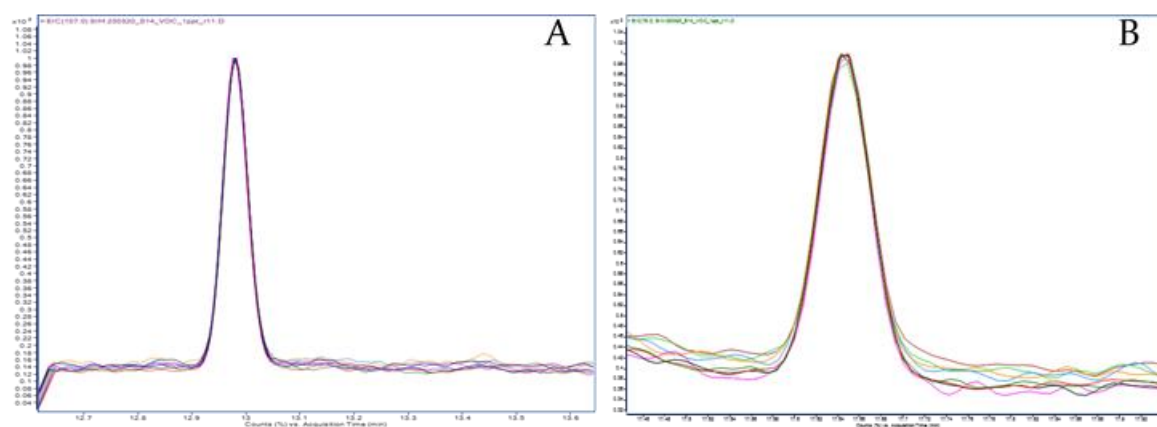


Figure 6.2: Overlapped chromatograms of the analytical replicates at 1 ppt for 1,2-dibromoethane (**A**) and 1,2,3-trichloropropane (**B**).

One of the evaluation criteria is the recovery of surrogate standards. Each sample is spiked with surrogates before analysis, and their recovery is monitored for every analytical batch. The two compounds selected as surrogates were p-bromofluorobenzene and 1,2-dichlorobenzene-D₄, which were added to each sample by injecting 5 µL of a solution at a concentration of 1000 ppm. Acceptable recovery ranges are typically set between 70% and 130%. Surrogate recoveries were to fall outside these limits, reanalysis or corrective actions would be required to ensure data accuracy and reliability. As shown in the **Table 6.7**, the results fall within the established limits.

Table 6.7: Results of the recoveries of surrogate standards (p-bromofluorobenzene, 1,2-dichlorobenzene-D₄) in calibration curves points

ppt	p-Bromofluorobenzene	1,2-Dichlorobenzene-D ₄
1	102.3	98.1
1	101.3	97.9
1	101.2	97.2
2.5	101.4	100.6
2.5	101.1	100.0
2.5	100.8	101.0
5	99.7	99.6
5	100.2	100.5
5	98.3	98.7
10	97.8	97.8
10	97.9	97.8
10	98.8	99.1
25	98.9	99.1
25	97.6	99.1
25	97.9	98.1
50	102.0	105.9
50	101.2	104.3
50	101.4	104.8

In addition, a series of 25 blanks was performed using helium as the purge gas and nitrogen as the purge gas. The stability of target compounds was examined to determine whether the change in purge gas affected their stability over time. The results are shown in **Figures 6.3** for helium and nitrogen, respectively. Despite the greater fluctuation observed with nitrogen compared to the reference gas, the %RSD values remain below 5%, demonstrating acceptable stability.

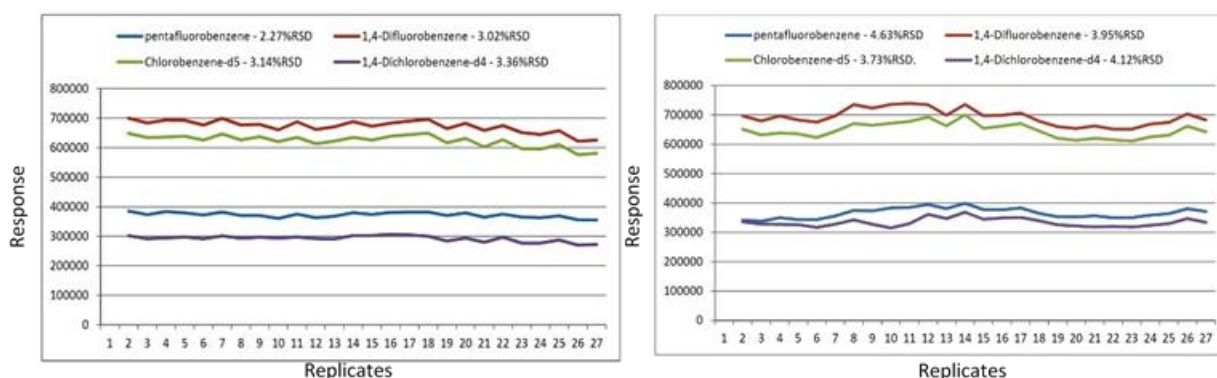


Figure 6.3: Responses of target compounds in 25 blanks: from right to left, using Helium and Nitrogen as purge gases.

Method 8260D utilizes multi-point calibration (typically five to seven levels) to construct a robust calibration curve for each VOC. The calibration standards should include the expected concentration range of the analytes in the samples. To assess linearity, analyses were performed at several concentration levels, specifically at 6 levels: 1 ppt, 2.5 ppt, 5 ppt, 10 ppt, 25 ppt, and 50 ppt. Each calibration level was conducted in triplicate, resulting in a total of 18 points used to generate the calibration curve. The chosen calibration curve type was linear. The R^2 values for the calibration curves for each analyte are presented in **Table 6.8**.

Table 6.8: R^2 values for the calibration curves for all target compounds.

Analite	R^2
Chloromethane	0.988
Vinyl chloride	0.989
1,1-Dichloroethylene	0.994
trans-1,2-Dichloroethylene	0.994
1,1-Dichloroethane	0.997
cis-1,2-Dichloroethylene	0.996
Chloroform	0.998
Benzene	0.999
1,2-Dichloroethane	0.993
Trichloroethylene	0.998
1,2-Dichloropropane	0.999
Bromodichloromethane	0.997
Toluene	0.999
1,1,2-Trichloroethane	0.998
Tetrachloroethylene	0.992
Dibromochloromethane	0.998
1,2-Dibromoethane	0.995
Chlorobenzene	0.997
Ethylbenzene	0.998
(m+p) Xylenes	0.999
Styrene	0.997
Bromoform	0.996
1,1,2,2-Tetrachloroethane	0.995
1,2,3-Trichloropropane	0.997
1,4-Dichlorobenzene	0.997
1,2-Dichlorobenzene	0.998
1,2,4-Trichlorobenzene	0.988
Hexachloro-1,3-butadiene	0.997
1,2,4,5-Tetrachlorobenzene	0.996

One of the crucial aspects of this methodology is the determination of the signal-to-noise ratio (S/N). For its evaluation, analyses were conducted at various concentrations. Typically, a concentration 10 times lower than the regulatory limit is assessed to verify the system's performance. The method determines detection limits by calculating the minimum detectable concentrations for each VOC, taking into account the instrument's noise level, blank sample results, and calibration response. These limits are essential for ensuring accurate reporting of data below regulatory action levels. The focus was on the compounds with the lowest detection limit (1 ppt) *i.e.* 1,2-dibromoethane and 1,2,3-trichloropropane.

Figure 6.4 illustrates the overlap of five analytical replicates (colored lines) of a 0.1 ppt solution and a blank (black line) extracting ion 75 for 1,2,3-trichloropropane (A) and ion 107 for 1,2-dibromoethane. (B).

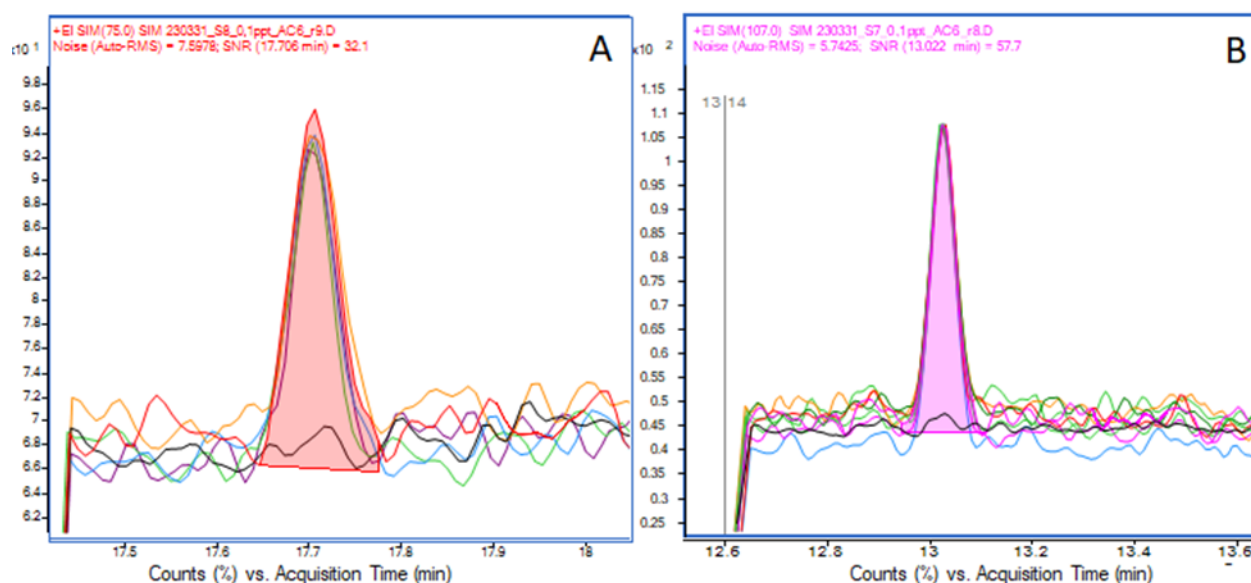


Figure 6.4: Overlap of five analytical replicates (colored lines) of a 0.1 ppt solution and a blank (black line) for 1,2,3-trichloropropane (A) and for 1,2-dibromoethane (B)

The S/N was automatically calculated using the Auto-RMS algorithm in MassHunter Qualitative Version 12.0 software (Agilent Technologies). In both cases, the results were satisfactory, and as shown in the picture of **Figure 6.4**, the distinction between the peaks and the blank analysis is clear. The resulting S/N values for 1,2-dibromoethane and 1,2,3-trichloropropane were 57.7 and 32.1, respectively.

In addition, the Instrument Detection Limit (IDL) was calculated based on the reference concentration of compounds 1,2-dibromoethane and 1,2,3-trichloropropane. Eight analytical replicates were performed at a concentration of 1 ppt. The chromatographic area for each replicate is reported in **Table 6.9**.

Subsequently, the IDL was calculated according to the formula specified in **Equation 6.1**.^[1]

$$IDL = t \times \frac{\%RSD}{100} \times \text{amount measured}$$

Where: t is the t-student

%RSD is the Relative Standard Deviation expressed in percentage

Table 6.9: Parameters and chromatographic areas used to calculate the IDL of the analytes regulated at 1 ppt.

	1,2-Dibromoethane	1,2,3-Trichloropropane
Number of replicates	8	
Confidence	0,99	
t(p,N-1)	2,998	
Amount	1 ppt	
Area replicates 1	3586	2338
Area replicates 2	3287	2233
Area replicates 3	3539	2295
Area replicates 4	3313	2117
Area replicates 5	3539	2317
Area replicates 6	3479	2159
Area replicates 7	3297	2105
Area replicates 8	3317	2093
RSD %	3.7	4.6

Based on **Equation 6.1** and adopting a confidence level of 0.99, the Student's $t(p, N-1)$ value for 8 analytical replicates is 2.998. As a result, the IDL values for analytes 1,2-dibromoethane, and 1,2,3-trichloropropane are determined to be 0.11 ppt and 0.14 ppt, respectively.

In conclusion, the implementation of nitrogen as a purge gas offers significant economic benefits. During the P&T sampling phases, helium consumption is as follows: 440 mL during the purging phase (11 minutes at 40 mL/min) and 1,000 mL during the bake-up phase (10 minutes at 100 mL/min). Replacing helium with nitrogen results in a savings of approximately 1,500 mL of helium per sample.

Considering that generally automated P&T system solution can operate continuously without operator supervision, with an analysis time of about 30 minutes per sample, the maximum productivity for VOC analysis is 48 samples per day. Based on this, the estimated helium consumption is about 72 liters per day, 500 liters per week, and 2,000 liters per month, which is equivalent to roughly one cylinder every four months, resulting in an annual savings of about 2.5 cylinders.

The P&T-GC-MS with nitrogen as purge gas meets the performance standards required by EPA Method 8260 and the legal limits specified by DLGS152/06. Calibration results meet all method criteria using both purge gases, with comparable MDL values. Excellent precision and accuracy were achieved with both helium and nitrogen as purge gases. Although nitrogen, due to its larger molecular size, extracted slightly more moisture during the purge phase compared to helium, this did not affect the experimental results.

6.7 Hydrogen as carrier gas in GC-MS

The analysis of VOCs in water traditionally relies on GC-MS, where helium is the carrier gas of choice due to its inertness and favorable chromatographic properties. However, the increasing scarcity and cost of helium have driven interest in hydrogen as a more sustainable, greener, and accessible alternative. While hydrogen offers advantages such as reduced analysis times and the ability to be generated on-site, its use in MS introduces specific challenges that necessitate careful evaluation.

Hydrogen as a carrier gas in MS presents several challenges that could critically impact analyses requiring ultra-low LOQ, such as 1 ppt for certain compounds in EPA Method 8260D. One of the main concerns is the increased background noise associated with hydrogen, which can significantly reduce the S/N. This issue becomes critical when detecting trace levels of analytes where even minor interferences can compromise data reliability. Additionally, hydrogen may promote secondary ionization reactions, leading to the formation of spurious peaks or spectral artifacts that further complicate the interpretation of complex mixtures. A specific concern in the current application is the potential hydrogenation of unsaturated compounds, many of which could be target analytes. Such reactions might alter the molecular structure of these compounds, affecting both qualitative and quantitative outcomes. These drawbacks highlight the need for careful evaluation of hydrogen's suitability for highly sensitive and precise analytical requirements.

To address these concerns, the present study explores the implementation of hydrogen as a carrier gas in the P&T-GC-MS method for VOC analysis. Utilizing the innovative HydroInert™ source, which has been specifically designed to mitigate the reactive properties of hydrogen, this work evaluates the system's performance in terms of chromatographic response, spectral fidelity, and sensitivity, comparing it directly to the conventional use of helium. By systematically assessing these parameters, this research aims to provide insights into the practical applicability and limitations of hydrogen as a carrier gas for environmental and industrial VOC analysis.

6.7.1 Evaluation of requested parameters

The P&T conditions utilized in this study were those previously tested and optimized, with results detailed in the preceding section (See section 6.6.1). The following results exclusively address the translation of the chromatographic method from helium to hydrogen as the carrier gas.

A 20 m × 0.18 mm × 1 µm DB-624UI column was installed in the GC, and the mass detector was configured to operate in SIM/SCAN mode. GC-MS 8890/5977C parameters are reported in **Table 6.10**.

Table 6.10: GC-MS 8890/5977C parameters

Parameters	Set values
Inlet	Split/Splitless
Inlet Temp.	220°C
Inlet Head Pressure	8.5032 psi
Mode	Split
Split Ratio	4:1
Column	DB-624UI 20 m x 180 µm x 1 µm film thickness
Oven Temp. Program	35°(1') to 80°C (2.75') @20°C/min to 240°C (1') @20°C/min
Column Flow Rate	1.0 mL/min
Gas	Hydrogen
Total Flow	5mL/min
Source Temp.	230°C
Quad Temp.	150°C
MS Transfer Line Temp.	220°C
Scan Range m/z	45-260 m/z
Scans	6.2 scans/sec
Solvent Delay	0.1 min

Through the use of hydrogen as the carrier gas and the application of method translation, the chromatographic analysis duration was reduced from 30.7 minutes to 15 minutes. **Figure 6.5** shows an example of an analysis conducted with hydrogen as carrier gas.

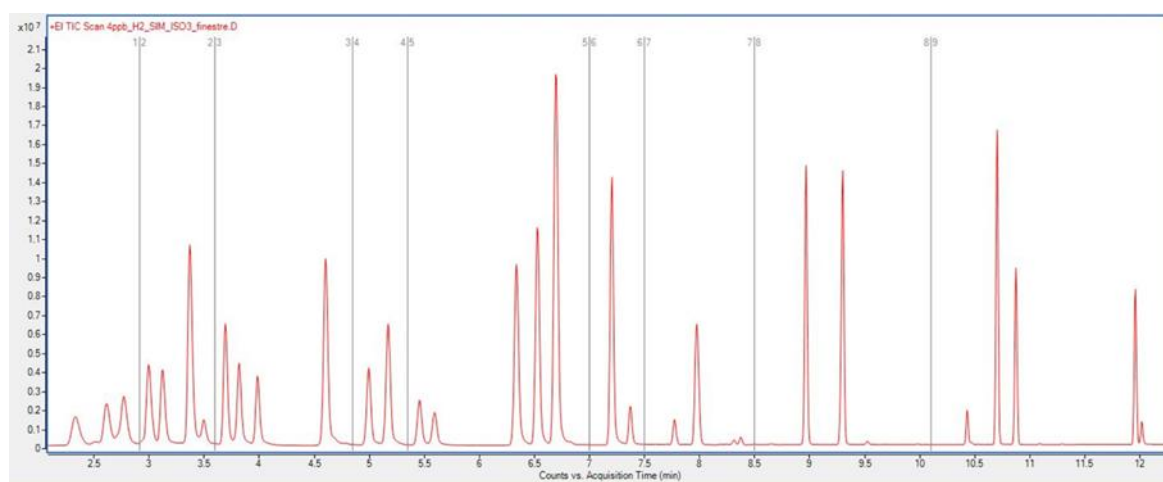


Figure 6.5: Chromatogram of target analytes mixture with hydrogen as carrier gas

Proceeding systematically, the initial evaluation focused on assessing the spectral quality of the analytes. The analysis of VOCs of various types in scan/SIM mode demonstrated that the use of hydrogen as the carrier gas had a minimal impact on spectral quality. For the compounds under investigation, the spectral profiles were largely preserved, showing a strong alignment with reference spectra obtained using helium and with commercial libraries. This allowed for a direct translation of the method without requiring significant modifications. As shown in **Table 6.11**, the responses of both quantifier and qualifier ions exhibited no significant variations when compared to the standard methodology employing helium. In the case of target analyses of this type, performed in SIM mode with the aid of certified standards, any discrepancies that might arise could be effectively addressed by optimizing the selection of ions or their relative ratios. Data were acquired in SIM mode with an Extractor™ MS source with both helium and hydrogen.

Table 6.11: Comparison of helium and hydrogen carriers, respectively, for the response of the qualifier ions versus quantifier ion signal (SIM). The results were obtained with the Ion Extractor™ source.

Compound	Helium			Hydrogen		
	Qualifier 1	Qualifier 2	Qualifier 3	Qualifier 1	Qualifier 2	Qualifier 3
trans-1,2-Dichloroethylene	89	66		88		
1,1-Dichloroethane	33	15		33	16	
cis-1,2-Dichloroethylene	66	80		66	81	
Chloroform	67			67		
Benzene						
1,2-Dichloroethane	23			23		
Trichloroethylene	143	66	146	122	66	126
1,2-Dichloropropane	7			7		
Bromodichloromethane	66	12		66	10	
Toluene	154			149		
1,1,2-Trichloroethane	132	66		134	66	
Tetrachloroethylene	76	73	130	85	81	129
Dibromochloromethane	76			78		
1,2-Dibromoethane	97	5		98	4	
Chlorobenzene	33	44		34	41	
Ethylbenzene	41			42		
(m+p) Xylenes	153			154		
Styrene	30			33		
Bromoform	50	13		49	13	
1,1,2,2-Tetrachloroethane	67	14		66	13	

Table 6.11: Continue

Compound	Helium	Hydrogen		32		
1,4-Dichlorobenzene	Qualifier 1	Qualifier 2	Qualifier 3	Qualifier 1	Qualifier 2	Qualifier 3
1,2-Dichlorobenzene	66	32		67	33	
1,2,4-Trichlorobenzene	98	24		98	26	
Hexachloro-1,3-butadiene	65	63		64	63	
1,2,4,5-Tetrachlorobenzene	79			78		

Sensitivity proved to be more critical in the perspective to substituting helium with nitrogen as carries gas. As expected, an increase in the baseline was observed. An example is shown in **Figure 6.6**, where chromatograms acquired with helium and hydrogen are presented side by side for the same two compounds at the same concentration (dibromochloromethane and 1,2-dichloroethane). It is clear that the intensity of the background noise increased in parallel with the decrease in sensitivity.

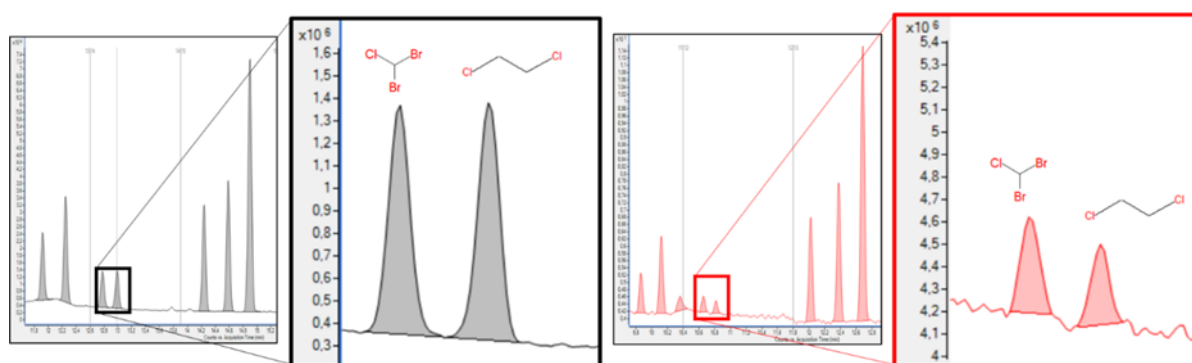


Figure 6.6. Detail of chromatograms (TIC) of the VOC mix with the Ion Extractor™ source with helium (black) and hydrogen (red) carrier gases.

Nevertheless, for some compounds, such as non-chlorinated aromatic compounds, the signal remained unchanged. However, for others, the transition to hydrogen as the carrier gas resulted in a significant loss of signal. This was particularly evident in the case of brominated compounds, which showed a marked decrease in response. This reduction is likely due to secondary reactions occurring in the ionization phase, particularly in the presence of an active gas such as hydrogen, which can influence the ionization process and lead to a diminished signal for these specific analytes.

For these reasons, a new ion source specifically designed for compatibility with hydrogen in GC-MS was tested to address the observed limitations. The HydroInert™ (Agilent Technologies) source, developed to ensure optimal performance with hydrogen as the carrier gas, demonstrated excellent results. An overview of the outcomes is summarized in **Table 6.12**. It lists data from analyses conducted using an Ion Extractor™ source with hydrogen as the carrier gas and the HydroInert™ source with hydrogen as the carrier gas. The responses have been normalized to those obtained with helium, which is considered the gold standard in GC-MS.

Table 6.12: Ratio between Extractor™ source with hydrogen and helium and HydroInert™ source with hydrogen and helium

Target compounds	Ion Extractor™ Source	HydroInert™ Source
Aromatics		
Benzene	92%	93%
Toluene	95%	89%
Chlorobenzene	101%	82%
Ethylbenzene	95%	105%
(m+p) Xylenes	97%	78%
Styrene	98%	101%
Average	96%	91%
Aromatic chlorinated compounds		
1,4-Dichlorobenzene	96%	83%
1,2-Dichlorobenzene	96%	83%
1,2,4-Trichlorobenzene	82%	65%
1,2,4,5-Tetrachlorobenzene	70%	66%
Average	86%	74%
Aliphatics brominated compounds		
Bromodichloromethane	50%	112%
Dibromochloromethane	39%	88%
1,2-Dibromoethane	56%	80%
Bromoform	38%	96%
Average	46%	94%
Aliphatics chlorinated compounds		
trans-1,2-Dichloroethylene	80%	73%
1,1-Dichloroethane	78%	109%
cis-1,2-Dichloroethylene	80%	79%
Chloroform	71%	111%
1,2-Dichloroethane	99%	160%
Trichloroethylene	82%	91%
1,2-Dichloropropane	78%	102%
1,1,2-Trichloroethane	69%	96%
Tetrachloroethylene	70%	68%
1,1,2,2-Tetrachloroethane	63%	105%
1,2,3-Trichloropropane	69%	118%
Hexachloro-1,3-butadiene	47%	73%
Average	74%	99%

Figure 6.7 visually presents the same results for selected target: dibromochloromethane (**A**, ion 129 m/z), 1,2-dibromoethane (**B**, ion 107 m/z), and 1,2,3-trichloropropane (**C**, ion 75 m/z), compounds in an intuitive manner. Specifically, the data in the overlaid chromatograms were acquired with helium as the carrier gas (black), hydrogen with the Extractor™ source (red), and hydrogen with the HydroInert™ source (green). These graphical representations highlight the differences in performance among the setups, making the impact of the HydroInert™ source readily apparent. As shown, the implementation of this technology effectively mitigates undesired secondary effects, such as those impacting brominated compounds, and achieves performance consistent with helium-based reference methods. These experimental findings strongly support the adoption of the HydroInert™ source in combination with hydrogen to ensure optimal performance, particularly when targeting low concentration levels, thereby meeting stringent analytical objectives. This overlay clearly demonstrates how the chromatographic response decreases when transitioning from helium to hydrogen as the carrier gas with a standard Extractor™ source. However, the HydroInert™ source effectively mitigates this effect.

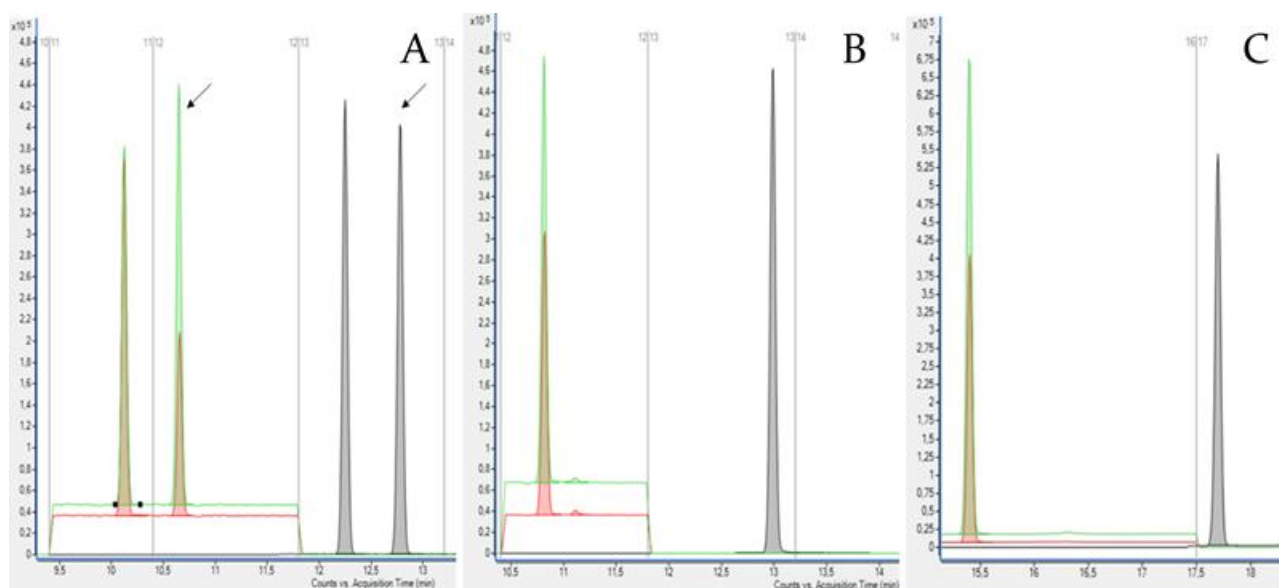


Figure 6.7: Overlay of SIM chromatograms for quantifier ions: dibromochloromethane (A, ion 129 m/z), 1,2-dibromoethane (B, ion 107 m/z), and 1,2,3-trichloropropane (C, ion 75 m/z), acquired with helium carrier gas (black), hydrogen with the Ion Extractor™ source (red), and hydrogen with the HydroInert™ source (green).

Additionally, **Figure 6.7A** highlights the analyte-dependent nature of the response decrease: for 1,1,2,2-tetrachloroethane ($t_{R,H_2}=10.19\text{min}$, and $t_{R,He}=12.25$), the response remains nearly identical with both hydrogen sources, whereas dibromochloromethane ($t_{R,H_2}=10.59\text{min}$ and $t_{R,He}=12.81$) shows a significant decrease with the standard Extractor™ source.

The last evaluation revisited spectral quality, this time focusing on the HydroInert™ source. As previously observed with the Extractor™ (Agilent Technologies) source in hydrogen, the HydroInert™ source also generally maintains consistent spectral quality. For certain analytes, as shown in **Figure 6.8**, the presence of additional fragments can be observed. Nevertheless, the identification, comparison, and matching with commercial libraries such as NIST yielded excellent results, with match scores exceeding 900 and reverse match scores above 950.

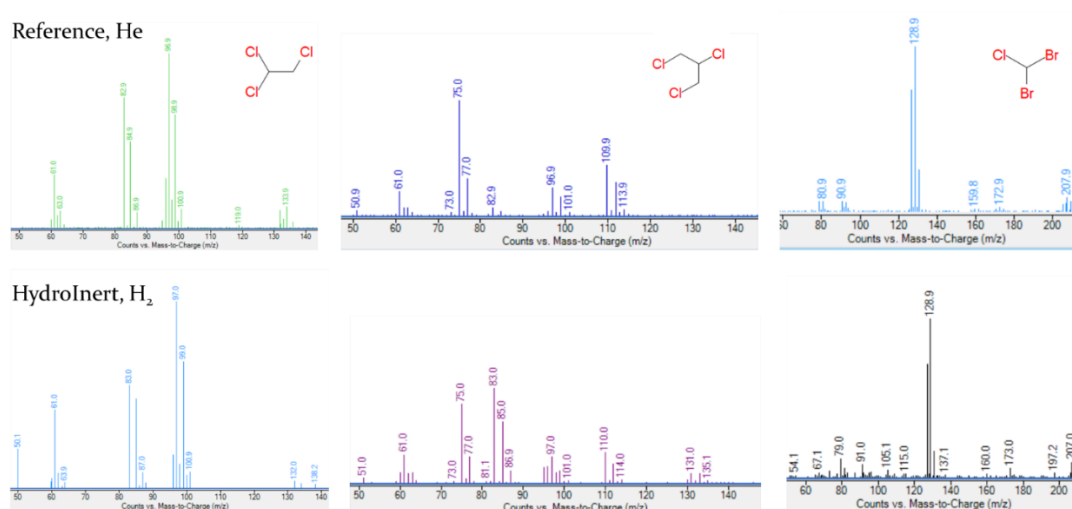


Figure 6.8: Examples of spectra acquired with hydrogen carrier and HydroInert™ source and comparison with commercial library. The molecules, from left to right, are 1,1,2-trichloroethane, 1,2,3-trichloropropane, and dibromochloromethane.

In conclusion, the evaluations conducted in this study were focused on applications involving target analytes, yielding promising results. However, the untargeted approach requires particular attention, as it is not possible to exclude a priori that certain chemical classes may experience unconventional fragmentation, leading to reduced spectral fidelity. This factor could compromise the reliability of the characterization process, especially in the absence of preliminary information on the chemical composition of the matrices under investigation. The results obtained convincingly support the benefits of the HydroInert™ source, ensuring good compatibility of the GC-MS technique with hydrogen as a carrier gas. While helium remains the preferred carrier gas in GC-MS, hydrogen proves to be a viable alternative in many application areas. Performance was observed to vary depending on the type of molecules analyzed: in some cases, the use of hydrogen affected the spectral quality or reduced sensitivity, while in others, the results remained consistent with those obtained using helium. The new HydroInert™ source, specifically designed for hydrogen use, has proven to be key in reducing unwanted secondary effects and enhancing system performance, confirming its potential in optimizing GC-MS analyses.

References

- [1] N. P. Parra and L. Taylor, "Why Instrument Detection Limit (IDL) is a Better Metric for Determining The Sensitivity of Triple Quadrupole LC/MS Systems: Technical Overview," *Agil. Technol.*, no. Idl, pp. 1-4, 2014.

Chapter 7

Validation of a High-Throughput Method for the Accurate Quantification of Secondary Products of Lipid Oxidation in High-Quality Hazelnuts (*Corylus avellana* L.): A Robust Tool for Quality Assessment

European hazelnuts (*Corylus avellana* L.) are a key ingredient for the confectionery industry with a distinctive flavor profile and high nutritional quality. However, the high-fat content makes raw kernels prone to autooxidation that takes place during storage. Fatty acids hydroperoxides (*i.e.*, primary products of oxidation) readily decompose by forming volatile secondary products with low odor thresholds and negative aroma notes. This study presents method performance parameters of an accurate quantitative methodology, based on multiple headspace solid phase microextraction (MHS-SPME) coupled to conventional and fast gas chromatography – mass spectrometry [(Fast)GC-MS], targeted to oleic and linoleic acids autooxidation products (*i.e.*, hexanal, heptanal, octanal, and nonanal). Results refer to good repeatability (max RSDr 12%) and intermediate precision (max RSDR 19.6%); linearity over four orders of magnitude ($R^2 \geq 0.995$ and calibration error $\leq 20\%$); good accuracy (error bias $\leq 20\%$); limit of detection and quantification compatible with real-world samples oxidative levels. Experimental results from method application on 358 industrial samples, confirm the method's effectiveness and its information capacity; moreover, for medium-to-high oxidized samples, the quantification of a larger set of products instead of a single marker (*i.e.*, hexanal), provides a more accurate picture on the actual quality/rancidity level.

This chapter was published as Original Research Paper in the Journal of Food Composition and Analysis:
S. Squara*, A. Caratti*, F.O. Gavilan, P. Bolzoni, N. Spigolon, G. Genova, G. Castello, M.G.B. González, L. Cuadros-Rodriguez, C. Bicchi, C. Cordero, Validation of a high-throughput method for the accurate quantification of secondary products of lipid oxidation in high-quality hazelnuts (*Corylus avellana* L.): A robust tool for quality assessment, J. Food Compos. Anal. 114 (2022) 104766. DOI: 10.1016/j.jfca.2022.104766.

*Simone Squara and Andrea Caratti equally contributed to this work.

7.1 Introduction

European hazelnut (*Corylus avellana* L.) is one of the most cultivated nut crops worldwide with a market portion in constant growth [1]. Agrofood and confectionery industry use hazelnut kernels as key-ingredient for dairy, bakery, confectionery, candy, and chocolate products due to their distinctive flavor profile.

According to the most recent FAO statistics [2], in 2019 the global production accounted for 1 million tons of in-shell hazelnuts with Turkey as the leading country with about the 67% of the global production, followed by Italy (12%), Azerbaijan (5%), and the USA (4%). Details on hazelnut world production and major cultivars of interest for confectionery have been recently reviewed by Squara *et al.* [1].

Hazelnut bulk composition accounts for a 50-73% of lipids [3], followed by carbohydrates (10-22%), proteins (10-24%), moisture ($\approx 5\%$), and ashes ($\approx 3\%$) [4]. The oxidative stability of the lipid fraction is a key element to preserve raw kernels quality over time. It is composed of a saponifiable portion consisting of a polar fraction (1.2%), and of an apolar one (98.8%) dominated by triacylglycerols (TAGs) [5]. Major fatty acids (FAs) are oleic acid (18:1 ω 9) representing 77.5-82.95% of the total, followed by linoleic acid (18:2 ω 6) with a 7.55-13.69%, and palmitic acid (16:0) reaching the 4.85-5.79% of the total.

Quality standards applied to select and valorize hazelnuts for direct consumption include the absence of damaged kernels [*e.g.*, insect-damaged, rotten, twin, and yellowed kernels are targeted by visual inspection [6]–[11]], a controlled moisture content, absence of rancidity, and some morphological parameters including shape and kernel size [12].

At industrial level, quality controls include additional parameters correlated to the desirable hedonic profile of finished products; cultivar, cultural techniques, geographical origin, harvesting time, post-harvest management, and shelf-life storage conditions are the major variables further considered for kernel selection [13]–[17]. Moreover, sensory analysis by expert panels supports hazelnut valorization by excluding kernels/batches with perceivable sensory defects and off-flavors [18], [19].

State-of-the-art studies provide proof of evidence that within the complex volatilome of raw hazelnuts, many external variables strongly affect its composition (*i.e.*, functional variables) and consequently hazelnuts industrial quality [1], [20], [21]. Cialiè Rosso *et al.* [20] analyzed the interrelation between cultivar/origin impact with post-harvest and storage conditions. Authors selected commercial samples of Tonda Gentile Romana and Ordu Turkish hazelnuts subjected to traditional sun-drying ($\approx 30/35^{\circ}\text{C}$) or artificial drying at low temperatures ($\approx 18/20^{\circ}\text{C}$) and then stored until twelve months at 5° and $18^{\circ}\text{C} \pm 0.1$ under *regular* (78% N_2 / 21% O_2) or *modified* (99% N_2 / 1% O_2) atmosphere. By explorative Principal Component Analysis (PCA) on 133 volatile organic compounds, cultivar/geographical origin and post-harvest drying conditions were identified as major drivers delineating diagnostic patterns within the whole detectable volatilome. A series of potent odorants with odor thresholds (OTs) up to 2,500 $\mu\text{g/L}$ were closely correlated ($r > 0.800$) to storage time. Of particular interest were fatty acids hydroperoxides degradation products: (*E*)-2-heptenal

(fatty, almond), hexanal (leaf-like, green), heptanal (fatty), octanal (fatty), and nonanal (tallowy, fruity) because of their unpleasant notes.

On the sensory quality degradation occurring during storage, many studies have confirmed the primary role of lipids [5], [9], [14], [20], [22]–[34]. Unsaturated FAs are subjected to autoxidation and, in presence of sensitizers and oxygen (O₂), they form hydroperoxide derivatives that can be readily degraded into secondary products with lower molecular weight and polarity (e.g., aldehydes, ketones, alcohols), and low odor threshold (OT) [35]. Moreover, if TAGs are hydrolyzed by endogenous esterases and lipases, the resulting free fatty acids (FFAs) accelerate quality degradation being up to 10 times more prone to autoxidation [36], [37].

Several studies have proposed volatile markers to monitor hazelnut shelf-life evolution and quality degradation [28]. Among volatile secondary products of lipid oxidation, hexanal has been proposed as a good indicator due to its high volatility and relative abundance which makes it very suitable for HS sampling. Pastorelli *et al.* [38] developed an HS-SPME-GC-FID method to profile hexanal and monitor its formation during storage while testing the reliability of e-nose sensors for faster screenings. The authors were successful in evaluating the role of oxygen scavengers in the autooxidation process posing a solid foundation for the adoption of fast and non-invasive methodologies [29].

Ghirardello *et al.* [31] made a step forward by developing a quantitative methodology based on stable isotope dilution for hexanal quantification. The study focused on two high-quality cultivars (i.e., *Tonda Gentile Trilobata* from Piedmont, Italy, and *Delisava* from Turkey) stored for twelve months at different temperature/atmosphere conditions (ambient temperature, 5°C, and –25°C with air, under vacuum or modified atmosphere). Of the monitored parameters, the hexanal content showed the highest consistency and robustness to assess the impact of storage on compositional quality.

Very recently, Pedrotti *et al.* [14] explored the possibility to predict the sensory quality of raw hazelnuts by rapid fingerprinting with Proton Transfer Reaction-Mass Spectrometry (PTR-MS). In their study spoiled hazelnuts showing perceivable sensory defects were screened for their volatile profile; then by machine learning, characteristic ions were selected and further correlated to their parent volatiles. Within marker compounds, several oxidation products were detected: octanal, 2,4-nonadienal (information on geometrical isomers was not available), and hexanal.

Considering the industrial quality control requirements for accurate and reliable methods, possibly including a confirmatory step, for the objective assessment of hazelnut rancidity to be used as a decision-making tool, this contribution presents performance evaluation of an accurate quantitative method targeted to major products of oleic and linoleic acids hydroperoxides cleavage, i.e., hexanal, heptanal, octanal, and nonanal. The method, in compliance with the principles of green analytical chemistry [39], includes a solvent-free extraction approach (i.e., SPME) implemented with multiple headspace extraction (MHE) followed by GC-MS to match for suitable specificity, selectivity, and quantification accuracy.

The validated procedure is then applied on a large set of hazelnut samples (n=358), differing for cultivar/geographical origin, harvest year, storage time and conditions. Results will be critically evaluated in light of method's information potential compared to existing/available screening techniques.

7.2 Materials and methods

7.2.1 Chemicals

Pure standard solution of *n*-alkanes (from *n*-C₇ to *n*-C₃₀) for Linear Retention Indices (*I*^T) calibration and system quality control were from Merck (Milan, Italy); they were used to prepare a reference *I*^T calibration mixture at 100 mg/L - "*n*-alkanes mixture". Pure reference standard solution of α -/ β -thujone from Merck was prepared in dibutyl phthalate (99% purity, Merck, Milan, Italy) and adopted as Internal Standard (IS) at the final concentration of 100 mg/L - "IS spiking mixture".

Standard reference compounds for calibration solutions of target aldehydes (hexanal, heptanal, octanal, nonanal from Merck) were prepared in dibutyl phthalate (99% purity, Merck) at a final concentration of 10.00 g/L ("reference solutions") and then diluted to suitable concentrations for linearity evaluation ("calibration mixtures") using dibutyl phthalate (99% purity, Merck, Milan, Italy) as solvent.

7.2.2 Reference solutions and calibration mixtures

Calibration curves were obtained by analyzing pure reference standards of target aldehydes (hexanal, heptanal, octanal, and nonanal) at different concentration levels within the range of 0.2 to 1,000 mg/L. Calibration mixtures were prepared by diluting suitable volumes of reference solutions in dibutyl phthalate (99% purity, Merck, Milan, Italy). **Supplementary Table ST1** lists available calibration levels and their actual amount referred to as targeted analytes

7.2.3 Hazelnut samples

Raw hazelnuts from different geographical/botanical origins were provided by Soremartec Italia Srl (Alba, Cuneo, Italy). Homogeneously ground raw materials (*i.e.*, *granella*) were immediately stored at -18°C away from UV exposure until analysis. For validation purposes, three samples were selected from those made available by Soremartec Italia Srl: Sample #A; Sample #B, and Sample #C.

Extensive method application on commercial quality-grade samples (n=358) was between harvesting years 2015 and 2020; samples were of different cultivars/geographical origins (*i.e.*, Giresun and Ordu and Ackakoca from Turkey, Tonda Gentile delle Langhe and Tonda Romana from Italy); post-harvest practices were classified by industry experts as "good" or "bad" according to the drying methodology/timing applied. As a further functional variable, storage time and conditions were selected to cover different rancidity levels (*i.e.*, from 0 to 24 months of storage at ambient temperature or refrigerated at 5°C with or without oxygen). Results on target analytes' accurate quantification are reported in the Supplementary Material; samples are listed with their unique #IDs.

7.2.4 Multiple headspace solid-phase microextraction: devices and conditions

Multiple HS-SPME was carried out using a Divinylbenzene/Carboxen/Polydimethylsiloxane (DVB/CAR/PDMS) d_f 50/30 μm fiber of variable length (Set-up A 1 cm long and Set-up B 2 cm long). Fibers were from Merck and, before use, they were conditioned as recommended by the manufacturer by thermal desorption at 270°C for 60 min in a split/splitless injector under constant carrier gas flow.

Before analyte extraction from hazelnut *granella*, the standard-in-fiber procedure was applied to pre-load IS on the SPME device [40]. This procedure was performed on 5 μL of IS mixture placed in a 20 mL glass sealed vial and submitted to HS-SPME at 50°C for 5 min.

MHS-SPME was conducted on differential amounts of hazelnut *granella* (i.e., 250, 150, 100, 50, and 25 mg $\pm$ 0.2 mg) at 50°C for 50 min for Set-up A and at 50°C for 10 min for Set-up B. Sampling conditions and analytical set-up are detailed in Table 7.1.

7.2.5 GC-MS: instrument set-up and conditions

The GC-MS system consisted of an MPS-2 multipurpose auto-sampler (Gerstel GmbH, Germany) integrated with an Agilent 7890A GC unit, coupled to an Agilent 5977B MS detector provided by a single-quadrupole analyzer and a High-Efficiency Source HES (Agilent, Little Falls, DE, USA) operating in EI mode at 70 eV. The GC transfer line was set at 250°C for Set-up A and 270°C for Set-up B while the MS scan range was 40-300 m/z with a scanning rate of 1,000 amu/s for Set-up A and 6,250 amu/s for Set-up B.

The capillary column was a 100% polyethylene glycol column (SolGel-Wax™ from Trajan Scientific and Medicals, Ringwood, Australia) (30 m $\times$ 0.25 mm dc, 0.25 μm df) in Set-up A and a 100% polydimethylsiloxane (DB1 from J&W, Little Falls, DE, USA) (10 m $\times$ 0.10 mm dc, 0.40 μm df) in Set-up B. Fiber thermal desorption into the GC injector port was under the following conditions: Set-up A split/splitless injector in split mode, split ratio 1:10, injector temperature 250°C and 5 min of thermal desorption; and Set-up B split/splitless injector in pulsed split mode, split ratio 1:10 and pulse pressure at 295 kPa for 2 min, injector temperature 250°C and 2 min of thermal desorption. The carrier gas was helium at a constant flow of 1.0 mL/min in Set-up A and 0.4 mL/min in Set-up B.

The temperature program was for Set-up A from 40°C (1 min) to 180°C at a rate of 3°C/min and to 260°C a rate of 15°C/min (5 min), and for Set-up B from 50°C (1 min) to 260°C a rate of 20°C/min and to 280°C a rate of 30°C/min (3 min). Analytical parameters are summarized in Table 7.1.

The *n*-alkanes liquid sample solution for IT determination (i.e., *n*-alkanes mixture) was analyzed under the following conditions: split/splitless injector in split mode, split ratio 1:50 for Set-up A and 1:100 for Set-up B, injector temperature 270°C, injection volume 1 μL .

Table 7.1: Analytical conditions applied to methods Set-up A and Set-up B

	Set-up A	Set-up B
SPME devices	Divinylbenzene/Carboxen/Polydimethylsiloxane (DVB/CAR/PDMS) df 50/30 μ m fiber 1 cm	Divinylbenzene/Carboxen/Polydimethylsiloxane (DVB/CAR/PDMS) df 50/30 μ m fiber 2 cm
SPME conditions	Standard-in-fiber pre-loading 5 μ L ISs MIXTURE 20 mL Vial 50°C - 5 min Hazelnut granella sampling 250 $\pm$ 0.5 mg granella 20 mL Vial 50°C - 50 min	Standard-in-fiber pre-loading 5 μ L ISs MIXTURE 20 mL Vial 50°C - 5 min Hazelnut granella sampling 25 $\pm$ 0.3 mg granella 20 mL Vial 50°C - 10 min
GC-MS	100% polyethylene glycol column (SolGel-Wax™ Trajan Scientific, Ringwood, Australia) (30 m $\times$ 0.25 mm dc, 0.25 μ m df) He carrier 1.0 mL/min Oven programming: 40°C (1 min) to 180°C at 3°C/min and to 260°C at 15°C/min (5 min) Injector Temperature: 250° Injection mode: split - split ratio 1:10 Fiber desorption time: 5 min	100% polydimethylsiloxane (DB1 from J&W, Little Falls, DE, USA) (10 m $\times$ 0.10 mm dc, 0.40 μ m df) He carrier 0.4 mL/min Oven programming: 50°C (1 min) to 260°C at 20°C/min and to 280°C at 30°C/min (3 min) Injector Temperature: 250° Injection mode: pulsed split - split ratio 1:10 Pulsed pressure 195 kPa - 2 min Fiber desorption time: 2 min
	El-MS at 70 eV 40-300 m/z at 1,000 amu/sec	El-MS at 70 eV 40-300 m/z at 6,200 amu/sec
Analysis time and samples	115 min for a single run 55 min in sequence overlap 26 samples/day	28 min for a single run 18 min in sequence overlap 80 samples/day

7.2.6 Multiple Headspace Extraction - principle of operation and reference equations

Multiple headspace extraction (MHE) consists of a dynamic, stepwise gas extraction technique carried out to accurately quantify volatiles in solid or heterogeneous samples. The approach adapted to HS-SPME is referred to as MHS-SPME [41]–[45]. It can be conducted as an external standard calibration and consists of three experimental steps:

- Exhaustive extraction of analytes from calibration solutions within a suitable range of concentrations/amounts.
- Exhaustive extraction of analytes from representative samples (solid, liquid, or heterogeneous) to define HS linearity boundaries.
- Application of the MHE procedure to samples of interest.

The first two steps, visualized in **Supplementary Figure 1 – SF1**, are necessary to build a cumulative instrumental response function after repeated consecutive extractions from the HS of the same aliquot of calibration solutions or representative samples: extractions must reach complete (*i.e.* exhaustive) analyte removal from the sample, and 4 to 6 consecutive extractions are recommended to validate the exhaustiveness of the extraction process for all analytes of interest [41], [46]. The decrease of the chromatographic peak area, after consecutive extractions, is exponential in HS linearity conditions [41], [45], [47]. This condition refers to the linear function between the analyte concentration in the sample (C_o) and its concentration in the gas phase (C_g), or between C_o and the instrumental response (A) obtained when analyzing an HS aliquot. The actual linear range depends on the analyte partition coefficient into the HS (K_{hs}) and its activity coefficient: it ranges generally between 0.1 and 1% in the sample and can be tuned by modifying sampling temperature, equilibration times, and the ratio between HS volume (V_h) and condensed phase volume (V_s).

The sum of the instrumental response, *i.e.*, chromatographic peak area (A_s) from each step of HS extraction corresponds to the total response (A_T) as it is generated by the analyte originally present in the sample.

Equation 7.1 is applied to obtain the cumulative instrumental response A_T :

$$\text{Equation 7.1 } A_T = \sum_{i=1}^{\infty} A_i = A_1 \frac{1}{(1 - e^{-q})} = \frac{A_1}{(1 - \beta)}$$

where A_T is the total estimated response, A_i is the analyte response from the first extraction, and q is a constant associated with the response exponential decay (β) through consecutive extractions. The sequential steps corresponding to the exhaustive extraction of an analyte from a sample in HS linearity conditions are illustrated in **Supplementary Figure 2 – SF2** for a MHS-SPME approach.

The term q can be obtained by plotting the natural logarithm of the chromatographic peak areas as a function of the number of extractions. From this, a linear regression equation (**Equation 7.2**) can be calculated as follows:

$$\text{Equation 7.2 } \ln A_i = a(i - 1) + b$$

where i is the number of extraction steps, b is the intercept on the y axis, and a is the slope.

β (e^{-q}) is analyte and matrix dependent, and it is constant for those samples exerting comparable matrix effects, *i.e.* the samples of the same matrix [48], [49]. Moreover, β values can be adopted to confirm or rebut the HS linearity assumption: its dependence on K_{hs} provides information on matrix retention behavior, and on the actual release of the target analyte under specified conditions [43].

When applied to calibration solutions, MHE provides data for external calibration. Curves can be adopted to estimate the analyte amount in the sample with a simplified procedure, where the analyte response after the first extraction step (A_i) is sufficient for accurate quantitation [41].

For validation purposes, experimental β values were estimated by consecutively extracting the tested samples up to 6-10 successive steps, and %RSD on exponential decay was evaluated. It is expected that β values from Set-up A and Set-up B will be different since analytes extraction is conducted in non-equilibrium under different sampling conditions.

7.2.7 Responses normalization and choice of internal standards

The use of single or multiple internal standards (IS) is a recommended practice in GC [50] and, regarding HS-SPME, it enables both the chromatographic response normalization for calibration purposes and the monitoring of sampling device extraction repeatability (intra-batch) and reproducibility (inter-batch). Moreover, it compensates for response fluctuations due to random instrumental errors.

In this study, the standard-in-fiber procedure introduced in 2005 by Wang *et al.* [40], was adopted. This procedure entails pre-loading the IS onto the SPME fiber coating before the sampling step. It informs about the reliability and efficiency of the accumulating polymer over

time, and better exploits the isotropy of absorption and desorption of an analyte onto and from the SPME device.

7.2.8 Validation parameters

The method performance parameters considered for the validation of the analytical procedure were selected within those describing its suitability for quality control purposes of marker constituents of raw hazelnuts (*Corylus avellana* L.). Reference guidelines adopted to design the experimental protocol were by Eurachem/Citac Institutions [51], [52] and EU Commission Decision 657/2002 [53].

7.2.9 Data acquisition

Data were acquired and processed by MassHunter version 10.0 Agilent Technologies (Wilmington, DE, USA). Statistical analysis and chemometrics were conducted by XLSTAT 2014 (Addinsoft, New York, NY, USA).

7.3 Results and Discussion

The next sections will introduce some preliminary considerations about the choice of targeted analytes and the method performance parameters evaluated according to consensus guidelines. Results, corresponding to a conventional GC-MS strategy (Set-up A) and an alternative fast-GC-MS approach (Set-up B), are briefly commented in light of method application in industrial control.

The validation scheme adopted was designed according to samples' analysis time (115 min Set-up A vs. 30 min Set-up B) and based on a "one-week" replicated analyses for Set-up A, or "three-days over one-week" replicates for Set-up B, as summarized in **Table 7.2**.

Out of the scope of this contribution is the description of the design of experiments that brought to the definition of the final experimental set-up and conditions. Next sections instead provide experimental proof of method's suitability while experimental data can be adopted to quickly implement method's procedure while also supporting its accreditation according to ISO 17025:2018 standards.

7.3.1 Target analytes selection and quantification strategy

The selection of target analytes was conducted keeping into consideration the autoxidation process of oleic and linoleic acids, the major FAs represented in hazelnut oil. According to the model studies conducted by Grosch, Schieberle, and co-workers [35], [54], [55], and considering the average relative amount of oleic acid ($\approx 80\%$) and linoleic acid ($\approx 11\%$) in hazelnut TAGs [56], the expected pattern of volatile secondary products should include as primary component hexanal (100%), followed by nonanal (53%), octanal (47%), and heptanal (8%). This estimation was based on the experimental data obtained by accurate quantification of volatile secondary products of oleic and linoleic hydroperoxides cleavage in a model system with 1 g of FA kept at 20°C and with a FA uptake of 0.5 mole oxygen/mole [35].

Table 7.2: Validation scheme applied to Set-up A and Set-up B.

	Set-up A	Set-up B		
	One-week	One-week		
Linearity	Day 1-5	Day 1	Day 2	Day 3
	Range 1-1,000 ng (absolute amount) Number of replicates per Calibration Level n=3	Range 1-5,000 ng (absolute amount) Number of replicates per Calibration Level n=3		
Precision				
	Sample #A (n=6)	Sample #A (n=3)	Sample #A (n=3)	Sample #A (n=3)
	Sample # B (n=6)	Sample #B (n=3)	Sample #B (n=3)	Sample #B (n=3)
		Sample #C (n=3)	Sample #C (n=3)	Sample #C (n=3)
Accuracy				
Exponential decay (β)	Sample #A up to 6 extraction steps (n=3)	Sample #A up to 10 extraction steps (n=3)		
Estimated amount	Sample #A (n=6)	Sample #A (n=3)	Sample #A (n=3)	Sample #A (n=3)
	Sample #B (n=6)	Sample #B (n=3)	Sample #B (n=3)	Sample #B (n=3)
		Sample #C (n=3)	Sample #C (n=3)	Sample #C (n=3)
Limit of Detection Limit of Quantitation	T2 samples normalized response > S/N mimicking blank sample	Sample #A, #B, #C and RIF T5 normalized response > S/N mimicking blank sample		

Regarding the quantification approach, of those available for HS analysis of samples of vegetal origin [44], MHE was here selected to overcome inaccuracy issues that might occur quantifying analytes released by heterogeneous solid samples. As recently estimated by Stilo *et al.* [57], the quantification error arising from internal standardization, a procedure largely adopted for responses normalization while erroneously considered an accurate quantification approach, could exceed one order of magnitude the actual amount present in the original sample. The influence of partition coefficient (K_{HS}) on the analyte distribution in the HS, and the detector response factor, both analyte's specific, cannot be assimilated to a single, or even to multiple ISs. For selected volatiles characteristic of the extra-virgin olive oil (*i.e.*, fifteen key analytes including potent odorants and geographical markers), the % relative error (RE%) between the MHS-SPME quantification with external standardization (taken as reference for comparison) and the HS-SPME quantification based on the IS, was on average 208% with a maximum of +538% for (*E*)-2-hexenal. This analyte, the most abundant in the extra-virgin olive oil volatile fraction, is highly informative being a key-aroma compound responsible for the *green* and *fruity* notes [58].

Moreover, with solid samples, as for the case of hazelnut *granella*, when sampling conditions (*e.g.*, sample amount vs. headspace volume) exceed the HS linearity and saturation occurs, the amount of the analyte partitioned in the HS does not reflect the actual amount in the sample thus generating a systematic error in the quantitative estimation.

Based on all these considerations and keeping in mind the needs of an industrial laboratory for a high-throughput method with a confirmatory role and suitable performances, MHE-SPME-GC-MS was identified as the most suitable approach. Conventional fast-GC-MS set-ups were both implemented at this validation stage to provide alternative solutions for food quality control laboratories.

Table 7.3: continue

Set-up B								
Sample #A			Sample #B			Sample #C		
Hexanal			Hexanal			Hexanal		
D1	D2	D3	D1	D2	D3	D1	D2	D3
6.27E-02	6.88E-02	7.21E-02	1.13E-01	1.32E-01	1.15E-01	1.35E-01	1.08E-01	1.25E-01
6.56E-02	6.81E-02	5.62E-02	1.05E-01	1.42E-01	1.26E-01	1.10E-01	1.44E-01	1.19E-01
7.28E-02	7.91E-02	6.37E-02	1.13E-01	1.33E-01	1.00E-01	1.19E-01	1.28E-01	1.19E-01
RSDr %	9.66		RSDr %	7.08		RSDr %	10.42	
RSDR %	9.66		RSDR %	12.98		RSDR %	10.42	
Heptanal			Heptanal			Heptanal		
D1	D2	D3	D1	D2	D3	D1	D2	D3
3.40E-02	3.40E-02	3.83E-02	4.08E-02	4.92E-02	4.20E-02	4.66E-02	4.55E-02	4.47E-02
3.56E-02	3.45E-02	3.54E-02	4.01E-02	4.64E-02	4.25E-02	4.29E-02	4.83E-02	4.05E-02
3.33E-02	3.92E-02	3.47E-02	4.09E-02	4.53E-02	4.11E-02	3.91E-02	4.36E-02	4.13E-02
RSDr %	5.93		RSDr %	2.91		RSDr %	6.57	
RSDR %	5.93		RSDR %	8.21		RSDR %	6.57	
Octanal			Octanal			Octanal		
D1	D2	D3	D1	D2	D3	D1	D2	D3
9.40E-03	9.60E-03	1.13E-02	1.61E-02	2.43E-02	1.73E-02	2.01E-02	2.08E-02	2.18E-02
1.03E-02	1.10E-02	8.10E-03	1.65E-02	2.28E-02	1.88E-02	2.00E-02	2.60E-02	2.12E-02
9.50E-03	1.23E-02	8.80E-03	1.64E-02	2.16E-02	1.64E-02	2.14E-02	2.36E-02	2.05E-02
RSDr %	12.73		RSDr %	5.58		RSDr %	7.43	
RSDR %	12.73		RSDR %	19.15		RSDR %	7.43	
Nonanal			Nonanal			Nonanal		
D1	D2	D3	D1	D2	D3	D1	D2	D3
1.77E-02	1.98E-02	1.88E-02	2.51E-02	3.25E-02	2.29E-02	3.82E-02	3.22E-02	3.25E-02
1.82E-02	1.99E-02	1.39E-02	2.43E-02	3.22E-02	2.49E-02	3.03E-02	3.79E-02	3.07E-02
1.84E-02	2.26E-02	1.66E-02	2.27E-02	3.41E-02	2.35E-02	3.22E-02	3.46E-02	3.19E-02
RSDr %	9.23		RSDr %	4.06		RSDr %	8.82	
RSDR %	9.23		RSDR %	19.6		RSDR %	8.82	

Response function and linearity within the working range

Calibration curves obtained by analyzing pure reference standards of target aldehydes (hexanal, heptanal, octanal, and nonanal) by MHS-SPME are needed to accurately determine the analyte concentration in real samples by interpolation. The MHE recommended practices [41] indicate, as a suitable procedure for method calibration, the exhaustive extraction of known amounts of reference standard compounds within the established method conditions. In the current validation, the response function was examined within a wide range of analytes amounts to account for the extensive oxidation of hazelnut kernels occurring during their entire shelf-life.

Figure 7.1 shows the chromatographic profile of a calibration mixture analyzed by Set-up A (1A) and Set-up B (1B). Overlaid chromatograms are obtained from multiple headspace extraction ($n=4$) of the sample calibration vial.

Based on available data from previous studies, calibration levels were set between 1.0 and 5,000 ng. According to SANTE Guidance document [59], a multi-level calibration (more than three concentrations) was preferred. Then the most appropriate calibration function was defined (*e.g.* linear or polynomial second order) and its suitability for accurate quantitation of real samples was examined through determination coefficient R^2 (acceptability criterion $R^2 \geq 0.995$ for linear regression and $R^2 \geq 0.99$ for polynomial functions) and deviation of the back-calculated concentrations of the calibration standards, *i.e.*, residues, not exceeding $\pm 20\%$ of error.

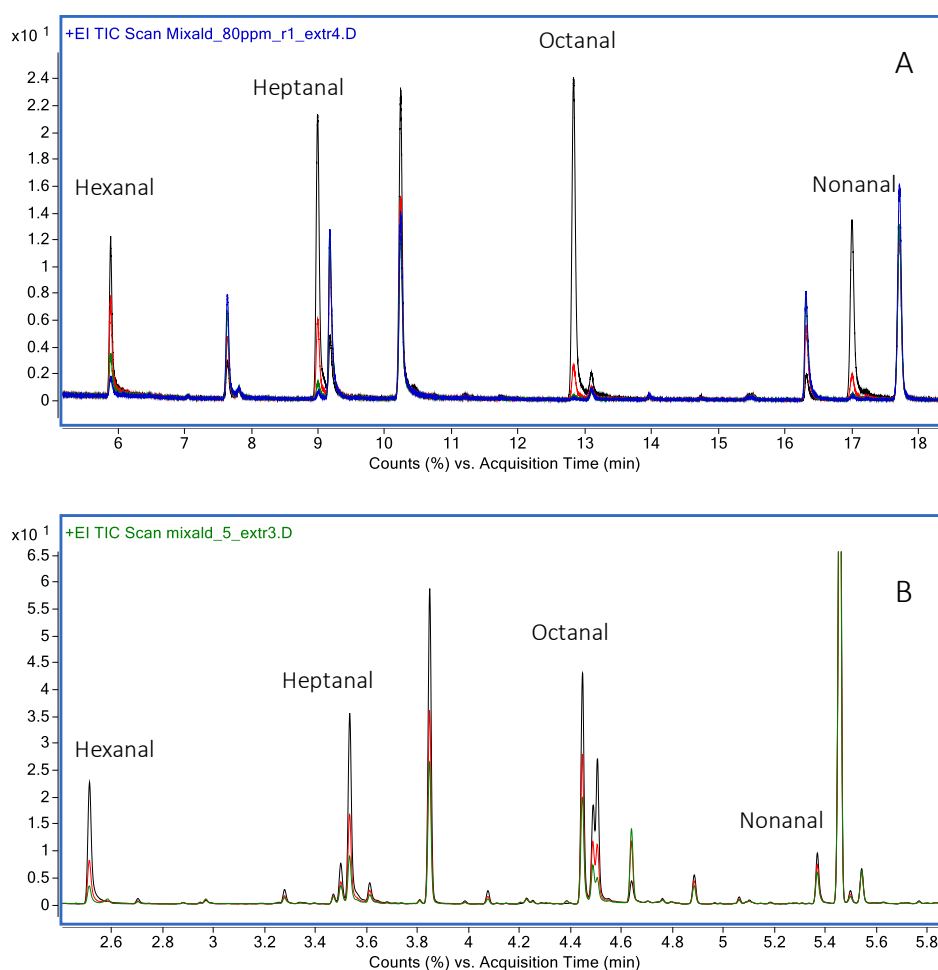


Figure 7.1: Chromatographic profiles of calibration mixtures analyzed with Set-up A (1A) and Set-up B (1B). Overlaid chromatograms with different colors show the chromatographic response decay along with successive headspace extractions (i.e., multiple headspace extraction) from the same calibration vial.

Supplementary Table ST₃ reports response functions and performance parameters (R^2) evaluated for Set-up A and Set-up B. To note, for Set-up B a wider range was considered due to the expected amounts of analytes in severely oxidized samples. Linearity was verified, analyte-by-analyte, in limited concentration ranges, and functions are reported in **Supplementary Table ST₃**. To note, in Set-up B for nonanal it was not possible to define a linearity range with a determination coefficient below the defined acceptability criterion (i.e., $R^2 = 0.995$).

Precision: repeatability (r) and intermediate precision (R)

"The closeness of agreement between independent test results obtained under stipulated (predetermined) conditions" as EU Commission Decision 657/2002 [53] defines method precision, was here expressed as coefficient of variation (CV) or RSD of the mean. In case of repeated analysis of a sample carried out under within-laboratory reproducibility conditions, the intra-laboratory CV or RSD of the mean, shall not exceed 20% in the range between 5/10 $\mu\text{g/kg}$ and 10% in the range between 10/20 mg/kg [53]. Based on the validation scheme designed for the two analytical Set-up (**Table 7.2**), in the case of a "three-days over one-week" validation, the intermediate precision (%RSDR) is also described.

Results are reported in **Table 7.3** and refer to normalized responses (*i.e.*, normalized chromatographic areas) calculated from replicated analyses of Sample #A, #B, and #C run within one-week or three-days over one-week validation. To note, for Set-up B, one-way ANOVA was applied on replicated measures within each validation day (D₁, D₂, D₃) and used to estimate inter-day repeatability (%RSDR) while rejecting/accepting the null hypothesis, *i.e.* replicated measures belong to the same statistical population. Thereby, a characteristic %RSDR within multiple days of method application is defined.

Results indicate a good method precision, always below 20%, for both analytical set-ups; inter-day repeatability (intermediate precision) estimated on Sample #B is generally higher than for other test samples suggesting that inhomogeneity of *granella* accompanied by shorter sampling times (10 min vs. 50 min) could have an impact on the method precision and accuracy.

Trueness/Accuracy

The accuracy of a method is determined by the combination of precision and trueness, with the latter referring to the closeness of agreement between the arithmetic mean of a large number of test results and the value from a Certified Reference Material (CRM). If there is no CRM available, the trueness has to be replaced by recovery. Recovery should be determined by experiments using a fortified blank matrix reasonably covering the working range of the method.

Since the current methodology is targeted to hazelnut constituents (secondary products of lipid oxidation, *i.e.*, fatty acids hydroperoxides degradation products) and a blank sample was not available, accuracy/recovery was evaluated by combining two strategies. At first, exhaustive extraction up to method LOD by MHE was conducted in both Step-up A and B on test samples hazelnut *granella* (Sample #A, #B and #C) and experimental decay functions were examined for their effective extraction and consistency. Estimated amounts (using the best fitting model concerning the normalized response of the sample), including method precision, were then tabulated and %RSD calculated to see if they were compliant with reference values as reported in SANTE Guidance document [59]. In a second step, a standard addition (SA) approach performed with Set-up B was applied on two spiking levels (*i.e.*, +25 ng and +250 ng) for test Sample #A to establish the actual amount of target aldehydes as resulting from an alternative methodology. To note, SA amounts were estimated by the MHS-SPME procedure due to the solid nature of the sample matrix. A quantification bias/error by over-estimation was expected since the spiking procedure does not reproduce the actual distribution of the analytes within the solid sample matrix. The acceptable range of (in)accuracy was established at $\pm 20\%$ of the predetermined amount, as indicated by reference guidance documents [53].

As an additional test, repeated multiple headspace extractions (up to 10 steps) were conducted on the test sample Sample #A ($n=3$) to estimate decay functions, evaluate their fitness with an exponential model and extrapolate β values. **Supplementary Table ST4** reports mean β values for targeted aldehydes in Sample #A together with %RSD on replicated experiments.

Table 7.4 reports estimated amounts for target analytes in test samples as a result of applying MHS-SPME with Set-up A and Set-up B, while **Figure 7.2A** shows the chromatographic profile of the test sample (Sample #B); hexanal, heptanal, and nonanal are highlighted in red rectangles. Enlarged areas of **Fig. 7.2B-D** include the first four extractions of the same sample vial.

Table 7.4: Accuracy data calculated for targeted aldehydes analyzed over one-week (Set-up A) or in a three-days (D1-D2-D3) over one-week period (Set-up B) according to **Table 7.2** experimental plan. One-way ANOVA was conducted on Set-up B measurements to verify the null hypothesis that the three-days measurements belong to the same statistical population and to evaluate recovery % versus Set-up A amounts (Rec.vs.A).

	Set-up A		Set-up B					
	Sample #A	Sample #B	Sample #A		Sample #B		Sample#C	
	Hexanal		Hexanal					
	Day 1-5		Day 1/2/3		Day 1/2/3		Day 1/2/3	
Mean	8.07E+02	1.55E+03	Mean	7.96E+02	Mean	1.72E+03	Mean	1.90E+03
RSD %	8.50	14.47	RSDr %	13.06	RSDr %	8.66	RSDr %	13.50
			RSDR %	13.06	RSDR %	16.25	RSDR %	13.50
			Rec. vs. A	98.80%	Rec. vs. A	110.60%		
			Rec. vs. SA	112.53%				
	Heptanal		Heptanal					
Mean	2.01E+02	3.50E+02	Mean	1.95E+02	Mean	3.57E+02	Mean	4.16E+02
RSD %	9.45	9.16	RSDr %	10.84	RSDr %	4.16	RSDr %	17.07
			RSDR %	10.84	RSDR %	19.81	RSDR %	17.07
			Rec. vs. A	97.20%	Rec. vs. A	102.20%		
			Rec. vs. SA	117.83%				
	Octanal		Octanal					
			Mean	2.21E+02	Mean	4.24E+02	Mean	5.02E+02
Mean	2.17E+02	3.90E+02	RSDr %	13.74	RSDr %	5.07	RSDr %	8.45
RSD %	6.47	6.72	RSDR %	13.74	RSDR %	18.2	RSDR %	8.45
			Rec. vs. A	101.90%	Rec. vs. A	108.80%		
			Rec. vs. SA	118.00%				
	Nonanal		Nonanal					
Mean	2.57E+02	4.27E+02	Mean	2.34E+02	Mean	4.41E+02	Mean	8.39E+02
RSD %	8.90	7.11	RSDr %	4.31	RSDr %	8.09	RSDr %	13.99
			RSDR %	14.56	RSDR %	13.29	RSDR %	13.99
			Rec. vs. A	91.10%	Rec. vs. A	103.30%		
			Rec. vs. SA	83.73%				

Cross-validation between quantitative data obtained by the two Set-up with MHS-SPME and SA was conducted. Linear saturated aldehydes amount estimated by applying MHS-SPME approach with Set-up A were arbitrarily taken as a reference value and results from Set-up B were compared (intra/inter-day data) to define a characteristic bias over the reference. As a further confirmatory test, SA with MHS-SPME conducted with Set-up B was also compared to the reference. Results are reported, together with precision data, in Electronic Supplementary Material – **Supplementary Figures SF3-10**.

Results are all below the critical threshold of $\pm 20\%$ RSD for intra and inter-day data. Moreover, the recovery % versus the confirmatory methodology (i.e., SA) never exceeds the $\pm 20\%$ indicating acceptable relative accuracy for both Set-up A and B.

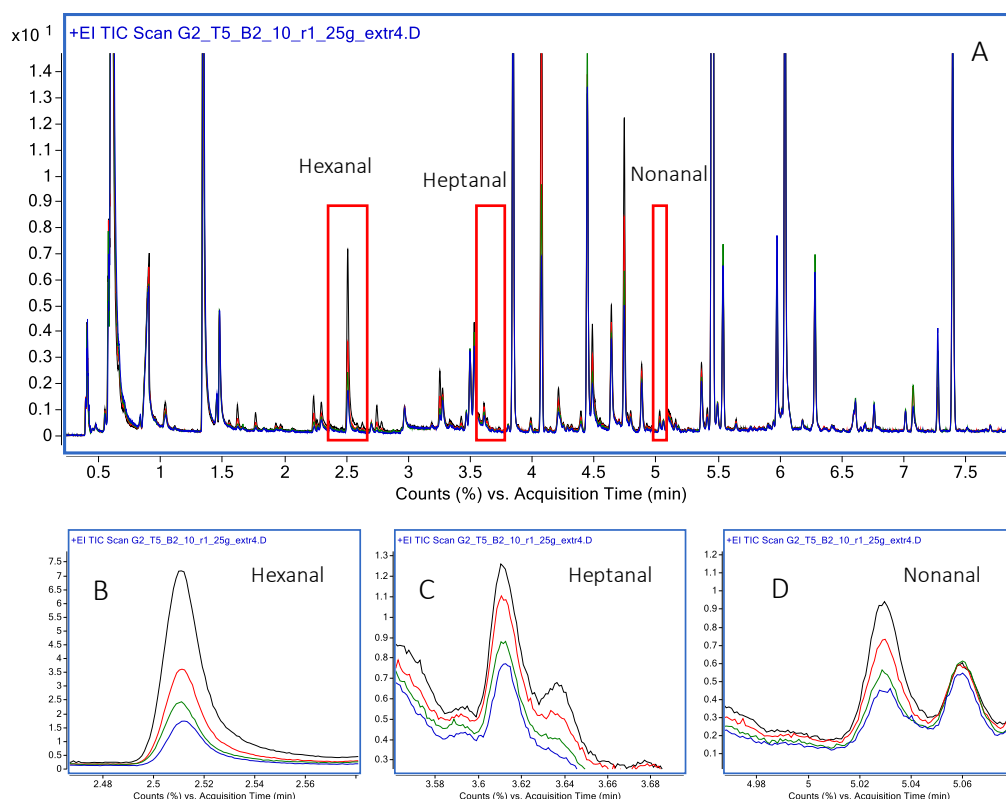


Figure 7.2: Chromatographic profile of the test sample acquired by Set-up B conditions; highlighted in red rectangles corresponding to the elution regions of hexanal, heptanal, and nonanal. Enlarged areas of **2B-D** include the first four extractions of the same sample vial.

Limit of detection and limit of quantitation

The limit of detection (LOD) of an analytical procedure is the smallest amount or concentration of analyte in the test sample that can be reliably distinguished from zero signal [60] while the limit of quantitation (LOQ) is the lowest amount of analyte in a sample that can be quantitatively determined with suitable precision and accuracy [61]. The estimation of LOD and LOQ involves the independent preparation of a suitable number of blanks or pseudo-blanks (samples spiked with a reasonably low amount of analyte) [62]. As discussed above, for this specific application, a blank matrix was not available and any fatty matrix mimicking the retention behavior of raw hazelnut *granella* would not be suitable for this experimental measure too. Therefore, according to guidance documents [62], LOD and LOQ were calculated by Equations 7.4 and 7.5 from normalized areas estimated from *T_i* signals exceeding the S/N of a factor of 3 or 10. For this purpose replicated measures on Sample #A, #B, and #C samples were computed together (n=12).

Equation 7.4:
$$X_{LOD} = 3.9 \times \frac{s_{y,b}}{b}$$

Equation 7.5:
$$X_{LOQ} = 3.3 \times X_{LOD}$$

Where X_{LOD} is the limit of detection and X_{LOQ} is the limit of quantitation; $s_{y,b}$ is the blank signal standard deviation calculated on replicated tests measurements (Set-up A n=12 and Set-

up B n=12) (above $3 \times S/N$) and b is the slope of the calibration curve at concentration levels closer to the LOD (i.e., levels 1-4).

Supplementary Table ST5 reports LOD and LOQ values (ng absolute amount) characteristic of the method and estimated for Set-up A and Set-up B. To note, the differential performance is strictly dependent on chromatographic efficiency. For this reason, when separation efficiency is higher and peak widths smaller (i.e., Set-up B performances), the LODs and LOQs are lower. To note: limits are not provided as relative to the matrix since the method, *per se*, is suitable to cover a very wide dynamic range for example by sampling a higher amount of matrix to compensate for the lower abundance of the targeted aldehyde.

Uncertainty associated with the results

Uncertainty was derived by the accuracy estimation that includes precision and calibration bias. For Set-up A, the uncertainty corresponds to the intermediate precision %RSDR on estimated amounts obtained by replicated measurements over one-week of analyses, while for Set-up B it refers to %RSDR characteristic of the method. In both cases, the largest estimated RSD was considered to define the relative uncertainty interval. **Table 7.5** lists relative uncertainty for targeted analytes quantified by Set-up A or Set-up B.

Table 5: Analyte relative uncertainty for the two method Set-up.

		Relative uncertainty %		
Set-up A		intra-week	Set-up B	intra-day intra-week
Hexanal	-	± 14.47	Hexanal	± 13.06 ± 16.25
Heptanal	-	± 9.45	Heptanal	± 10.84 ± 19.81
Octanal	-	± 6.72	Octanal	± 13.74 ± 18.20
Nonanal	-	± 8.90	Nonanal	± 8.09 ± 14.56

7.3.3 Real-world samples application results and remarks

The quantification method, realized with Set-up B conditions over a five-years time interval, was then applied to a large set of real-world samples provided by Soremertec Italia Srl. Samples (n=358) were from different geographical origins, harvesting years, shelf-life storage time and conditions. Their oxidation status was therefore very different, spanning a wide range of concentrations between method's LOQ up to 23,148 ng/g relative to the sum of target aldehydes expressed as hexanal equivalents. Results are reported in **Supplementary Table 6 – ST6**.

By applying linear regression analysis, the correlation between the sum of target quantified aldehydes (y), expressed as ng/g of hexanal equivalents, and hexanal amount (x) was verified. **Figure 7.3A** shows the regression function together with confidence interval (95%) boundaries for the estimated model. A random validation (n=30) was also performed (white-colored circles). The model function ($y = 61.28 + 2.38 x$) was characterized by a R^2 of 0.7840 indicating, as expected, a poor linear correlation between the variables. As it appears by observations/samples dispersion over the Cartesian plane, for samples at higher rancidity status (i.e., higher amount of target aldehydes) the correlation between variables is lower resulting in larger values for standardized residuals as shown in **Figure 7.3B**.

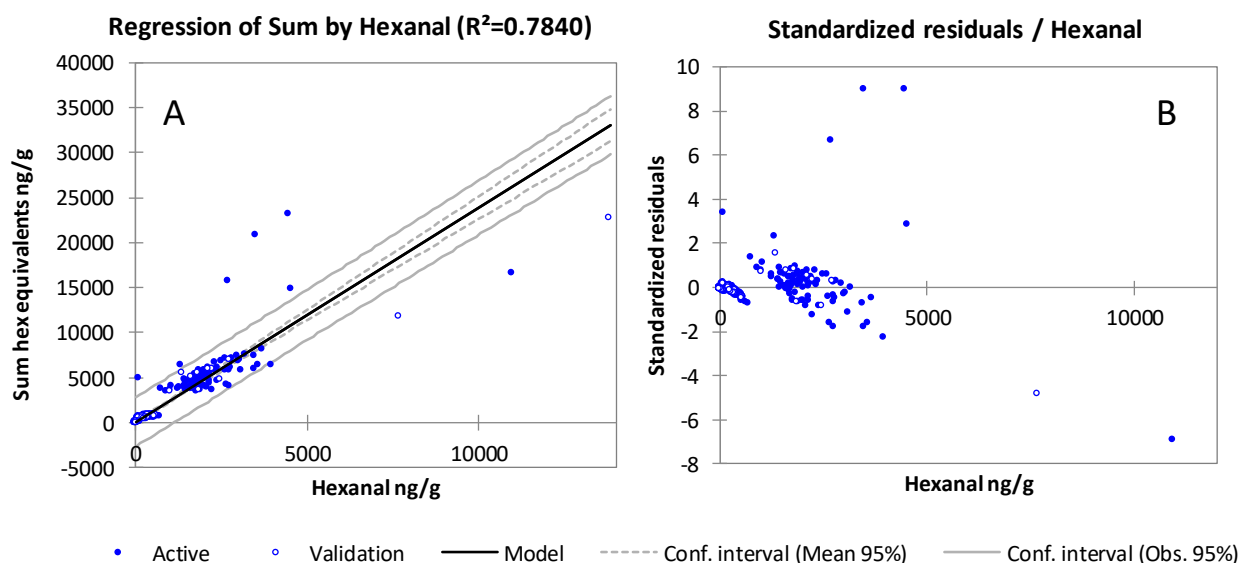


Figure 7.3: Linear regression analysis (A) on the sum of target aldehydes (ng/g) expressed as hexanal equivalents vs. hexanal amounts (ng/g) and standardized residuals (B) over hexanal amounts (ng/g). Blue indicators correspond to observations (samples) computed in the model while white-colored indicators represent validation samples ($n=30$).

A closer look at the quantitative data by Pearson correlation (r) test, confirms that hexanal and octanal are poorly correlated ($r = 0.5880$ with a P -value < 0.0001) as for hexanal and heptanal ($r = 0.6691$ with a P -value < 0.0001). This evidence indicates that to assess rancidity status by secondary products of fatty acids hydroperoxides degradation, the computation of all major reaction products provides more accurate information than a single target analyte. By the sum of C6-C9 aldehydes, a theoretical coverage of about 75% of secondary products is provided [35], making this parameter a good decision-maker for quality assessment and storage strategies.

7.4 Conclusions

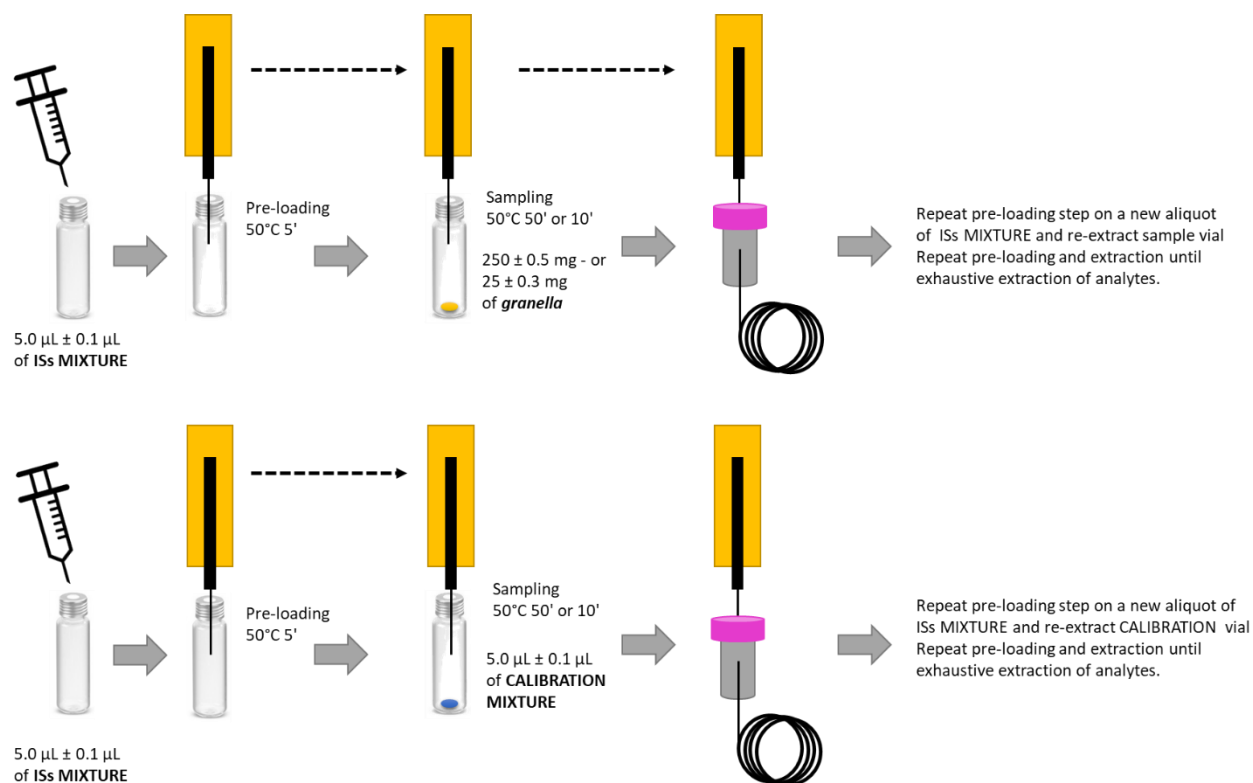
A fully automated, high-throughput procedure based on MHS-SPME combined to conventional and fast gas chromatography – mass spectrometry [(fast)GC-MS) was here validated, based on reference guidelines for food quality and safety assessment (Eurachem/Citac and DG-SANCO protocols), and its fitness-for-purpose demonstrated on a large set of hazelnut samples of industrial interest ($n=358$). The evaluation of performance parameters indicated a better precision and accuracy for the conventional set-up (*i.e.*, Set-up A), where sampling time is longer (50 min *vs.* 10 min). However, fast-GC (*i.e.*, Set-up B) analytical conditions enabled the achievement of suitable precision and intra-day accuracy that are both below (or very close-to) acceptability criteria. The advantages provided by a faster methodology (*i.e.*, Set-up B conditions) with its higher throughput of samples per unit time, might compensate for larger inaccuracy by increasing the number of replicated samples per batch.

The study, for the first time, proposes a truly accurate procedure, with confirmatory attitudes, for the quantitative determination of major secondary product of oleic and linoleic

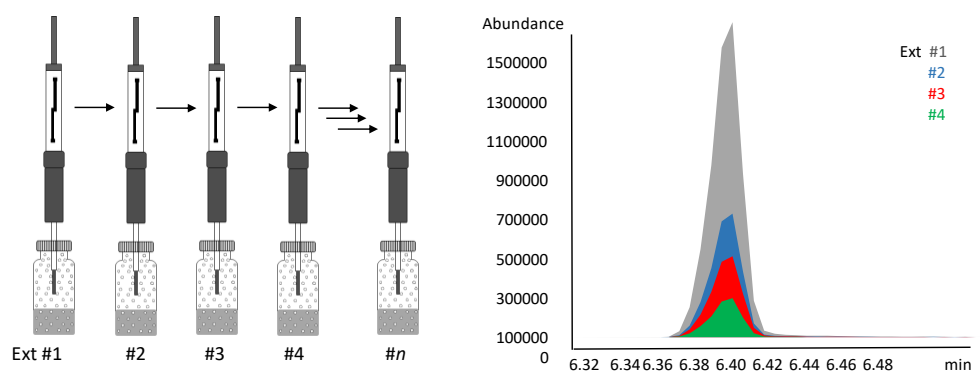
acid hydroperoxides degradation to be used as benchmarking tool for hazelnut rancidity assessment and/or as a decision-maker for market screening and storage strategies.

Compared to existing methodologies, characterized by a poor-accuracy (e.g., Internal Standard normalization or response evaluation) or by a limited information capacity (e.g., single analyte quantification vs. four major target), this study makes a step ahead while also offering a solid basis for method accreditation according to ISO 17025:2017 standards [63]. By this methodology, quality assessment is supported by an objective measure of rancidity and industrial strategies can be driven toward a better/strategic management of hazelnut quality.

Electronic Supplementary Material



Supplementary Figure 1. Sequential steps corresponding to the exhaustive extraction of an analyte from a sample in HS linearity conditions for a HS-SPME approach. Adapted from [44].



Supplementary Figure 2. Sequential steps corresponding to the exhaustive extraction of an analyte from a sample in HS linearity conditions for a HS-SPME approach. Adapted from [44].

[illegible]

Amount ng/g					
Ref - Set-up Ax 1.55E+03	1 1.62E+03	2 2.07E+03	3 1.55E+03		
	1.70E+03	2.07E+03	1.73E+03		
	1.53E+03	1.86E+03	1.32E+03		
	Day 1	Day 2	Day 3		
n	3	3	3		
S2 - Variance	7391.70	15139.53	43711.85	Dev.Std	
Mean	1616.13	1998.48	1534.91	247.5487	
Precision					
Relative Standard Deviation Repeatability	RSDr		8.66 %		
Relative Standard Deviation Intermediate precision	RSDR		16.25 %		
Analysis of Variance Inter-group (A)					
Significance test					
Source of variation		Sum of variances Q	DF	Estimated q	
Intermediate precision (inter-group)		367682.19	2	183841.09	qR
Repeatability (intra-group)		132486.18	6	22081.03	qr
Total		500168.37	11	62521.05	
F3 =	8.326	A meaningful	s2 =	(qA - qr) / n	
F(0.05;2;6)	5.14325285		inter-day variance	40440	

Supplementary Figure SF5: Analysis of Variance and Accuracy estimation for heptanal in Sample #A. Reference amount arbitrarily taken from Set-up A measurement (2.01E+02 ng/g). Accuracy 97.2% - verified.

Supplementary Figure SF6: Analysis of Variance and Accuracy estimation for heptanal in Sample #B. Reference amount arbitrarily taken from Set-up A measurement (3.50E+02 ng/g). Accuracy 102.2% - verified.

[illegible]

Supplementary Figure SF7: Analysis of Variance and Accuracy estimation for octanal in Sample #A. Reference amount arbitrarily taken from Set-up A measurement ($2.17\text{E}+02$ ng/g). Accuracy 101.9% - verified.

[illegible]

Supplementary Figure SF8: Analysis of Variance and Accuracy estimation for octanal in Sample #B. Reference amount arbitrarily taken from Set-up A measurement (3.90E+02 ng/g). Accuracy 108.8% - verified

Supplementary Figure SF9: Analysis of Variance and Accuracy estimation for nonanal in Sample #A. Reference amount arbitrarily taken from Set-up A measurement ($2.57\text{E}+02$ ng/g). Accuracy 91.1% - verified

Supplementary Figure SF10: Analysis of Variance and Accuracy estimation for nonanal in Sample #B. Reference amount arbitrarily taken from Set-up A measurement ($4.27\text{E}+02$ ng/g). Accuracy 103.3% - verified

Supplementary Table ST1: Calibration levels, calibration mixture concentration (referred to each targeted aldehyde) and absolute amounts analyzed by MHS-SPME for linearity evaluation in Set-up A and Set-up B.

Calibration Level	CALIBRATION MIXTURE mg/L	Absolute amount in 5µL
1	0.2	1.0 ng
2	0.5	2.5 ng
3	1.0	5.0 ng
4	2.5	10 ng
5	5.0	25 ng
6	10	50 ng
7	25	100 ng
8	50	250 ng
9	100	500 ng
10	250	1,000 ng
11	500	2,500 ng
12	1,000	5,000 ng

Supplementary Table ST2: Identity confirmation results: retention times and diagnostic *m/z* ratios stability.

Set-up A		IS - α -Thujone	Hexanal	Heptanal	Octanal	Nonanal
	<i>t_R</i> min	17.55	5.53	8.74	12.64	16.84
	RSD% (n=12)	0.02	0.05	0.03	0.02	0.02
	<i>t'_R</i> min	-	0.32	0.50	0.72	0.96
	RSD% (n=12)	-	0.04	0.02	0.01	0.02
	Ti/Q ₁ (m/z)	81/67 - 72%	56/72 - 28%	70/96 - 14%	84/56 - 133%	57/96 - 25%
	RSD% (n=12)	0.6	4.3	5.4	3.7	7.4
	Ti/Q ₂ (m/z)	81/95 - 54%	56/82 - 22%	70/55 - 63%	84/69 - 96%	57/82 - 38%
	RSD% (n=12)	0.5	1.8	3.7	10.9	4.7
	Ti/Q ₃ (m/z)	81/110 - 80%	56/57 - 70%	70/71 - 28%	84/81 - 51%	57/98 - 38%
	RSD% (n=12)	0.4	3.3	5.5	3.5	4.8
Set-up B		IS - α -Thujone	Hexanal	Heptanal	Octanal	Nonanal
	<i>t_R</i> min	5.45	2.51	3.53	4.48	5.36
	RSD% (n=27)	0.04	0.06	0.04	0.03	0.03
	<i>t'_R</i> min	-	0.46	0.65	0.82	0.98
	RSD% (n=27)	-	0.05	0.03	0.03	0.02
	Ti/Q ₁ (m/z)	81/67 - 78%	56/72 - 23%	70/96 - 12%	84/56 - 132%	57/96 - 16%
	RSD% (n=27)	3.2	3.7	5.0	13.0	3.1
	Ti/Q ₂ (m/z)	81/95 - 50%	56/82 - 17%	70/55 - 83%	84/69 - 68%	57/82 - 26%
	RSD% (n=27)	1.4	3.7	7.1	12.8	3.6
	Ti/Q ₃ (m/z)	81/110 - 67%	56/57 - 69%	70/71 - 32%	84/81 - 115%	57/98 - 25%
	RSD% (n=27)	2.1	3.0	6.2	11.2	3.3

Supplementary Table ST₃: Response functions estimated within the entire interval. Performance parameters (R^2) are all within the acceptable threshold (* Levels corresponds to different absolute amounts according to **Table ST₁**).

Set-up A	Levels	R^2	Function
Hexanal	Lv 3-10	0.996	$y = -3E-06x^2 + 0.0193x + 0.08$
Heptanal	Lv 3-10	0.998	$y = -4E-07x^2 + 0.0187x + 0.4191$
Octanal	Lv 3-10	0.997	$y = -2E-07x^2 + 0.0126x - 0.188$
Nonanal	Lv 3-10	0.996	$y = -2E-06x^2 + 0.0169x - 0.2633$
Hexanal	Lv 3-10	0.996	$y = 0.0153x + 0.5151$
Heptanal	Lv 3-10	0.998	$y = 0.0182x + 0.4757$
Octanal	Lv 3-10	0.997	$y = 0.0123x - 0.1555$
Nonanal	Lv 3-10	0.995	$y = 0.0148x - 0.0335$
Set-up B	Levels	R^2	Function
Hexanal	Lv 1-12	0.996	$y = -6E-07x^2 + 0.0085x + 0.5283$
Heptanal	Lv 1-9	0.998	$y = -1E-05x^2 + 0.0147x + 0.4222$
Octanal	Lv 1-9	1.000	$y = -6E-06x^2 + 0.0082x + 0.0024$
Nonanal	Lv 1-9	0.999	$y = 7E-06x^2 + 0.0085x + 0.1691$
Hexanal	Lv 1-7	0.997	$Y = 0.01581x + 0.19024$
Heptanal	Lv 1-7	0.998	$Y = 0.01480x + 0.40561$
Octanal	Lv 1-8	0.999	$Y = 0.00693x + 0.02232$
Nonanal	Lv 1-9	0.993	$Y = 0.01174x + 0.07321$

Supplementary Table ST₄: β values estimated from multiple extractions on test Sample #A analyzed by Set-up A and Set-up B. The consistency of the exponential decay is by RSD%.

Set-up A		
Sample #A - up to 6 steps Number of replicates = 3	β	RSD%
Hexanal	0.87	1.60
Heptanal	0.94	0.71
Octanal	0.91	1.23
Nonanal	0.95	0.95
Set-up B		
Sample #A - up to 10 steps Number of replicates = 3	β	RSD%
Hexanal	0.87	1.01
Heptanal	0.92	0.54
Octanal	0.79	6.99
Nonanal	0.91	0.41

Supplementary Table ST5: LOD and LOQ as estimated by Equations 4 and 5 for the two method Set-up.

	Set-up A			
	Hexanal	Heptanal	Octanal	Nonanal
$S_{y,b}$	1.20E-02	6.61E-02	1.27E-02	2.61E-02
b	5.00E-02	5.16E-02	3.19E-02	4.56E-02
LOD ng	7.95E-01	4.23E+00	1.31E+00	1.89E+00
LOQ ng	2.38E+00	1.27E+01	3.94E+00	5.66E+00
	Set-up B			
	Hexanal	Heptanal	Octanal	Nonanal
$S_{y,b}$	1.85E-03	9.76E-04	7.39E-04	3.48E-03
b	1.47E-02	1.14E-02	8.60E-03	1.21E-02
LOD ng	4.15E-01	2.83E-01	2.84E-01	9.48E-01
LOQ ng	1.25E+00	8.48E-01	8.51E-01	2.84E+00

Supplementary Table ST6: Estimated amounts of target aldehydes (ng/g) in real-world samples (n=358). Results are expressed as hexanal equivalents.

Sample #ID	Hexanal ng/g	Heptanal* ng/g	Octanal* ng/g	Nonanal* ng/g	Sum hexanal equivalents ng/g
1	4468.10	2230.25	9317.86	7131.94	23148.15
2	13825.28	1223.98	4361.49	3387.83	22798.58
3	3497.84	2024.60	11979.51	3356.70	20858.65
4	10967.00	715.09	2642.01	2332.55	16656.65
5	2697.92	2373.99	8514.23	2133.80	15719.94
6	4559.22	1335.52	5478.57	3523.56	14896.87
7	7688.31	471.48	1883.99	1708.60	11752.38
8	3701.73	577.03	1292.75	2645.77	8217.27
9	3189.97	1048.28	1222.40	2176.62	7637.28
10	2967.15	1042.51	1183.51	2204.87	7398.05
11	3475.40	675.61	1188.05	2043.17	7382.23
12	2819.23	502.42	2182.05	1689.69	7193.39
13	2620.00	1413.69	948.52	2158.59	7140.81
14	2750.00	1254.08	851.23	2201.20	7056.52
15	3048.10	523.25	1253.08	2176.17	7000.59
16	2770.00	1047.73	853.79	2329.01	7000.53
17	2520.00	1185.68	906.28	2258.00	6869.96
18	3013.27	594.72	1165.19	2037.78	6810.97
19	2310.00	1092.19	867.88	2400.02	6670.08
20	1333.27	1028.92	2996.82	1114.95	6473.97
21	3607.93	292.93	1245.76	1299.60	6446.21
22	3971.01	252.65	1104.62	1108.94	6437.22
23	2120.00	1197.08	824.35	2030.78	6172.22
24	2390.00	1208.48	764.19	1789.36	6152.03
25	2766.32	826.51	838.73	1711.47	6143.03
26	2799.67	762.70	909.67	1624.19	6096.23
27	2240.00	1151.48	768.03	1888.77	6048.28
28	2270.00	1130.96	788.51	1789.36	5978.83
29	3489.76	234.45	1064.40	1148.00	5936.61
30	2130.00	1140.08	800.03	1846.17	5916.27
31	2180.00	841.38	794.91	2087.59	5903.88
32	2370.00	1073.95	794.91	1661.55	5900.41
33	3103.83	461.54	1290.69	1042.26	5898.31
34	1860.00	1162.88	803.87	2030.78	5857.53
35	2040.00	1101.31	806.43	1874.57	5822.31
36	2269.53	485.69	1593.20	1467.64	5816.05
37	2630.72	770.10	934.66	1449.36	5784.84
38	2766.32	459.91	838.73	1711.47	5776.42
39	1910.00	1265.48	733.47	1704.15	5613.11
40	1830.00	1056.85	759.07	1902.97	5548.89
41	2090.00	998.71	739.87	1718.35	5546.93
42	1371.26	450.32	1868.06	1818.90	5508.54
43	2020.00	965.64	775.71	1704.15	5465.51

Supplementary Table ST6: continue

Sample #ID	Hexanal ng/g	Heptanal* ng/g	Octanal* ng/g	Nonanal* ng/g	Sum hexanal equivalents ng/g
44	1920.00	1174.28	747.55	1618.95	5460.77
45	2160.00	1093.33	766.75	1434.33	5454.41
46	2010.00	1084.21	668.19	1633.15	5395.55
47	1744.45	487.18	1783.89	1338.36	5353.87
48	2160.00	916.62	675.87	1547.94	5300.43
49	1946.38	573.20	841.50	1827.19	5188.27
50	1820.00	1048.87	680.99	1576.34	5126.20
51	1980.00	1002.13	661.79	1476.93	5120.85
52	1640.00	1091.05	736.03	1590.54	5057.63
53	1800.00	897.24	707.87	1633.15	5038.26
54	90.60	49.01	801.45	4035.50	4976.56
55	1780.00	962.22	705.31	1505.34	4952.87
56	2038.43	553.43	802.83	1507.59	4902.27
57	2483.51	285.56	933.59	1158.70	4861.36
58	1730.00	848.22	660.51	1604.74	4843.47
59	1920.00	947.40	670.75	1283.80	4821.95
60	1810.00	994.15	625.95	1347.70	4777.79
61	1460.00	1080.79	673.31	1533.74	4747.84
62	1690.00	795.77	634.91	1618.95	4739.62
63	2418.53	523.23	1357.98	391.75	4691.49
64	1540.00	884.70	709.15	1547.94	4681.79
65	2163.13	270.40	1156.92	1067.52	4657.98
66	1580.00	997.57	664.35	1408.77	4650.68
67	1790.00	893.82	637.47	1319.30	4640.58
68	1494.08	412.22	1702.05	948.88	4557.22
69	2166.77	564.35	666.82	1029.17	4427.11
70	1761.18	577.49	710.06	1365.40	4414.13
71	1760.00	843.66	573.46	1129.00	4306.12
72	1676.63	573.92	704.73	1349.57	4304.85
73	2666.07	153.76	669.28	683.61	4172.72
74	1856.44	535.51	633.17	1135.52	4160.64
75	2760.12	153.20	795.21	436.65	4145.19
76	1730.43	221.76	1042.82	1136.79	4131.80
77	2036.52	566.89	647.23	878.33	4128.98
78	1060.00	1036.33	540.18	1491.13	4127.65
79	2011.14	493.49	636.80	930.78	4072.22
80	1450.00	836.82	579.86	1191.49	4058.17
81	1430.00	845.94	561.94	1146.04	3983.92
82	1586.64	436.69	637.92	1310.91	3972.16
83	1280.00	784.37	572.18	1332.08	3968.64
84	1492.39	420.26	1298.94	741.62	3953.21
85	1574.47	426.46	652.51	1254.57	3908.00
86	2099.74	460.32	597.41	745.41	3902.88
87	1687.52	480.71	633.78	1090.17	3892.17
88	1634.98	494.58	651.89	1069.39	3850.84
89	758.06	487.92	1256.29	1308.02	3810.29
90	1270.00	753.59	504.34	1218.47	3746.40
91	1723.96	426.40	585.40	997.21	3732.97
92	2248.81	132.11	608.14	721.20	3710.26
93	1902.66	505.78	571.25	681.21	3660.90
94	1877.27	535.77	564.25	661.89	3639.18
95	1037.76	396.27	1926.20	245.35	3605.58
96	1781.43	442.99	563.83	809.94	3598.19
97	1476.87	430.88	591.10	1074.87	3573.73
98	1795.49	432.37	601.54	710.99	3540.39
99	1030.00	728.51	430.10	1339.18	3527.79
100	930.42	375.01	1988.58	216.75	3510.76
101	452.07	30.98	165.54	280.22	928.81
102	589.76	94.73	62.25	181.54	928.28
103	491.23	63.64	262.79	108.00	925.67
104	590.50	113.33	72.34	149.18	925.34
105	535.89	174.44	81.63	132.36	924.33
106	503.22	52.50	113.30	239.61	908.63

Supplementary Table ST6: continue

Sample #ID	Hexanal ng/g	Heptanal* ng/g	Octanal* ng/g	Nonanal* ng/g	Sum hexanal equivalents ng/g
107	582.81	98.19	64.55	161.21	906.76
108	322.75	1.14	93.16	486.06	903.10
109	416.78	132.42	165.89	185.57	900.66
110	452.09	68.70	240.10	131.32	892.20
111	418.10	32.76	233.80	202.76	887.42
112	528.56	114.06	82.18	159.95	884.76
113	318.37	1.14	36.44	522.70	878.65
114	375.98	1.14	135.88	354.86	867.86
115	512.61	156.19	91.44	99.39	859.63
116	513.48	101.36	73.15	153.77	841.76
117	387.40	1.14	98.99	346.00	833.53
118	390.72	46.31	136.84	256.42	830.29
119	530.41	97.99	70.87	130.86	830.13
120	466.29	166.37	53.00	140.91	826.57
121	449.77	166.55	84.14	123.49	823.95
122	583.73	71.31	29.11	138.71	822.86
123	297.69	1.14	25.19	497.47	821.50
124	321.32	54.79	87.57	357.17	820.85
125	448.42	1.14	51.97	318.98	820.51
126	519.77	14.05	75.91	209.62	819.35
127	408.52	27.10	59.96	318.28	813.86
128	511.02	105.13	70.09	125.44	811.69
129	344.53	37.89	152.50	275.91	810.83
130	538.69	96.40	62.42	112.28	809.79
131	437.40	1.14	89.51	280.92	808.97
132	408.14	168.11	116.23	115.67	808.15
133	531.73	98.00	56.71	118.91	805.35
134	255.81	109.76	305.55	131.44	802.56
135	418.94	160.15	93.99	125.45	798.53
136	434.42	104.87	124.51	133.50	797.30
137	469.61	150.13	71.31	102.86	793.91
138	469.08	130.00	84.42	104.64	788.14
139	357.66	38.24	152.77	237.04	785.71
140	403.99	65.39	92.60	223.64	785.62
141	389.31	60.95	84.67	243.45	778.38
142	509.01	1.14	56.46	211.44	778.05
143	435.79	26.04	194.42	106.17	762.41
144	262.94	88.91	302.00	108.00	761.85
145	381.89	26.07	197.51	154.03	759.50
146	657.01	1.14	2.56	82.11	742.82
147	359.83	65.38	213.88	97.76	736.85
148	551.65	1.14	8.58	172.45	733.82
149	703.20	1.14	2.56	22.04	728.94
150	483.88	16.82	162.31	61.57	724.59
151	535.66	44.59	21.28	122.07	723.59
152	416.55	90.49	71.22	144.96	723.22
153	476.60	80.20	57.70	108.38	722.88
154	329.80	1.14	22.67	365.78	719.40
155	273.92	23.10	126.00	291.28	714.30
156	241.06	1.14	2.56	468.70	713.46
157	343.17	114.37	103.59	150.80	711.93
158	310.02	71.79	62.72	267.08	711.62
159	378.71	1.14	27.59	302.33	709.77
160	400.43	94.68	65.50	148.25	708.85
161	311.02	60.72	66.58	250.37	688.69
162	236.44	1.14	25.98	421.87	685.44
163	244.98	1.14	26.60	408.49	681.20
164	208.14	58.39	196.49	218.00	681.02
165	114.42	131.35	367.16	67.16	680.08
166	469.11	1.14	66.45	142.64	679.34
167	501.08	60.24	27.26	89.53	678.10
168	317.60	28.00	110.78	221.27	677.64
169	428.63	1.14	50.66	194.63	675.06

Supplementary Table ST6: continue

Sample #ID	Hexanal ng/g	Heptanal* ng/g	Octanal* ng/g	Nonanal* ng/g	Sum hexanal equivalents ng/g
170	263.64	1.14	17.67	391.12	673.57
171	366.18	1.14	46.20	258.01	671.53
172	347.50	65.54	79.30	176.89	669.22
173	205.65	1.14	234.43	223.57	664.79
174	370.43	1.14	72.97	219.52	664.07
175	347.51	1.14	77.00	236.24	661.89
176	259.39	1.14	16.76	372.46	649.75
177	352.32	25.38	56.66	210.54	644.89
178	231.68	12.92	164.73	232.41	641.74
179	386.59	1.14	27.46	221.49	636.68
180	115.57	50.51	223.16	242.73	631.98
181	312.60	1.14	25.32	285.07	624.13
182	406.28	86.33	34.92	94.43	621.95
183	178.56	1.14	342.84	97.68	620.22
184	424.90	1.14	34.74	158.98	619.76
185	438.34	88.73	38.03	53.77	618.88
186	356.49	111.55	29.82	120.39	618.24
187	536.76	1.14	29.02	51.28	618.19
188	432.99	1.14	45.87	130.60	610.61
189	378.32	1.14	2.56	221.08	603.10
190	234.87	1.14	13.82	353.26	603.08
191	374.15	1.14	176.91	49.95	602.15
192	242.65	1.14	21.56	333.36	598.71
193	327.65	34.31	59.06	177.59	598.61
194	279.55	1.14	24.77	292.89	598.35
195	242.10	1.14	2.56	352.54	598.35
196	264.46	1.14	43.90	288.10	597.60
197	243.47	1.14	9.45	343.12	597.18
198	268.24	150.73	30.42	147.64	597.03
199	227.87	57.58	55.77	254.97	596.19
200	324.31	1.14	47.55	219.89	592.89
201	270.86	1.14	27.43	290.69	590.13
202	185.92	105.02	50.50	247.50	588.95
203	365.44	91.13	31.93	99.89	588.38
204	557.08	1.14	2.56	23.47	584.24
205	104.57	8.00	134.99	335.54	583.11
206	314.61	35.51	25.77	206.11	582.00
207	155.16	75.32	52.10	293.53	576.10
208	328.78	1.14	24.88	218.68	573.48
209	342.40	87.74	28.36	111.91	570.41
210	246.58	1.14	16.47	303.03	567.22
211	223.06	18.63	175.36	148.71	565.77
212	372.29	42.74	63.73	84.85	563.60
213	238.00	1.14	2.56	321.08	562.78
214	320.76	1.14	2.56	237.26	561.73
215	232.41	36.15	195.39	96.10	560.05
216	217.18	1.14	217.13	123.67	559.12
217	281.68	1.14	49.89	226.24	558.96
218	316.31	62.41	33.26	146.22	558.19
219	246.51	74.74	56.05	180.63	557.93
220	182.22	36.53	130.51	207.83	557.09
221	237.50	1.14	2.56	314.80	555.99
222	336.32	109.49	22.00	83.94	551.76
223	298.26	1.14	19.72	231.38	550.50
224	286.08	15.83	46.80	201.72	550.42
225	301.59	142.68	35.44	67.46	547.17
226	316.86	1.14	30.77	197.04	545.81
227	338.28	19.15	30.56	155.00	542.99
228	316.16	68.24	66.27	89.32	539.98
229	296.06	123.07	23.97	89.08	532.18
230	185.35	31.20	110.40	204.10	531.05
231	227.84	9.20	66.86	226.48	530.38
232	279.77	1.14	13.96	234.77	529.64

Supplementary Table ST6: continue

Sample #ID	Hexanal ng/g	Heptanal* ng/g	Octanal* ng/g	Nonanal* ng/g	Sum hexanal equivalents ng/g
233	394.75	1.14	23.16	110.59	529.64
234	77.73	8.57	177.31	261.95	525.57
235	245.65	1.14	2.56	275.55	524.90
236	286.62	80.16	23.61	134.12	524.52
237	270.17	1.14	40.06	211.24	522.61
238	260.34	81.44	61.65	117.84	521.27
239	191.62	1.14	30.98	296.87	520.62
240	175.12	1.14	2.56	341.49	520.32
241	215.71	1.14	19.40	283.66	519.91
242	114.08	54.59	54.55	293.93	517.16
243	277.77	90.63	13.33	135.04	516.77
244	269.22	1.14	2.56	242.38	515.30
245	367.54	1.14	140.63	4.26	513.57
246	299.77	1.14	26.05	186.41	513.36
247	242.65	45.46	20.66	201.22	509.99
248	270.45	136.54	38.10	61.04	506.12
249	76.32	1.14	129.62	298.68	505.76
250	210.34	1.14	21.58	269.65	502.71
251	258.55	126.97	24.21	89.69	499.42
252	329.41	64.32	14.76	89.75	498.24
253	307.21	1.14	15.14	174.36	497.85
254	292.24	97.45	38.54	68.25	496.49
255	256.92	1.14	2.56	231.89	492.50
256	210.36	1.14	2.56	276.52	490.58
257	321.22	1.14	38.32	120.48	481.17
258	205.69	68.63	58.51	146.55	479.38
259	90.40	1.00	12.82	55.32	159.54
260	150.67	1.14	2.56	4.26	158.63
261	82.71	1.14	31.01	43.44	158.30
262	111.52	1.14	2.56	42.43	157.65
263	122.38	1.14	11.32	22.33	157.17
264	120.53	1.14	2.56	32.81	157.04
265	86.14	2.98	29.24	38.48	156.83
266	120.47	1.14	2.56	32.02	156.19
267	141.84	1.14	2.56	10.03	155.58
268	29.01	1.14	82.53	42.89	155.57
269	115.99	1.14	2.56	33.96	153.65
270	75.91	1.14	2.56	71.03	150.63
271	66.09	1.14	36.54	46.44	150.21
272	141.25	1.14	2.56	4.26	149.22
273	51.65	1.00	29.85	66.34	148.84
274	50.84	1.00	45.72	49.53	147.09
275	96.54	1.14	2.56	43.34	143.58
276	48.54	4.34	43.15	44.98	141.01
277	49.55	1.00	20.05	68.87	139.47
278	46.86	1.00	5.57	85.23	138.66
279	51.21	1.00	8.99	76.70	137.90
280	50.43	1.00	22.74	63.68	137.85
281	57.26	1.14	52.63	25.61	136.64
282	64.63	1.14	66.09	4.26	136.11
283	73.16	1.14	2.56	58.91	135.77
284	64.54	1.00	18.25	49.69	133.48
285	48.25	1.00	29.41	53.86	132.52
286	71.21	1.14	19.45	39.75	131.55
287	47.95	5.89	39.61	37.17	130.62
288	21.62	1.00	1.00	105.37	128.99
289	53.95	1.00	9.23	62.75	126.94
290	104.92	1.14	2.56	17.22	125.84
291	71.22	1.00	24.49	28.53	125.24
292	103.99	1.14	2.56	16.61	124.31
293	17.87	1.14	14.66	89.56	123.22
294	41.35	1.14	2.56	74.79	119.84
295	110.62	1.14	2.56	4.26	118.58

Supplementary Table ST6: continue

Sample #ID	Hexanal ng/g	Heptanal* ng/g	Octanal* ng/g	Nonanal* ng/g	Sum hexanal equivalents ng/g
296	25.27	1.00	12.27	78.23	116.78
297	25.69	1.00	24.44	61.38	112.50
298	88.60	1.14	2.56	20.19	112.49
299	31.92	1.00	11.92	66.33	111.17
300	47.95	1.00	17.67	44.53	111.15
301	77.77	1.14	2.56	26.84	108.30
302	28.87	1.00	11.64	66.72	108.23
303	58.46	1.14	2.56	45.80	107.96
304	54.00	1.00	22.72	29.76	107.48
305	26.07	1.00	25.85	54.57	107.48
306	99.36	1.14	2.56	4.26	107.32
307	8.14	1.14	9.93	86.51	105.72
308	45.91	1.14	2.56	55.82	105.43
309	74.96	1.14	2.56	26.51	105.17
310	54.80	1.00	11.51	36.67	103.98
311	22.98	1.00	1.62	77.32	102.92
312	67.64	1.14	2.56	31.47	102.80
313	50.76	1.00	6.61	42.78	101.14
314	41.05	1.00	16.13	42.34	100.51
315	75.80	1.14	2.56	18.45	97.96
316	33.77	1.00	1.00	60.49	96.26
317	28.87	1.00	26.47	38.62	94.96
318	85.75	1.14	2.56	4.26	93.71
319	44.54	1.00	12.55	35.54	93.63
320	27.61	1.00	1.00	62.14	91.76
321	41.76	1.00	1.00	47.77	91.53
322	65.00	1.14	2.56	22.17	90.88
323	43.30	1.14	2.56	43.34	90.33
324	68.61	1.14	2.56	15.50	87.81
325	34.73	1.00	1.18	50.60	87.51
326	42.65	1.14	2.56	37.94	84.29
327	48.54	1.00	6.85	27.62	84.02
328	71.57	1.14	2.56	8.62	83.89
329	26.30	1.00	11.57	43.63	82.50
330	71.26	1.14	2.56	4.26	79.22
331	23.03	1.00	17.41	36.84	78.29
332	21.43	12.40	2.56	41.83	78.22
333	31.19	1.00	11.61	34.11	77.92
334	68.28	1.14	2.56	4.26	76.24
335	20.77	1.00	5.96	48.34	76.07
336	28.82	1.00	11.18	34.03	75.02
337	43.39	1.14	2.56	27.86	74.95
338	22.26	1.00	4.07	44.51	71.84
339	49.08	1.14	2.56	17.57	70.35
340	22.50	1.00	19.72	26.76	69.97
341	57.81	1.14	2.56	4.26	65.77
342	5.00	1.14	2.56	56.83	65.53
343	23.12	1.00	5.85	35.49	65.46
344	55.36	1.14	2.56	4.26	63.32
345	35.38	1.14	2.56	23.71	62.79
346	39.40	1.14	2.56	16.50	59.60
347	24.26	1.14	2.56	30.72	58.68
348	21.18	1.14	2.56	33.58	58.46
349	26.08	1.00	2.05	26.76	55.89
350	21.34	1.14	2.56	29.03	54.07
351	16.99	1.14	2.56	33.11	53.80
352	38.92	1.14	2.56	4.26	46.88
353	28.00	1.00	1.00	16.69	46.69
354	22.94	1.00	3.29	15.96	43.18
355	32.97	1.14	2.56	4.26	40.93
356	11.24	1.14	2.56	19.94	34.89

Supplementary Table ST6: continue

Sample #ID	Hexanal ng/g	Heptanal* ng/g	Octanal* ng/g	Nonanal* ng/g	Sum hexanal equivalents ng/g
357	5.00	1.14	2.56	25.60	34.30
358	9.57	1.14	2.56	19.02	32.29

References

- [1] S. Squara *et al.*, “Corylus avellana L. Aroma Blueprint: Potent Odorants Signatures in the Volatilome of High Quality Hazelnuts,” *Front. Plant Sci.*, vol. 13, no. March, pp. 1–25, Mar. 2022.
- [2] FAO, “FAOSTAT - Production quantities of hazelnuts (2013-2017),” 2019. .
- [3] A. I. Köksal, N. Artik, A. Şimşek, and N. Güneş, “Nutrient composition of hazelnut (*Corylus avellana* L.) varieties cultivated in Turkey,” *Food Chem.*, vol. 99, no. 3, pp. 509–515, 2006.
- [4] V. Cristofori, S. Ferramondo, G. Bertazza, and C. Bignami, “Nut quality and sensory evaluation of hazelnut cultivars,” *Acta Hortic.*, vol. 845, pp. 657–664, 2009.
- [5] C. Alasalvar *et al.*, “Turkish Tombul hazelnut (*Corylus avellana* L.). 2. Lipid characteristics and oxidative stability,” *J. Agric. Food Chem.*, vol. 51, no. 13, pp. 3797–3805, 2003.
- [6] S. A. Mehlenbacher, D. C. Smith, and L. K. Brenner, “Variance Components and Heritability of Nut and Kernel Defects in Hazelnut,” *Plant Breed.*, vol. 110, no. 2, pp. 144–152, 1993.
- [7] N. Göncüoğlu Taş and V. Gökmen, “Maillard reaction and caramelization during hazelnut roasting: A multiresponse kinetic study,” *Food Chem.*, vol. 221, pp. 1911–1922, 2017.
- [8] B. Yuan, M. Lu, K. M. Eskridge, L. D. Isom, and M. A. Hanna, “Extraction, identification, and quantification of antioxidant phenolics from hazelnut (*Corylus avellana* L.) shells,” *Food Chem.*, vol. 244, no. August 2017, pp. 7–15, 2018.
- [9] S. Belviso *et al.*, “Chemical, mechanical and sensory monitoring of hot air- and infrared-roasted hazelnuts (*Corylus avellana* L.) during nine months of storage,” *Food Chem.*, vol. 217, pp. 398–408, 2017.
- [10] A. Caligiani *et al.*, “Application of ¹H NMR for the characterisation and authentication of “Tonda Gentile Trilobata” hazelnuts from Piedmont (Italy),” *Food Chem.*, vol. 148, pp. 77–85, 2014.
- [11] A. Giraudo, R. Calvini, G. Orlandi, A. Ulrici, F. Geobaldo, and F. Savorani, “Development of an automated method for the identification of defective hazelnuts based on RGB image analysis and colourgrams,” *Food Control*, vol. 94, no. July, pp. 233–240, 2018.
- [12] United Nations, “United Nations Economic Commission for Europe Marketing and commercial quality standard of hazelnut kernels - Standard DDP-04,” Geneva, Switzerland, 2011.
- [13] L. F. Ciarmiello *et al.*, “Analysis of different European hazelnut (*corylus avellana* L.)

cultivars: Authentication, phenotypic features, and phenolic profiles,” *J. Agric. Food Chem.*, vol. 62, no. 26, pp. 6236–6246, Jul. 2014.

- [14] M. Pedrotti *et al.*, “Quality control of raw hazelnuts by rapid and non-invasive fingerprinting of volatile compound release,” *LWT*, vol. 143, no. February, p. 111089, 2021.
- [15] M. Locatelli *et al.*, “Chemotype and genotype chemometrical evaluation applied to authentication and traceability of ‘tonda Gentile Trilobata’ hazelnuts from Piedmont (Italy),” *Food Chem.*, vol. 129, no. 4, pp. 1865–1873, 2011.
- [16] S. Klockmann, E. Reiner, N. Cain, and M. Fischer, “Food Targeting: Geographical Origin Determination of Hazelnuts (*Corylus avellana*) by LC-QqQ-MS/MS-Based Targeted Metabolomics Application,” *J. Agric. Food Chem.*, vol. 65, no. 7, pp. 1456–1466, 2017.
- [17] F. Stilo, M. Cialie Rosso, S. Squara, C. Bicchi, C. Cordero, and C. Cagliero, “*Corylus avellana* L. Natural Signature: Chiral Recognition of Selected Informative Components in the Volatilome of High-Quality Hazelnuts,” *Front. Plant Sci.*, vol. 13, no. April, pp. 1–15, Apr. 2022.
- [18] C. Alasalvar, E. Pelvan, B. Bahar, F. Korel, and H. Ölmez, “Flavour of natural and roasted Turkish hazelnut varieties (*Corylus avellana* L.) by descriptive sensory analysis, electronic nose and chemometrics,” *Int. J. Food Sci. Technol.*, vol. 47, no. 1, pp. 122–131, 2012.
- [19] C. Alasalvar, A. Z. Odabasi, N. Demir, M. Ö. Balaban, F. Shahidi, and K. R. Cadwallader, “Volatiles and flavor of five Turkish Hazelnut varieties as evaluated by descriptive sensory analysis, electronic nose, and dynamic headspace analysis/gas chromatography-mass spectrometry,” *J. Food Sci.*, 2004.
- [20] M. Cialie Rosso *et al.*, “Evolution of potent odorants within the volatile metabolome of high-quality hazelnuts (*Corylus avellana* L.): evaluation by comprehensive two-dimensional gas chromatography coupled with mass spectrometry,” *Anal. Bioanal. Chem.*, vol. 410, no. 15, pp. 3491–3506, 2018.
- [21] F. Stilo *et al.*, “An effective chromatographic fingerprinting workflow based on comprehensive two-dimensional gas chromatography – Mass spectrometry to establish volatiles patterns discriminative of spoiled hazelnuts (*Corylus avellana* L.),” *Food Chem.*, vol. 340, p. 128135, Mar. 2021.
- [22] M. A. Koyuncu, A. Islam, and M. Küçük, “Fat and fatty acid composition of hazelnut kernels in vacuum packages during storage,” *Grasas y Aceites*, vol. 56, no. 4, pp. 263–266, 2005.
- [23] A. Turan, “Effect of drying methods on fatty acid profile and oil oxidation of hazelnut oil during storage,” *Eur. Food Res. Technol.*, vol. 244, no. 12, pp. 2181–2190, 2018.
- [24] S. M. Momchilova, S. P. Taneva, D. Zlatanov, G. A. Antova, M. J. Angelova-Romova, and E. Blagoeva, “Fatty acids, tocopherols and oxidative stability of hazelnuts during storage,” *Bulg. Chem. Commun.*, vol. 49, no. August, pp. 65–70, 2017.
- [25] J. L. Kinderlerer and S. Johnson, “Rancidity in hazelnuts due to volatile aliphatic aldehydes,” *J. Sci. Food Agric.*, vol. 58, no. 1, pp. 89–93, 1992.
- [26] A. Memoli *et al.*, “Effect of bug damage and mold contamination on fatty acids and

sterols of hazelnut oil," *Eur. Food Res. Technol.*, vol. 243, no. 4, pp. 651–658, Apr. 2017.

- [27] M. H. Azarbad and H. Jeleń, "Determination of Hexanal—an Indicator of Lipid Oxidation by Static Headspace Gas Chromatography (SHS-GC) in Fat-Rich Food Matrices," *Food Anal. Methods*, 2015.
- [28] F. Shahidi, "Headspace Volatile Aldehydes as Indicators of Lipid Oxidation in Foods," in *Advances in Experimental Medicine and Biology*, 2001, pp. 113–123.
- [29] S. Pastorelli, L. Torri, A. Rodriguez, S. Valzacchi, S. Limbo, and C. Simoneau, "Solid-phase micro-extraction (SPME-GC) and sensors as rapid methods for monitoring lipid oxidation in nuts," *Food Addit. Contam.*, vol. 24, no. 11, pp. 1219–1225, 2007.
- [30] D. Ghirardello *et al.*, "Effect of storage conditions on chemical and physical characteristics of hazelnut (*Corylus avellana* L.)," *Postharvest Biol. Technol.*, vol. 81, pp. 37–43, Jul. 2013.
- [31] D. Ghirardello *et al.*, "Phenolic composition, antioxidant capacity and hexanal content of hazelnuts (*Corylus avellana* L.) as affected by different storage conditions," *Postharvest Biol. Technol.*, vol. 112, pp. 95–104, 2016.
- [32] M. Locatelli, J. D. Coisson, F. Travaglia, M. Bordiga, and M. Arlorio, "Impact of Roasting on Identification of Hazelnut (*Corylus avellana* L.) Origin: A Chemometric Approach," *J. Agric. Food Chem.*, vol. 63, no. 32, pp. 7294–7303, 2015.
- [33] M. Özdemir, F. Açıktur, M. Yildiz, G. Biringen, T. Gürcan, and M. Löker, "Effect of roasting on some nutrients of hazelnuts (*Corylus Avellana* L.)," *Food Chem.*, vol. 73, no. 2, pp. 185–190, 2001.
- [34] M. Özdemir, "FACTORS INFLUENCING SHELF LIFE OF HAZELNUT," *Basımı-Published Gıda Teknol.*, vol. 3, no. 3, pp. 66–71, 1998.
- [35] H. D. Belitz, W. Grosch, and P. Schieberle, *Food chemistry*, 4th edition. Springer, Berlin, Heidelberg, 2009.
- [36] M. Cialiè Rosso, F. Stilo, S. Mascrez, C. Bicchi, G. Purcaro, and C. Cordero, "Shelf-Life Evolution of the Fatty Acid Fingerprint in High-Quality Hazelnuts (*Corylus avellana* L.) Harvested in Different Geographical Regions," *Foods*, vol. 10, no. 3, p. 685, Mar. 2021.
- [37] N. Frega, M. Mozzon, and G. Lercker, "Effects of free fatty acids on oxidative stability of vegetable oil," *JAOCS, J. Am. Oil Chem. Soc.*, vol. 76, no. 3, pp. 325–329, 1999.
- [38] S. Pastorelli, S. Valzacchi, A. Rodriguez, and C. Simoneau, "Solid-phase microextraction method for the determination of hexanal in hazelnuts as an indicator of the interaction of active packaging materials with food aroma compounds," *Food Addit. Contam.*, vol. 23, no. 11, pp. 1236–1241, Nov. 2006.
- [39] M. Tobiszewski, "Metrics for green analytical chemistry," 2016.
- [40] Y. Wang, J. O'Reilly, Y. Chen, and J. Pawliszyn, "Equilibrium in-fibre standardisation technique for solid-phase microextraction," *J. Chromatogr. A*, vol. 1072, no. 1, pp. 13–17, Apr. 2005.
- [41] B. Kolb and L. S. Ettre, *Static headspace-gas chromatography : theory and practice*. New York: Wiley-VCH, 2006.

- [42] C. Cagliero, C. Bicchi, C. Cordero, P. Rubiolo, B. Sgorbini, and E. Liberto, "Fast headspace-enantioselective GC-mass spectrometric-multivariate statistical method for routine authentication of flavoured fruit foods," *Food Chem.*, vol. 132, no. 2, pp. 1071–1079, 2012.
- [43] C. Cordero *et al.*, "Odorants quantitation in high-quality cocoa by multiple headspace solid phase micro-extraction: Adoption of FID-predicted response factors to extend method capabilities and information potential," *Anal. Chim. Acta*, vol. 1052, pp. 190–201, 2019.
- [44] B. Sgorbini, C. Cagliero, E. Liberto, P. Rubiolo, C. Bicchi, and C. Cordero, "Strategies for Accurate Quantitation of Volatiles from Foods and Plant-Origin Materials: A Challenging Task," *J. Agric. Food Chem.*, vol. 67, no. 6, pp. 1619–1630, Feb. 2019.
- [45] L. Nicolotti *et al.*, "Quantitative fingerprinting by headspace-Two-dimensional comprehensive gas chromatography-mass spectrometry of solid matrices: Some challenging aspects of the exhaustive assessment of food volatiles," *Anal. Chim. Acta*, vol. 798, pp. 115–125, 2013.
- [46] F. Stilo *et al.*, "Chromatographic Fingerprinting Enables Effective Discrimination and Identification of High-Quality Italian Extra-Virgin Olive Oils," *J. Agric. Food Chem.*, vol. 69, no. 31, pp. 8874–8889, Aug. 2021.
- [47] B. Kolb and L. S. Ettre, "Theory and practice of multiple headspace extraction," *Chromatographia*, vol. 32, no. 11–12, pp. 505–513, 1991.
- [48] A. Griglione *et al.*, "High-quality Italian rice cultivars: Chemical indices of ageing and aroma quality," *Food Chem.*, vol. 172, pp. 305–313, 2015.
- [49] C. Bicchi *et al.*, "Quantitative analysis of volatiles from solid matrices of vegetable origin by high concentration capacity headspace techniques: Determination of furan in roasted coffee," *J. Chromatogr. A*, vol. 1218, no. 6, pp. 753–762, 2011.
- [50] H. Brevard *et al.*, "Guidelines for the quantitative gas chromatography of volatile flavouring substances, from the Working Group on Methods of Analysis of the International Organization of the Flavor Industry (IOFI)," *Flavour Fragr. J.*, vol. 26, no. 5, pp. 297–299, 2011.
- [51] Eurachem, *Eurachem Guide: The Fitness for Purpose of Analytical Methods – A Laboratory Guide to Method Validation and Related Topics*. 2014.
- [52] Citac and Eurachem, *Quantifying Uncertainty in Analytical Measurement*, vol. 2nd. 2000.
- [53] Commission Decision 657/2002, "Commission Decision 2002/657/EC implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results," *Off. J. Eur. Union*, vol. L221, no. 23 May 1996, pp. 8–36, 2002.
- [54] M. Raczyk, D. Kmiecik, P. Schieberle, R. Przybylski, H. Jeleń, and M. Rudzińska, "Model studies on the formation of volatile compounds generated by a thermal treatment of steryl esters with different fatty acid moieties," *Food Res. Int.*, 2017.
- [55] P. Schieberle and W. Grosch, "Detection of monohydroperoxides with unconjugated diene systems as minor products of the autoxidation of methyl linoleate," *Z. Lebensm.*

Unters. Forsch., 1981.

- [56] J. S. Amaral, S. Casal, I. Citová, A. Santos, R. M. Seabra, and B. P. P. Oliveira, "Characterization of several hazelnut (*Corylus avellana* L.) cultivars based in chemical, fatty acid and sterol composition," *Eur. Food Res. Technol.*, vol. 222, no. 3-4, pp. 274-280, 2006.
- [57] F. STILO *et al.*, "Delineating the extra-virgin olive oil aroma blueprint by multiple headspace solid phase microextraction and differential-flow modulated comprehensive two-dimensional gas chromatography," *J. Chromatogr. A*, vol. 1650, p. 462232, Aug. 2021.
- [58] A. Neugebauer, M. Granvogl, and P. Schieberle, "Characterization of the Key Odorants in High-Quality Extra Virgin Olive Oils and Certified Off-Flavor Oils to Elucidate Aroma Compounds Causing a Rancid Off-Flavor," *J. Agric. Food Chem.*, vol. 68, no. 21, pp. 5927-5937, Apr. 2020.
- [59] European Commission, "Guidance document on analytical quality control and method validation for pesticide residues analysis in food and feed SANTE 11312/2021," *Sante/11312/2021*, 2021.
- [60] W. M. Irvine, "IUPAC," in *Encyclopedia of Astrobiology*, vol. 67, no. 10, Berlin, Heidelberg, Heidelberg: Springer Berlin Heidelberg, 2018, pp. 1-1.
- [61] D. A. Armbruster, M. D. Tillman, and L. M. Hubbs, "Limit of detection (LOD)/limit of quantitation (LOQ): Comparison of the empirical and the statistical methods exemplified with GC-MS assays of abused drugs," *Clin. Chem.*, vol. 40, no. 7 I, pp. 1233-1238, Jul. 1994.
- [62] T. Wenzl, J. Haedrich, A. Schaechtele, P. Robouch, and J. Stroka, *Guidance Document on the Estimation of LOD and LOQ for Measurements in the Field of Contaminants in Feed and Food. EUR 28099 EN*. 2016.
- [63] ISO International Organization for Standardization, "ISO/IEC 17025:2017 General requirements for the competence of testing and calibration laboratories." ISO/IEC, 2017.

Chapter 8

Enhanced Selectivity VOCs Detection Using Novel GC-QEPAS - EVOQUE Project

This chapter discusses the contribution to the European project entitled “EVOQUE” - Enhanced Selectivity VOCs Detection Using Novel GC-QEPAS. The project aims to develop a compact, cost-effective analyzer for detecting volatile organic compounds (VOCs) by integrating gas chromatography with a photonic sensor/detector. Activities planned during the first year of the project's four-year duration were conducted as part of this PhD. The industrial partner, SRA Instruments, is a member of the EVOQUE Project Consortium, alongside the University of Turin unit.

The contribution focuses on activities aimed at optimizing chromatographic methods, specifically adapting the GC-QEPAS system to operate with nitrogen as the carrier gas while ensuring compatibility with the QEPAS detector under development. Preliminary results from initial tests are presented, emphasizing the performance of various column geometries and the translation of methods from helium to nitrogen. These results include assessments of sensitivity, linearity, chromatographic efficiency, and elution temperature modulation, demonstrating the potential of the system. The chapter also outlines ongoing developments and future steps within the project.

8.1 Objectives and background of the EVOQUE project

The EVOQUE project is a novel initiative with the primary objective of developing an advanced photonic-based sensory system that outperforms current standards in various applications. The innovative EVOQUE system is designed to address the stringent requirements of at-line, on-line, and in-field monitoring for Agriculture, Food, Environmental Pollution, and Industrial Emissions. By leveraging newly developed photonic technologies, EVOQUE aims to revolutionize the analysis of complex gas and liquid mixtures, particularly volatile organic compounds (VOCs). Central to this system is the integration of Gas Chromatography (GC) with innovative photonic components, including Quartz Enhanced Photo-Acoustic Spectroscopy (QEPAS), long-wavelength Quantum Cascade Lasers (QCLs), and Mid-Infrared Metasurfaces.

8.1.1 The importance of VOCs

Plants emit various VOCs in response to changes in environmental conditions or abiotic and biotic stresses, making VOCs a key component of the plant metabolome, often referred to as the plant *volatiome* [1]. During postharvest storage, crops undergo metabolic processes such as senescence, sprouting, and ripening, which can lead to the production of stress-induced VOCs due to enzymatic or auto-catalytic reactions. These VOCs, emitted in response to stimuli like high water activity, elevated temperatures, or microbial activity, serve as indicators of quality and freshness. Odor-active VOCs, responsible for food aroma (pleasant or unpleasant), interact with olfactory receptors in the nose, triggering the sense of smell. When their concentration exceeds the odor threshold, they can attract (pleasant aromas) or repel (off-odors) humans.

VOCs are critical for assessing food quality and safety, as they evolved to signal safe, nutrient-rich food or warn against spoiled or contaminated items [2]. Microbial activity on crops produces VOC patterns that indicate spoilage and help identify the specific organisms responsible [3]. Proper storage and warehousing of crops are essential for ensuring year-round availability of high-quality materials, mitigating risks from climate or geopolitical events, and supporting food security policies [4]. Effective monitoring of VOCs can quickly identify quality or safety issues, enabling prompt corrective actions.

However, VOCs also pose environmental and health risks. High exposure levels can cause headaches, nausea, dizziness, and irritation, while prolonged exposure is linked to severe health effects, including organ damage, nervous system disorders, and cancer. Hazardous VOCs, such as benzene, toluene, and hexane [5], are emitted during combustion or from unburned natural gas, contributing to air pollution and the formation of ground-level ozone, which can cause respiratory problems.

In food processing, particularly during high-temperature methods like roasting or blanching, monitoring VOCs is crucial for detecting the early formation of harmful contaminants. This enables proactive adjustments to processing parameters, ensuring product safety. Additionally, VOCs formed during food processing contribute to the product

aroma blueprint [6], and their monitoring can guide process design to achieve desired sensory characteristics. In summary, the importance of VOCs spans over a wide range of applications due to their different roles:

- **Industrial processes:** VOCs serve as process indicators.
- **Food production:** They act as quality markers, technological indicators, and process contaminants.
- **Plant phenotyping:** VOCs function as botanical/geographical tracers and authenticity markers.
- **Crop storage:** VOCs provide shelf-life indicators.

8.2 *The EVOQUE project applications*

EVOQUE has identified four key challenges and linked use cases. Ensuring quality control during seed storage is crucial to maintaining optimal storage conditions and preserving seed quality. Hazelnuts, a key ingredient in the confectionery industry, are particularly vulnerable to quality degradation due to their high fat content, which makes them prone to auto-oxidation during storage. To safeguard quality and prevent contamination or spoilage, regular sampling and testing are necessary. Current industry practices involve physical inspections to identify defects such as rotten, insect-damaged, twin, or yellowed kernels [7]. These evaluations are conducted by trained personnel using panel tests on representative samples from incoming batches to determine acceptance or rejection. However, this method has significant limitations, including reliance on operator expertise, variability in human judgment, subjectivity in assessments, and time inefficiency.

While GC-MS has proven effective in providing objective measurements of hazelnut rancidity and off-flavors to complement sensory panel evaluations [8], it is impractical for at-line or on-line applications due to its high cost and the need for specialized laboratories. To address these challenges, EVOQUE will introduce at-line and on-line VOC analyzers capable of detecting and quantifying marker VOCs directly from the ambient air within storage containers. These VOCs, which are dynamically emitted during storage, will be monitored continuously over the entire storage period. Linear saturated aldehydes, known markers of storage quality, will be a focus of the study, particularly their formation through autocatalytic oxidation of unsaturated fatty acids and lipid hydrolysis into free fatty acids. The presence of worms, bacteria, or mold exacerbates these processes, leading to the production of VOCs with unpleasant odors[9].

As hazelnuts undergo natural lipid oxidation when stored in ambient conditions with air exposure, rancidity develops, accompanied by an increase in VOCs such as hexanal, heptanal, octanal, nonanal, and decanal [3]. Despite these advancements, the current standard quality control methods in the industry remain visual inspection and sensory testing, which have inherent limitations. EVOQUE's innovative VOC analyzers aim to revolutionize quality control by providing real-time, precise monitoring of storage conditions and improving the efficiency and reliability of quality assessment processes.

High-resolution fingerprinting of roasted coffee and hazelnuts is essential for ensuring quality control, authentication, and traceability of premium products. Roasting is a critical step in the production process for both coffee and hazelnuts, as it develops their distinctive flavor and aroma profiles. During roasting, heat induces complex reactions in non-volatile precursors, whose distribution in raw beans or kernels is characteristic of their botanical or geographical origin (*i.e.*, phenotype). These reactions produce compounds such as carbonyls (aldehydes and ketones), heterocycles (pyrazines, furans, pyrans), and aromatics (phenols), which contribute to the aroma blueprint and serve as a unique chemical fingerprint of the product. However, high temperatures or specific roasting technologies can also lead to the formation of process contaminants, including acrylamide, furan, and 2-methylfuran, which are regulated under EU policies [10]–[13].

Roasting causes both physical and chemical transformations in raw beans and kernels, allowing process monitoring through color indexing [14], while product qualification is typically conducted by trained sensory panels [15][15][15][15][15][15][15][15], [16] assessing the flavor profile. Despite the importance of quality control, current methods lack molecular resolution tools [17] capable of targeting specific process indicators or botanical tracers, such as VOCs, that provide insights into product roasting quality and authenticity. GC-MS profiling methods, while effective, are applied only to selected production batches in analytical laboratories, which delays the identification of non-conforming batches (*i.e.*, those with VOC fingerprints deviating from the reference). This limitation can negatively impact profitability by failing to ensure timely corrective actions. The ability to perform at-line checks for product compliance and alignment with reference benchmarks is increasingly valuable, particularly in the context of Industry 4.0 [18].

To address these challenges, the EVOQUE analyzer will enable at-line, high-resolution VOC fingerprinting to identify the specific chemical signature of products, facilitating decision-making and conformity verification directly at the roasting plant. This innovative tool will replace traditional at-line methods that infer molecular information from physical properties like temperature and color, while also complementing sensory panel evaluations. By providing precise fingerprinting for authentication, qualification, and discrimination, the EVOQUE analyzer will enhance efficiency and reliability in quality control processes.

Monitoring VOCs plays a crucial role in tracking environmental pollution. In the crude oil industry, VOCs such as H₂S and BTEX can escape at various stages, from extraction and transportation to storage and refining. Production and refinery sites are considered the second-largest source of VOC emissions, following vehicle exhaust in the transportation sector. While geothermal power plants are promoted as environmentally friendly energy sources, the steam they release contains non-condensable gases (NCGs), including H₂S, which can pose safety, health, and environmental risks if left uncontrolled. Additionally, H₂S is a by-product of natural gas purification in processing plants and is generated during the hydrodesulfurization of petroleum feedstocks in refineries. Beyond H₂S, VOCs such as ethane, propane, butane, pentane, and hexane, emitted from crude oil, can interact with NO_x in the atmosphere, contributing to the formation of ground-level ozone, also known as tropospheric ozone.

The detection of these VOCs is complicated by overlapping absorption bands and interference from atmospheric compounds, which limits the effectiveness of optical spectroscopy alone. The GC-QEPAS solution offers a robust alternative by separating compounds based on their retention times. For instance, while spectral interference between H₂S and CH₄ is unavoidable in optical detection, GC-QEPAS can elute H₂S at a time far removed from hydrocarbons and atmospheric interferents, effectively eliminating spectral interference. Similarly, detecting toluene optically requires complex multivariate data analysis due to CO₂ interference. With GC-QEPAS, quasi real-time measurements of such VOCs become feasible, providing a more reliable and efficient approach to environmental monitoring.

Discriminating alkanes such as methane (C₁), ethane (C₂), and propane (C₃) is crucial for monitoring reservoir health, improving emissions tracking in the oil and gas industry, and understanding the impacts of global warming, including the stability of natural gas hydrates. These alkanes fall within the "hydrocarbon region" (3000–4000 nm), where precise identification is hindered by significant spectral overlap, particularly with propane, which poses cross-sensitivity challenges due to both spectral and non-spectral interference. Real-time natural gas analysis is a valuable tool for enhancing the efficiency of oil well forecasting, minimizing the environmental impact of petroleum exploration, and advancing knowledge about global warming's effects. By leveraging the time separation of ethane and propane achieved through a GC column and utilizing long-wavelength QCLs, EVOQUE will provide accurate detection under optimal conditions, eliminating the need for extensive spectral scans and complex post-processing. This innovative approach ensures efficient and reliable monitoring of these critical hydrocarbons.

8.3 Current state of the art vs. the EVOQUE system

Gas Chromatography coupled with Mass Spectrometry (GC-MS) represents the current gold standard for VOC analysis. This analytical technique excels in profiling and fingerprinting VOCs through its superior specificity, sensitivity, and quantification accuracy. Despite its robustness, GC-MS has significant drawbacks:

- **Cost and complexity:** GC-MS systems are expensive and require specialized equipment.
- **Operational challenges:** Skilled operators are needed for analysis and maintenance.
- **Stationary constraints:** The high-vacuum requirements and intricate ion sources limit the portability and real-time applicability of GC-MS.

EVOQUE aims to overcome the limitations of GC-MS by combining GC with QEPAS, exploiting the strengths of both technologies:

- **GC:** Effective separation and concentration of VOCs from complex mixtures.
- **QEPAS:** High sensitivity, specificity, and quantification accuracy through infrared spectral analysis.

This combination enables the EVOQUE system to achieve two-dimensional selectivity, making it capable of identifying trace chemicals in the presence of interfering substances. Unlike simpler GC detectors such as Flame Ionization Detectors (FID) or Thermal

Conductivity Detectors (TCD), QEPAS provides molecular recognition based on the unique spectral features of VOCs. Furthermore, by separating compounds through GC before detection, the system eliminates common interferences caused by overlapping absorption features or atmospheric compounds like water vapor (H₂O) and carbon dioxide (CO₂). The EVOQUE analyzer's fundamental advantages over GC-MS include:

- **Operational flexibility:** Operating at pressures >200mB with simple pumping systems.
- **Maintenance efficiency:** Utilizing maintenance-free QCLs instead of complex ion sources.
- **Portability and scalability:** Enabling at-line and on-line deployment for real-time decision-making in quality control processes.

These attributes make EVOQUE a game-changing tool, particularly in scenarios where GC-MS remains impractical due to its size, cost, or maintenance demands. By addressing these barriers, EVOQUE ensures broader adoption and usability across strategic sectors.

8.4 EVOQUE system components

8.4.1 Portable Fast-GC

Conventional GCs rely on large air-bath ovens to heat the column, facilitating component separation through controlled temperature ramp rates. However, in portable GCs, where minimizing power consumption is critical, resistive heating is employed. This approach involves passing an electric current through a conductor, known as the heating element, which is in contact with the separation column. This enables rapid temperature changes and efficient cooling. In the EVOQUE system, the column will be integrated with a heated gas sampling valve at the front end to allow direct injection of gas samples ranging from 250 µL to 500 µL. At the outlet, a specialized coupling interface will be implemented to minimize dead volume, preserve separation efficiency, and prevent peak broadening, while carefully managing the column's flow and pressure parameters in relation to the sensor or detector.

8.4.2 QEPAS detector

Conventional photoacoustic spectroscopy (PAS) is a well-established trace gas detection method based on the photoacoustic effect. In this method, light energy absorbed by a gas is converted to heat that generally occurs via molecular collision-induced non-radiative relaxation of excited states. This absorption causes the gas to heat up and expand. If the light is chopped or intensity-modulated, the expansion produces pressure waves that can be detected using a microphone. The common approach used to detect the acoustic signal utilizes an acoustic resonator filled with the gas sample. In this way the absorbed laser power is accumulated in the acoustic mode of the resonator. However, background signals can originate from nonselective absorption of the gas cell windows and external acoustic noise, which can be reduced by means of a proper isolation of the cell from any mechanical vibrations. As alternative, QEPAS sensors are developed since 2002 and are based on the use of quartz tuning forks (QTFs) as optoacoustic transducers. This represents the first step of compaction for a photoacoustic sensor, replacing a microphone-equipped resonant cell with an acoustic detection module containing a QTF. The tuning fork coupled with resonator tubes

acting as pressure wave amplifier constitute a spectrophone. The QTF's resonance frequency and geometry have been further customized to better match the relaxation rates of the target analytes and provide a high adaptability to the spatial characteristics of the laser beam. An ADM with an internal volume $\sim 7 \text{ cm}^3$ is commercialized by Thorlabs GmbH. These sizes are not compatible with typical volumes (few microliters) and flow rates (few mL/min) of a compact-GC. Typically, wavelength modulation techniques are adopted to improve the signal-to-noise ratio (S/N) both in continuous scan mode across the absorption frequency or in line-locked mode. Recently, beat frequency QEPAS (BF-QEPAS) approach was developed as an alternative to QEPAS to simultaneously speed up and validate data acquisitions. Comparing BF-QEPAS with standard QEPAS, the frequency of light modulation is slightly detuned from the natural frequency of the signal transducer, and illumination rather triggers an acoustic pulse than a continuous acoustic wave. This subsequently causes free instead of forced vibrations of the signal transducer, which are demodulated at the frequency of light modulation. The demodulated signal will thus oscillate at a frequency equal to the difference of light modulation frequency and QTF resonance frequency. The benefit of such transient signal analysis is the simultaneous identification of analyte concentration and the real-time monitoring of the QTF resonance parameters, necessary to validate measurements. Moreover, BF-QEPAS reduces the overall scanning time by more than one order of magnitude with respect to standard QEPAS, requiring few seconds to scan a single absorption feature and identify the analyte. In EVOQUE, a novel ultra-compact ADM will be developed, utilizing a custom QTF optimized for BF-QEPAS and designed to operate with ultralow sample volumes and flow rates. Interband Cascade Lasers and long wavelength QCLs.

Lasers for spectroscopy applications need to exhibit single mode emission, which can be achieved in semiconductor lasers through Distributed Feedback (DFB) technology where a grating is used to modulate either the refractive index or the losses of the structure. DFB sources are very compact, but fabrication is complex and their tunability is limited (a few cm¹) which requires precise grating modelling and fabrication. EVOQUE's laser sources will be based on the innovative InAs/AlSb DFB-QCLs and InAs/GaInSb DFB-ICLs materials family especially well-suited for the mid-IR range. In particular, the InAs/AlSb QCL technology is the only semiconductor laser technology allowing efficient emission in the long-wave IR range between 12-20 μm , whereas the InAs/GaInSb ICL technology is the only semiconductor laser technology allowing efficient emission in the 3-4 μm range. Both technologies have already been successfully employed for spectroscopy, however their performances are not yet suitable for EVOQUE. The source developments will require working on the laser design (both material design and optical design), material growth, laser fabrication and laser characterization. Once the initial target wavelengths have been confirmed, the active zones will be designed to provide a large gain around the target wavelengths. They rely on complex heterostructures where adjacent layers have no common atoms, which requires particular attention in both the design and the epitaxial growth. A specific active zone and a specific DFB design will be needed for each wavelength. The active zone modelling will be performed using either proprietary or commercial softwares. Also the waveguide structure will be optimized using electromagnetic modelling.

The laser heterostructures will be grown by molecular-beam epitaxy (MBE), the only technique giving laser quality structures with these material systems. Fabry-Pérot QCLs and ICLs will first be fabricated at each wavelength to qualify the material and extract the parameters needed for designing the DFBs (gain width, operating conditions, losses, etc.). Given the properties of both technologies, we expect that all target wavelengths will be reached. However, dedicated work will be needed to achieve CW operation of the QCLs, which – depending on the wavelength – may require to adapt their design and processing steps, and/or to accept a compromise on other performances such as the operating temperature or the optical power. DFBs require to concomitantly design both the laser heterostructures (spacers, SCH, upper cladding thickness) and the DFB grating. The proposed solution is a surface grating on the top of the ridges, with many important parameters to be optimized: grating period, coupling with active region, ridge width, lateral structuration, grating depth, etc. COMSOL Multiphysics simulations will be performed to simulate the DFB behaviour, and the following parameters will be calculated: losses, coupling factor, confinement factor, Q-factor, lasing wavelength, effective refractive index. In parallel, a batch of QCLs and a batch of ICLs with different structures will be fabricated and studied to benchmark the simulations. Thorough characterizations of the first DFB will allow us to refine the model and to optimize the design. This will require several runs of fabrication/characterizations for each of the required wavelength. This is a large resource and material consuming task. The fabricated DFB lasers will be integrated with the QEPAS Acoustic Detection Module. The first QEPAS studies and developments will rely on already existing lasers sources. The developments carried out during the project will result in new laser sources targeting the specific wavelengths of the three use cases and delivering higher optical power ($P > 10$ mW). The usual solution is to operate the lasers in CW with a low frequency small signal modulation. The required electronic drivers and signal acquisition cards will be developed in the electronics work package. The system will consist in a DFB laser, an integrated electronic driver and a feedback loop that will lock the wavelength to the target wavelength. In such a system, the laser chip operates in pulsed mode, using very short current pulses (typically 10ns) with a large repetition frequency (typically 10 MHz). It will include low frequency modulation capability and will be simply driven by a Modbus interface and easily integrable in systems. Because of the stabilized wavelength this source can be used in a QEPAS detection system without the need to scan the wavelength. As a result, the response time of the detector, usually of the order of seconds, can be significantly improved by more than one order of magnitude. The acquisition frequency could then be increasing above 10 Hz. This frequency will allow to integrate the QEPAS sensor to a small volume GC column where the elution times are short. This innovative laser source is an enabling technology towards the realization of miniaturized GC-QEPAS analysers.

8.4.3 Metasurfaces

Metasurfaces are planar layers of nano and microstructures that can manipulate light and provide functionalities that go beyond what is provided by natural materials. Metasurfaces are particularly important in Mid IR due to the limited range of materials with suitable optical properties, while the longer wavelength makes the required lithography easier. In 2016 researchers at MTU realized a mid IR high contrast grating based on the germanium/strontium fluoride material system showing the potential of Mid IR gratings and

metasurfaces. Recently, free standing metasurfaces operating in the Mid IR have been demonstrated. In EVOQUE, a new fabrication process will develop and refine, it will be based on this latter process to realise polysilicon and germanium metasurfaces on transparent free-standing dielectric membranes. The process will be based on Deep Ultra-Violet lithography and the metasurfaces will be designed to be compatible with the process' resolution. Pillars and gratings on a thin film of Silicon nitride will compose the metasurface. Deep reactive ion etching from the backside will create an opening to realise the membrane. The diameter of the posts will be varied to control the amplitude and phase of the scattered light (any value in the $0-2\pi$ range). The divergence angles of the excitation laser will be determined and the metalens design adjusted to provide collimation or focusing and the desired beam diameter. The metasurface will be designed to improve the beam quality of the output beam based on the characteristics of the output beam of the QCL.

8.5 Aims of the EVOQUE project

In summary, the goal of the EVOQUE project is to develop a compact, user-friendly, and cost-effective analyzer for VOCs complex mixtures. The system is designed to offer performance comparable to laboratory-based GC-MS but without the high maintenance costs and the need for a skilled operator. Photonic sensors, such as those used in the project, offer the advantage of being non-invasive, non-destructive, and easy to operate. However, current photonic technologies, such as laser absorption spectroscopy, are not yet capable of competing with GC-MS, as they require large sample volumes to achieve high sensitivities, even for simple gases. While GC-MS remains the dominant technique for complex molecules like VOCs, which absorb weakly and could have complex IR spectra, EVOQUE's innovative combination of technologies (QEPAS, GC, long-wavelength lasers, and metasurfaces) provides a unique solution to the VOC detection challenge. The QEPAS detector is non-destructive and can be integrated with other GC systems, such as coupling the EVOQUE detector with a thermal conductivity detector, offering a complementary product line for industrial partners and enhancing the GC industry. This project provides a three-generation development approach. The *Gen zero* stage involves the integration of the existing QEPAS detector with a standard chromatographic system, where the focus will be on testing the discrimination of alkanes such as methane, ethane, and propane. The *Gen one* stage will focus on assembling a portable system for GC applications, featuring a more compact and miniaturized version of the detector compared to Gen zero. This stage will test selected VOC mixtures to verify GC performance using certified standard solutions, covering the full dynamic range expected for the target application. Additionally, the suitability of LODs and LOQs for the tested configurations and applications will be assessed. Finally, the *Gen two* stage will involve the creation of the final prototype, incorporating further miniaturization of the detector and preparing it for future at-line and online tests. Through these stages, the project aims to create a fully optimized, portable, and cost-effective system for VOC detection.

8.6 Gen zero GC-QEPAS development: Testing column geometries for targeted VOCs separation

The experimental phase of this project was designed to provide a comprehensive framework for the development and validation of an analytical methodology tailored to the specific requirements of GC-QEPAS technology. The primary objectives included the identification of optimal conditions for analyzing target samples using the first and second generations of GC-QEPAS, alongside an in-depth evaluation of chromatographic performance and sensitivity across various column configurations. A particular focus was on evaluating the transferability of methods and the consistency of column performance across different setups and geometry, ensuring reliability and reproducibility of results under different experimental conditions. To achieve these aims, model samples, including the volatile fraction of hazelnuts and particularly linear saturated aldehydes (C6-C10), were selected to provide a representative test case for the system's capabilities. This section delves into the experimental strategies, procedures, and technical considerations that were implemented to achieve these goals, highlighting the systematic approach adopted throughout the study.

In order to develop and assess the conditions to analyze target samples, several parameters were evaluated, divided into two categories: linearity&sensitivity and chromatographic efficiency. Metrics for linearity & sensitivity evaluation were as follows:

- **Linearity** is the ability of the method to elicit test results that are directly, or by a well-defined mathematical transformation, proportional to analyte concentration or quantity within a given range;
- **Limit of detection (LOD)** is the lowest concentration of an analyte in a sample that can be detected, not quantified;
- **Limit of quantification (LOQ)** is the lowest concentration of an analyte in a sample that can be determined (quantified) with acceptable precision and accuracy under the stated operational conditions of the methods.

While for chromatographic efficiency:

- **Alkanes discrimination** is a parameter relating to the performance of the gas chromatography system. It is calculated as the ratio of *n*-eicosane to other linear alkanes. The relative response (peak area) to *n*-eicosane should be at least 0.8. Otherwise, the discrimination of the injection system is too high and the injection system should be optimized or replaced;
- **Resolution** quantifies the separation between two peaks in relation to their widths, providing a measure of the column's ability in separating between two analytes [20]. It is calculated using Equation 8.1

Equation 8.1:

$$R_s = \frac{2(t_2 - t_1)}{W_1 + W_2}$$

Where t_1 and t_2 : the retention times of analytes 1 and 2

W_1 and W_2 : the peak widths of analytes 1 and 2

- **Peak width** is the length of the interpolated segment at the intersection of the tangents to the Gaussian flexes and the baseline. It indicates how broad or narrow a peak is, which in turn provides insight into the efficiency and resolution of the separation process. It also influences the sensitivity of the method; as wider peaks will result in a lower S/N ratio.
- **Separation Measure** is a chromatographic parameter defined by *Blumberg et al.*[21], used to evaluate the capacity of a chromatographic system to separate different compounds. Unlike resolution, which considers both the distance between peaks and their width, it focuses more on the efficiency of separation, taking into account various aspects such as peak symmetry and interference between peaks. It is calculated using the following equation.

Equation 8.2:

$$S = \frac{t_2 - t_1}{\sigma}$$

Where t_1 and t_2 : the retention times of analytes 1 and 2
 σ : standard deviation of the peak

The tests were conducted focusing just on one of the previously mentioned applications. Specifically, the volatile fraction of hazelnuts was used as the test matrix. Since hazelnuts themselves are a complex matrix, the focus was directed on four linear saturated aldehydes. Standard solutions at different concentrations of hexanal, heptanal, octanal, and nonanal were prepared, with concentrations of 0.5 ppm, 1 ppm, 5 ppm, 10 ppm, 50 ppm, and 100 ppm. The analyses were carried out on an Agilent 8890 system, with an FID detector used for these preliminary results, operating at an acquisition frequency of 20 Hz. The split/splitless injector was in pulsed-split mode at 250 kPa until 1 min with 1:5 split ratio, injector temperature 250 °C. For liquid injections, the injected volume was 1 µL. Regarding the sampling of real samples, the headspace solid-phase microextraction (HS-SPME) was performed using a Divinylbenzene/Carboxen/Polydimethylsiloxane (DVB/CAR/PDMS) fiber with a 50/30 µm film thickness and a length of 2 cm. The fibers, obtained from Merck, were conditioned prior to use as recommended by the manufacturer, through thermal desorption at 270°C for 60 minutes in a split/splitless injector under constant carrier gas flow. HS-SPME was performed on 1.00 ±0.05 g of finely ground hazelnuts (*i.e.*, *granella*) placed in a 20 mL glass sealed vial. Sampling was at 50 °C for 30 min. The carrier gas was helium at a constant flow depending on the columns' geometry. The stationary phase adopted in these preliminary analyses was a non-polar HP-5 ((5%-phenyl)-methylpolysiloxane). Column geometry, selected flows for helium used as carrier gas, and oven temperature ramp are listed in **Table 8.1**.

Table 8.1: Flows settings for nitrogen as carrier gas and oven temperature ramp for each column's geometry.

Column geometry	Helium Flow	Oven temperature ramp
20m × 0.18mm d _c , 0.18µm d _f	1 mL/min	35°C (3') to 160°C @ 20°C/min to 280°C (3') @ 35°C/min
30m × 0.25mm d _c , 0.25µm d _f	1.7 mL/min	35°C (5') to 55°C @ 5°C/min to 160°C @ 20°C/min to 280°C (3') @ 35°C/min
30m × 0.32mm d _c , 0.25µm d _f	2.6 mL/min	35°C (2') to 240°C (2') @ 25°C/min to 280°C (2') @ 35°C/min
30m × 0.53mm d _c , 3µm d _f	5 mL/min	35°C (1') to 95°C @ 5°C/min to 195°C @ 25°C/min to 260°C (1') @ 35°C/min
30m × 0.53mm d _c , 3µm d _f	10 mL/min	35°C (1') to 95°C @ 5°C/min to 195°C @ 25°C/min to 260°C (1') @ 35°C/min

An example of the chromatographic profile obtained using a 20m × 0.18mm d_c, 0.18μm d_f column is shown in **Figure 8.1**. Despite the complexity of the *volatilome* of raw hazelnuts, adequate separation was achieved for the four target compounds (C₆–C₉ aldehydes).

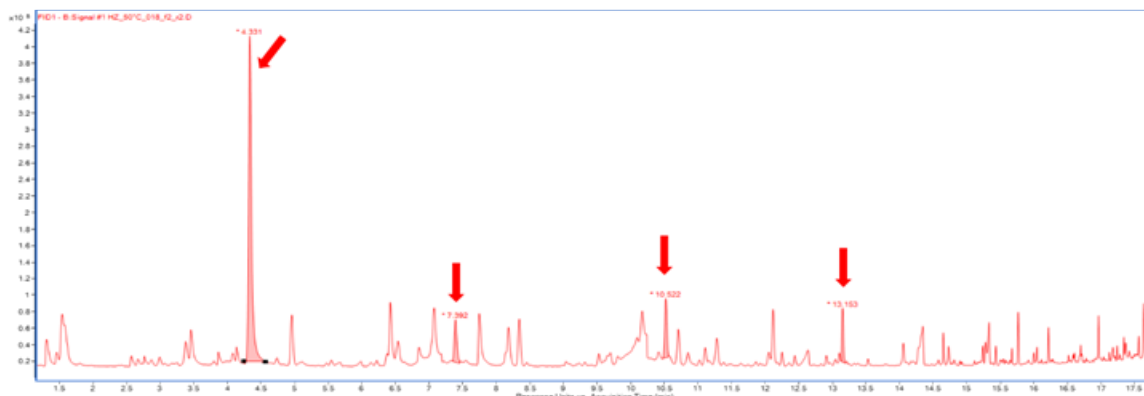


Figure 8.1: Example of raw hazelnuts volatilome chromatogram with a 20m × 0.18mm d_c, 0.18μm d_f column

8.6.1 Linearity

To assess linearity, a multi-point calibration curve was constructed. Analyses were performed at 6 concentration levels: 0.5 ppm, 1 ppm, 5 ppm, 10 ppm, 50 ppm, and 100 ppm. Each calibration level was acquired in triplicate, resulting in a total of 18 points used to generate the calibration curve. The chosen calibration curve type was linear. The equation of the curve and R² values for the calibration curves for each analyte are reported in **Table 8.2**.

Table 8.2: Equation of the calibration curve and R² values for each analyte in the different column geometries.

Column geometry	Compound	Equation	R ²
20m × 0.18mm d _c , 0.18μm d _f	Hexanal	y = 13068x + 2169.1	0.999
	Heptanal	y = 13592x + 9712.5	0.995
	Octanal	y = 13592x + 9712.5	0.995
	Nonanal	y = 12580x + 9857.3	0.995
30m × 0.25mm d _c , 0.25μm d _f	Hexanal	y = 15342x - 6.2186	0.999
	Heptanal	y = 14201x + 5868.4	0.999
	Octanal	y = 15784x + 9146.2	0.999
	Nonanal	y = 14249x + 9745.4	0.999
30m × 0.32mm d _c , 0.25μm d _f	Hexanal	y = 14563x - 6580.1	0.999
	Heptanal	y = 15834x - 4645.6	0.999
	Octanal	y = 16127x + 1071.1	0.999
	Nonanal	y = 16088x - 1123.1	0.999
30m × 0.53mm d _c , 3μm d _f (5 mL/min)	Hexanal	y = 14004x - 6018.2	0.998
	Heptanal	y = 15558x + 12963	0.998
	Octanal	y = 15313x - 3430	0.998
	Nonanal	y = 15182x - 4972	0.998
30m × 0.53mm d _c , 3μm d _f (10 mL/min)	Hexanal	y = 14036x - 5828.7	0.998
	Heptanal	y = 14837x - 8996.1	0.997
	Octanal	y = 14552x - 5580.5	0.998
	Nonanal	y = 13794x - 5947.3	0.997

The linearity results demonstrated consistent reliability across all tested setups, maintaining high performance even as the column geometry was progressively varied to 0.53mm d_c column. This indicates that the method is adaptable and robust, regardless of the column dimensions evaluated.

8.6.2 LOD and LOQ

The evaluation of the LOD and LOQ was conducted starting from the lowest concentration in the calibration curve. Dilutions were progressively prepared and analyzed to identify the concentration at which the signal-to-noise ratio reached a minimum of 3 for the LOD and 10 for the LOQ. This approach ensured accurate determination of the system's sensitivity while maintaining reproducibility. The process also provided valuable insights into the method's capability to detect and quantify low levels of the target analytes, highlighting its flexibility across varying concentration ranges. **Table 8.3** lists LOD and LOQ for each analyte in each column geometry.

Table 8.3: LOD and LOQ for each analyte in the different column geometries. Results are expressed in absolute amount in the sample.

Column geometry	Compound	LOD (ng)	LOQ (ng)
20m × 0.18mm d _c , 0.18μm d _f	Hexanal	1.00E-02	2.00E-02
	Heptanal	5.00E-03	1.50E-02
	Octanal	4.00E-03	1.20E-02
	Nonanal	2.00E-03	6.00E-03
30m × 0.25mm d _c , 0.25μm d _f	Hexanal	2.00E-02	8.00E-02
	Heptanal	1.50E-02	5.00E-02
	Octanal	4.00E-03	1.20E-02
	Nonanal	4.00E-03	1.20E-02
30m × 0.32mm d _c , 0.25μm d _f	Hexanal	5.00E-02	1.80E-01
	Heptanal	2.00E-02	5.00E-02
	Octanal	1.00E-02	3.00E-02
	Nonanal	1.00E-02	3.00E-02
30m × 0.53mm d _c , 3μm d _f (5 mL/min)	Hexanal	5.00E-02	2.00E-01
	Heptanal	3.00E-02	1.00E-01
	Octanal	2.00E-02	5.00E-02
	Nonanal	2.00E-02	5.00E-02
30m × 0.53mm d _c , 3μm d _f (10 mL/min)	Hexanal	5.00E-02	2.00E-01
	Heptanal	3.00E-02	1.00E-01
	Octanal	2.00E-02	5.00E-02
	Nonanal	2.00E-02	5.00E-02

As the diameter of the chromatographic columns increased from 0.18 mm d_c to 0.53mm d_c, a progressive peak broadening was observed, which directly impacted the sensitivity of the analysis. This broadening resulted in a decrease in the S/N, leading to higher absolute amounts required to reach the defined LOD and LOQ thresholds. Despite this trend, the results remained satisfactory across all tested geometries, demonstrating the method suitability for the scope. The observed increase in LOD and LOQ values highlights the inherent trade-off between column diameter and analytical sensitivity. Overall, the method proved adaptable and reliable, providing consistent performance even with the use of wider column diameters.

8.6.3 Alkanes discrimination

The following section focuses on the results on alkane discrimination, marking a shift from sensitivity-related parameters to those evaluating chromatographic efficiency. Alkane discrimination, a key indicator of system performance, is assessed as the ratio of the peak area of *n*-eicosane to other linear alkanes, with a minimum acceptable value of 0.8 for *n*-eicosane. Ratios below this threshold suggest excessive discrimination by the injection system, necessitating optimization or replacement. This evaluation emphasizes the system's ability to

achieve consistent and reliable separation, providing insights into the interplay between injection system performance and chromatographic efficiency. The analytes used for testing consisted of a mixture of linear alkanes ranging from C₉ to C₂₅. The results are reported in **Table 8.4**.

Table 8.4: Alkanes discrimination in each tested column geometry.

Analyte	20m × 0.18mm d _c , 0.18µm d _f	30m × 0.25mm d _c , 0.25µm d _f	30m × 0.32mm d _c , 0.25µm d _f	30m × 0.53mm d _c , 3µm d _f (5 mL/min)	30m × 0.53mm d _c , 3µm d _f (10 mL/min)
Nonane	94%	96%	112%	97%	90%
Decane	99%	101%	118%	100%	96%
Undecane	104%	106%	125%	107%	102%
Dodecane	106%	108%	127%	109%	105%
Tridecane	106%	108%	127%	110%	105%
Tetradecane	107%	109%	130%	111%	107%
Pentadecane	106%	108%	128%	108%	107%
Hexadecane	105%	106%	125%	107%	115%
Heptadecane	103%	104%	120%	108%	115%
Octadecane	102%	102%	114%	103%	105%
Nonadecane	100%	100%	107%	100%	100%
Icosane	100%	100%	100%	100%	111%
Heneicosane	101%	101%	98%	98%	99%
Docosane	99%	100%	97%	97%	98%
Tricosane	94%	94%	94%	90%	91%
Tetracosane	97%	97%	93%	89%	90%
Pentacosane	96%	97%	85%	89%	90%

As indicated by the results, all values were above the 80% threshold, indicating that no significant discrimination was present in the system. This confirms the reliability and consistency of the injection and separation processes for the alkane mixture analyzed. Overall, the method demonstrates adequate performance, ensuring accurate and reproducible results without bias toward specific compounds.

8.6.4 Resolution

The following section presents the results of the chromatographic resolution calculation for the target analytes. Resolution was assessed both in the standard mixture and in the real sample, focusing on the separation of adjacent peaks. In the standard mixture, resolution values were calculated specifically for the pairs of aldehydes: hexanal-heptanal, heptanal-octanal, and octanal-nonanal. In the real sample, the evaluation was based on the peaks closest to the aldehydes of interest, ensuring a comprehensive assessment of the method's performance in a complex matrix. An example of the peaks considered for resolution calculation is illustrated in **Figure 8.2**, providing a visual reference.

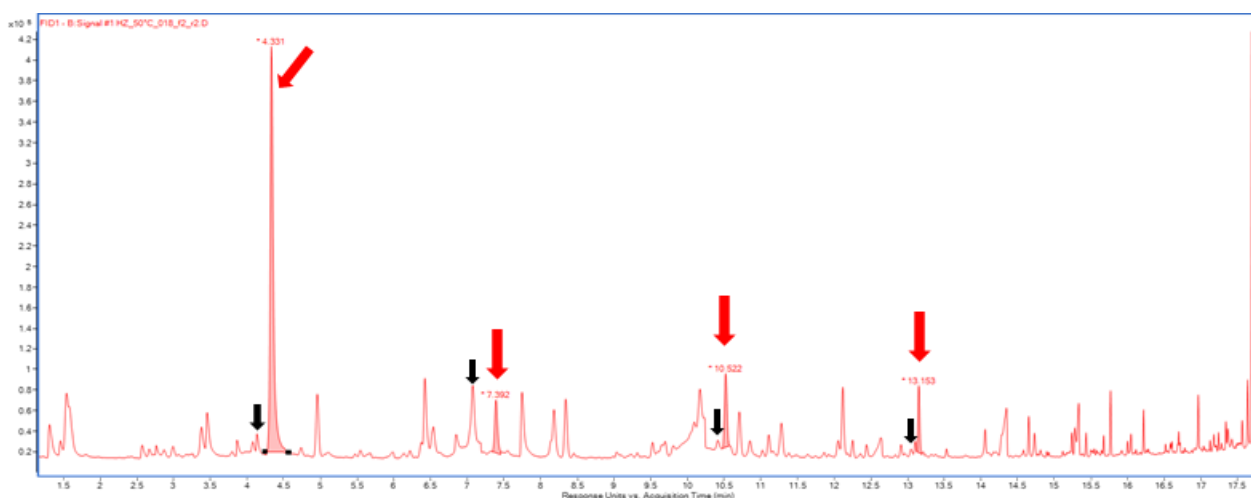


Figure 8.2: Chromatogram of the volatiles of raw hazelnuts obtained using a 20m × 0.18mm d_c, 0.18µm d_f. The target aldehydes are highlighted with red arrows, while black arrows indicate the adjacent peaks used for the resolution calculation.

The detailed results are listed in **Table 8.5** and **Table 8.6**, highlighting the method's ability to achieve sufficient separation under varying conditions. **Table 8.5** presents the results for the standard mixture, while **Table 8.6** shows the results for the real sample.

Table 8.5: Resolution calculated for the pairs of aldehydes in the standard solution.

Column geometry – Helium carrier gas	Pairs of aldehydes	Resolution R
20m × 0.18mm d _c , 0.18µm d _f	Hexanal-heptanal	48.14
	Heptanal-octanal	46.38
	Octanal-nonanal	42.84
30m × 0.25mm d _c , 0.25µm d _f	Hexanal-heptanal	58.65
	Heptanal-octanal	47.29
	Octanal-nonanal	42.07
30m × 0.32mm d _c , 0.25µm d _f	Hexanal-heptanal	23.64
	Heptanal-octanal	25.22
	Octanal-nonanal	27.08
30m × 0.53mm d _c , 3µm d _f (5 mL/min)	Hexanal-heptanal	27.45
	Heptanal-octanal	21.31
	Octanal-nonanal	20.12
30m × 0.53mm d _c , 3µm d _f (10 mL/min)	Hexanal-heptanal	26.99
	Heptanal-octanal	24.52
	Octanal-nonanal	20.11

Table 8.6: Resolution calculated for target aldehydes in the real sample.

Column geometry – Helium carrier gas	Compound	Resolution R
20m × 0.18mm d _c , 0.18µm d _f	Hexanal	2.93
	Heptanal	4.08
	Octanal	1.61
	Nonanal	1.20
30m × 0.25mm d _c , 0.25µm d _f	Hexanal	3.51
	Heptanal	1.90
	Octanal	2.93
	Nonanal	2.11
30m × 0.32mm d _c , 0.25µm d _f	Hexanal	1.45
	Heptanal	3.46
	Octanal	0.96
	Nonanal	2.11

Table 8.6: continue.

Column geometry – Helium carrier gas	Compound	Resolution R
30m × 0.53mm d _c , 3µm d _f (5 mL/min)	Hexanal	0.72
	Heptanal	0.88
	Octanal	0.69
	Nonanal	0.64
30m × 0.53mm d _c , 3µm d _f (10 mL/min)	Hexanal	0.86
	Heptanal	0.40
	Octanal	0.67
	Nonanal	0.56

In terms of resolution in the standard mixture, as expected a decrease in resolution values was observed with the increase in column diameter. However, an improvement was seen with the 30m × 0.53mm d_c, 3µm d_f column (thicker film) when a higher flow rate was used, increasing from 5 to 10 mL/min, resulting in values comparable to those obtained with the 30m × 0.32mm d_c, 0.25µm d_f column. Despite these variations, all values exceeded the 1.5 threshold [22], which is conventionally considered the optimal value for good chromatographic resolution. When focusing on real samples, the trend remained consistent with that observed for the standard mixture. However, as the column diameter increased, an increase in co-elutions was noted, as evidenced by resolution values falling below 1.5. These values are highlighted in bold in the table. A more detailed evaluation will be required unless the detector proves to be sufficiently sensitive and selective, as is the case with MS and ideally with QEPAS. In such cases, the need for further optimization of chromatographic conditions may be minimized.

8.6.5 Efficiency

Peak width is a critical parameter in chromatography, as it directly impacts both the efficiency and resolution of the separation process. It refers to the length of the interpolated segment at the intersection of the tangents to the Gaussian curves and the baseline, providing a measure of how broad or narrow a peak is. The width of a peak influences the sensitivity of the method, with wider peaks leading to a reduced S/N. Therefore, optimizing peak width is essential for achieving high-resolution separations while maintaining adequate sensitivity in chromatographic analyses. In this stage of the evaluation process, the width at half-height (w_{0.5}) was specifically calculated to assess the impact of different column geometries. This parameter is particularly crucial for understanding how peak width changes across various column configurations. **Table 8.7** lists the results for each analyte in the standard mix and the real sample for each configuration. For clearer interpretation, the results are presented in seconds.

Table 8.7: Width at half-height for target analytes in the standard mix and the real sample for each configuration. The results are presented in seconds.

Column geometry – Helium carrier gas	Compound	w _{0.5} standard solution (s)	w _{0.5} real sample (s)
20m × 0.18mm d _c , 0.18µm d _f	Hexanal	1.08 s	2.92 s
	Heptanal	0.85 s	2.57 s
	Octanal	0.76 s	2.31 s
	Nonanal	0.72 s	1.56 s
30m × 0.25mm d _c , 0.25µm d _f	Hexanal	2.82 s	3.86 s
	Heptanal	1.51 s	2.13 s
	Octanal	1.02 s	1.60 s
	Nonanal	0.87 s	1.48 s

Table 8.7: Continue

Column geometry – Helium carrier gas	Compound	w _{0.5} standard solution (s)	w _{0.5} real sample (s)
30m × 0.32mm d _c , 0.25µm d _f	Hexanal	3.00 s	7.07 s
	Heptanal	1.64 s	2.79 s
	Octanal	1.07 s	2.05 s
	Nonanal	0.87 s	1.71 s
30m × 0.53mm d _c , 3µm d _f (5 mL/min)	Hexanal	4.18 s	4.40 s
	Heptanal	2.70 s	2.56 s
	Octanal	1.77 s	1.90 s
	Nonanal	1.59 s	1.80 s
30m × 0.53mm d _c , 3µm d _f (10 mL/min)	Hexanal	4.40 s	4.39 s
	Heptanal	4.57 s	4.46 s
	Octanal	2.28 s	2.33 s
	Nonanal	1.84 s	2.03 s

The w_{0.5} showed a progressively increasing trend as the column diameter increased, in keeping with resolution results. Additionally, as the boiling point of the compounds increased, as well as the time spent within the column, the peak width decreased due to the focusing effect of the peak band within the chromatographic column that compensated for the band-broadening imparted by the injection dynamics and helium carrier gas compressibility and performance characteristics. To note, the temperature programming was tuned according to the method translation principles [23], the reference method was that validated for hazelnuts rancidity assessment discussed in Chapter 7. It is noteworthy that, for real samples, the w_{0.5} was consistently greater than that observed for the standards analyzed by liquid injection. This phenomenon is likely attributed to the time required for the analytes to desorb from the fiber and transfer into the column.

8.6.6 Separation measure

The Separation measure (S), as already mentioned, is a key parameter used to evaluate the efficiency of a chromatographic system in discriminating different compounds. Unlike resolution, which focuses on the distance and width of peaks, S emphasizes the overall effectiveness of the separation process. It takes into account factors such as peak symmetry and the degree of interference between adjacent peaks. In this study, the S was calculated using the first and last target aldehydes, specifically hexanal and nonanal, as the reference points for the first and last targeted peaks. This parameter was evaluated both in the standard mixture and in the real sample. The results are listed in **Table 8.8**, providing a comprehensive view of the system's separation efficiency under different conditions.

Table 8.8: Separation measure in standard solution and real samples. It is calculated using hexanal and nonanal as reference peaks.

Column geometry – Helium carrier gas	S standard solution	S real sample
20m × 0.18mm d _c , 0.18µm d _f	502	558
30m × 0.25mm d _c , 0.25µm d _f	488	468
30m × 0.32mm d _c , 0.25µm d _f	234	212
30m × 0.53mm d _c , 3µm d _f (5 mL/min)	248	234
30m × 0.53mm d _c , 3µm d _f (10 mL/min)	314	298

In conclusion, a decrease in separation measure was observed as the internal diameter of the columns increased. As with peak width, the difference between the values calculated for the standard solution and the real sample can likely be attributed to the effects of liquid

injection vs. SPME fiber analyte desorption in the S/SL injector. For the real sample, the S consistently yielded lower values compared to those obtained from the standard solution. This may be due to the additional time required for desorption and transfer of the analytes from the fiber to the column, leading to peak broadening and, consequently, less efficient separation.

8.7 Preliminary conclusions and future work

In the previous section (Section 8.6), preliminary results were presented regarding the transferability of the method and the performance of different column geometries, using helium as the carrier gas, which is considered a gold standard for GC-MS. Building on these initial findings, the focus of this second phase of the study shifted toward developing chromatographic methods that would meet the specific requirements for coupling the chromatographic system with the QEPAS detector. This phase needed a more customized approach, addressing the unique requirements of the system to ensure optimal performance.

A critical aspect of this phase was the flow rate required for the two generations of the QEPAS detector. For Gen one, a flow rate of 10 mL/min is necessary within the detector cell, while for Gen two, a reduced flow rate of 5 mL/min could be applied. As a result, the attention turned to selecting column geometries that support these flow rates without compromising chromatographic efficiency. Specifically, columns with an inner diameter of 0.53 mm d_c were chosen for Gen one, and columns with an internal diameter of 0.32 mm d_c were selected for Gen two. This setting was based on the need to balance the flow rate requirements with the desired chromatographic performance.

Another key consideration in this phase was the use of nitrogen as the carrier gas. It is well known that, according to the Van Deemter curve, nitrogen offers less flexibility in linear velocity within the column compared to helium and hydrogen while maintaining the same chromatographic efficiency. To address this challenge, a compromise was made between resolution and chromatographic efficiency to accommodate the flow rate requirements. Specifically, for Gen one, the flow rate of 10 mL/min and for Gen two, 5 mL/min were achieved, while maintaining acceptable chromatographic performance.

In addition to these considerations, a maximum temperature limit of 120°C was imposed for the QEPAS detector. This temperature constraint is significant, as it restricts the elution of semi-volatile analytes. To mitigate this issue, a back-flush system will be implemented to prevent these analytes from remaining in the column and eluting during subsequent chromatographic runs. The temperature limit is particularly significant in GC, as elevated temperature rates are often employed to enhance resolution and sensitivity. However, this approach can result in analytes eluting at temperatures significantly higher than their boiling points, which is not a viable option at this stage of the study. Therefore, a careful tuning of the temperature programming was conducted to achieve the best chromatographic performance while complying for the temperature limit of the QEPAS detector.

Finally, an additional compromise was required concerning the acquisition frequency of the new detector. The QEPAS detector operates with a maximum acquisition frequency of

10 Hz. This necessitated a careful evaluation of peak width. For proper resolution in terms of acquisition, a peak (specifically a Gaussian peak) requires at least 10 to 12 data points [24]. As a result, highly compressed peaks, which would lead to narrower and thinner bands, could present challenges in achieving good resolution with a 10 Hz acquisition frequency. To ensure optimal performance, it was determined that peaks should have a minimum width of at least 2 seconds to ensure the acquisition frequency was sufficient for accurate detection and resolution.

This stage of the study highlights the need for a careful balance between various chromatographic parameters, including flow rate, temperature, and acquisition frequency, to ensure the optimal performance of the QEPAS detector in combination with the chromatographic system. The results of these adjustments will be critical for advancing the analytical capabilities of the system in future applications.

The sampling conditions used in this study were those previously tested and optimized, with results detailed in the preceding section (Section 8.6). The parameters remained consistent in this phase, during which the second food application, coffee (Section 8.2), was also tested. For the analysis of real samples, 1 ± 0.05 g of ground coffee and hazelnuts were used, ensuring uniformity across tests. Additionally, for the evaluation of standard mixtures, a solution was prepared containing selected analytes tailored to each application. Four linear aldehydes were used for the hazelnut application, while six different analytes, previously identified as discriminative for the geographical origins of Arabica coffee [25], were utilized for the coffee application. The coffee mixture included the following analytes: 2-methylpyrazine, furfural, furfuryl alcohol, 2,5-dimethyl pyrazine, 5-methyl furfural, and furfuryl acetate. Standard solutions at different concentrations were prepared, with concentrations of 0.5 ppm, 1 ppm, 5 ppm, 10 ppm, 50 ppm, and 100 ppm in cyclohexane as solvent. The analyses were carried out on an Agilent 8890 system coupled with an FID, operating at an acquisition frequency of 20 Hz. The split/splitless injector was in split mode with 1:5 split ratio, injector temperature of 250 °C. For liquid injections, the injected volume was 1 µL. The carrier gas was nitrogen at a constant flow depending on the columns' geometry. The stationary phase adopted in these analyses was an apolar HP-1 (methylpolysiloxane). Considering the flow requirements, which necessitated the use of 0.53 mm d_c and 0.32 mm d_c columns, the results presented in this section will focus exclusively on these two column geometries. However, a preliminary screening was conducted using a 0.18 mm d_c column with nitrogen as the carrier gas for both applications. This initial evaluation served as a benchmark method for assessing chromatographic performance under nitrogen-based conditions. Following this screening, the method was adapted and translated to the aforementioned column geometries for both applications, ensuring compatibility with the specific requirements of each system. Column geometry, selected flows, application, and oven temperature ramps are listed in **Table 8.9**.

Table 8.9: Flows settings and oven temperature ramp for each column geometry (hazelnut and coffee).

Column	Application	Flow rate	Oven temperature ramp
20m × 0.18mm d _c , 0.18μm d _f	Hazelnut	1 mL/min*	40°C (3') to 80°C @10°C/min to 160 @15°C/min to 280°C (3') @35°C/min
20m × 0.18mm d _c , 0.18μm d _f	Coffee	1 mL/min*	40°C (3') to 160 @15°C/min to 280°C (3') @35°C/min
30m × 0.53mm d _c , 0.5μm d _f	Hazelnut	10 mL/min	40°C (2.5') to 80°C @ 12°C/min to 160°C @ 17°C/min to 280°C (3') @ 42°C/min
30m × 0.53mm d _c , 0.5μm d _f	Coffee	10 mL/min	35°C (2.5') to 80°C @ 12°C/min to 160°C @ 17°C/min to 280°C (3') @ 42°C/min
30m × 0.32mm d _c , 0.25μm d _f	Hazelnut	5 mL/min	40°C (2') to 80°C @ 15°C/min to 160°C @ 22°C/min to 280°C (3') @ 52°C/min
30m × 0.32mm d _c , 0.25μm d _f	Coffee	5 mL/min	40°C (2') to 160°C @ 22°C/min to 280°C (3') @ 52°C/min

*conditions used in the benchmark method tests

An example of chromatograms obtained using the benchmark methods with nitrogen as the carrier gas is shown in **Figure 8.3** for hazelnuts (A) and coffee (B), respectively. These benchmark methods were subsequently adapted and translated to the selected column geometries (0.53 mm d_c and 0.32 mm d_c) compatible with Gen one and Gen two systems, which were then used for further analyses specific to each application.

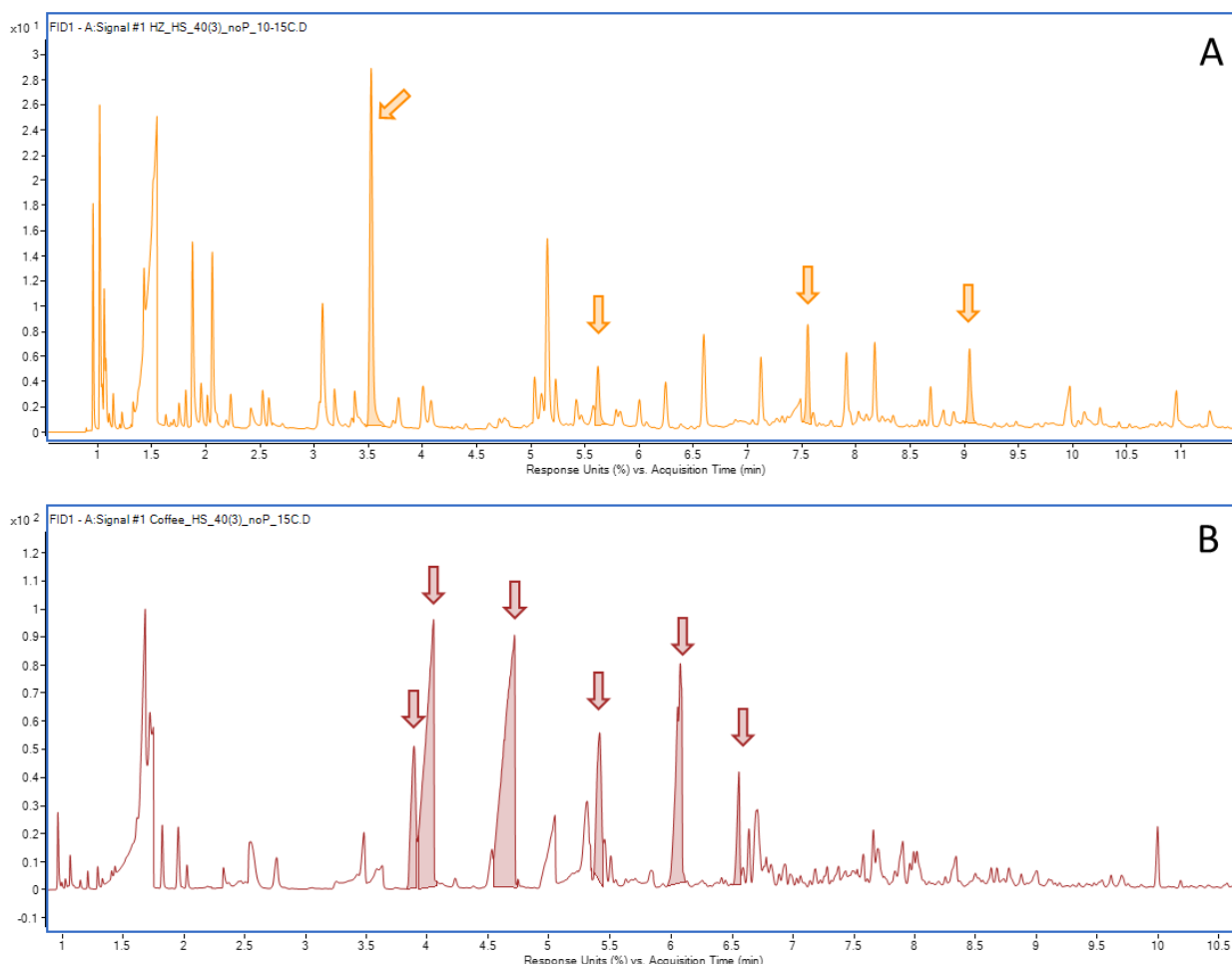


Figure 8.3: chromatograms of hazelnuts (A) and coffee (B) obtained using the benchmark methods with nitrogen as the carrier gas. Target analytes in both applications are highlighted with arrows

The use of method translation facilitated the transition from the benchmark method, employing a narrow-diameter column, to methods utilizing wider-diameter columns while preserving the elution order and original method's resolution. The tests were subsequently

conducted, and the results are summarized in **Table 8.10** for the hazelnut application and in **Table 8.11** for the coffee application.

Table 8.10: Elution temperature, LOD, LOQ, equation of the curve, R^2 , and average peak width of aldehydes in 0.53mm d_c , and 0.32mm d_c column for hazelnut application.

Column geometry	Compound	$T_{elution}$	LOD	LOQ	Equation (1-100 ppm)	R^2	Av. Peak width (s)
30m×0.53mm d_c , 0.5 μ m d_f	Hexanal	46°C	0.4 ng	1 ng	$y = 1277x - 4051$	0.997	4.38 s
	Heptanal	66°C	0.4 ng	1 ng	$y = 1051x - 2348$	0.999	5.28 s
	Octanal	86°C	0.2 ng	0.6 ng	$y = 1087x - 683$	0.999	6.54 s
	Nonanal	107°C	0.1 ng	0.4 ng	$y = 1201x + 1323$	1	7.38 s
30m×0.32mm d_c , 0.25 μ m d_f	Hexanal	48°C	0.02 ng	0.6 ng	$y = 1621x + 2123$	0.999	2.67 s
	Heptanal	68°C	0.02 ng	0.6 ng	$y = 1440x - 2456$	0.998	3.12 s
	Octanal	89°C	0.02 ng	0.6 ng	$y = 1324x + 473$	0.999	3.96 s
	Nonanal	110°C	0.1 ng	0.2 ng	$y = 1477x - 840$	0.999	3.75 s

Table 8.11: Elution temperature, LOD, LOQ, equation of the curve, R^2 , and average peak width of aldehydes in 0.53mm d_c , and 0.32mm d_c column for coffee application.

Column geometry	Compound	$T_{elution}$	LOD	LOQ	Equation (1-100 ppm)	R^2	Av. Peak width (s)
30m×0.53mm d_c , 0.5 μ m d_f	2-Methyl-pyrazine	49°C	0.1 ng	0.3 ng	$y = 14292 x - 29178$	0.998	8.64 s
	Furfural	50°C	0.1 ng	0.3 ng	$y = 14004 x + 4533$	0.999	7.68 s
	Furfuryl alcohol	57°C	0.1 ng	0.2 ng	$y = 11276 x - 36763$	0.994	7.68 s
	2,5-Dimethyl pyrazine	67°C	0.02 ng	0.1 ng	$y = 16329 x - 36634$	0.993	7.86 s
	5-Methyl furfural	75°C	0.02 ng	0.1 ng	$y = 17429 x - 44498$	0.992	9.06 s
	Furfuryl acetate	83°C	0.02 ng	0.1 ng	$y = 10425 x - 15973$	0.999	8.82 s
30m×0.32mm d_c , 0.25 μ m d_f	2-Methyl-pyrazine	55°C	0.1 ng	0.3 ng	$y = 15313 x - 28330$	0.994	2.16 s
	Furfural	57°C	0.1 ng	0.3 ng	$y = 13378 x - 10867$	0.997	3.45 s
	Furfuryl alcohol	65°C	0.02 ng	0.1 ng	$y = 13172 x - 25791$	0.994	4.17 s
	2,5-Dimethyl pyrazine	76°C	0.01 ng	0.02 ng	$y = 17689 x - 23989$	0.995	5.25 s
	5-Methyl furfural	86°C	0.01 ng	0.02 ng	$y = 18672 x - 30890$	0.994	6.38 s
	Furfuryl acetate	94°C	0.01 ng	0.02 ng	$y = 10993 x - 16526$	0.994	4.29 s

The results obtained from the EVOQUE project demonstrate that all the key requirements outlined in the initial phases have been successfully met. The chromatographic system was optimized to operate with nitrogen as the carrier gas, maintaining flow rates of 10 mL/min and 5 mL/min, as specified. Sensitivity parameters, including LOD and LOQ, showed highly promising outcomes, even though further validation with the final detector is still required. The linearity of the system proved to be excellent. The peak widths consistently exceeded the critical threshold of 2 seconds, ensuring compatibility with the detector's acquisition frequency of 10 Hz, a fundamental aspect of the system's design.

Temperature constraints were also respected, with maximum elution temperatures remaining well within the 120°C limit. The only exception was nonanal, which approached this threshold with values of 107°C and 110°C but still remained compliant.

These results represent a significant milestone in the development of the EVOQUE project. As an initiative in its early stages, it lays a solid foundation for further advancements and refinements, which will continue to evolve in the coming years, addressing both current challenges and future opportunities.

References

- [1] C. Bicchi and M. Maffei, "The plant volatilome: Methods of analysis," *Methods Mol. Biol.*, vol. 918, no. August, pp. 289–310, 2012.
- [2] S. A. Goff and H. J. Klee, "Plant volatile compounds: Sensory cues for health and nutritional value?," *Science* (80-.), vol. 311, no. 5762, pp. 815–819, 2006.
- [3] F. Stilo *et al.*, "An effective chromatographic fingerprinting workflow based on comprehensive two-dimensional gas chromatography – Mass spectrometry to establish volatiles patterns discriminative of spoiled hazelnuts (*Corylus avellana* L.)," *Food Chem.*, vol. 340, p. 128135, Mar. 2021.
- [4] Khartoum, "Towards Strategic Planning in Food-Grain Warehousing: the Role of Large-Scale Grain Warehouse Facilities on Markets' Mechanisms and Impact on Food Security Sudan Integrated Food Security Information for Action (Fao-Sifsia and Fsts)," no. June, 2011.
- [5] D. R. Michanowicz *et al.*, "Home is Where the Pipeline Ends: Characterization of Volatile Organic Compounds Present in Natural Gas at the Point of the Residential End User," *Environ. Sci. Technol.*, vol. 56, no. 14, pp. 10258–10268, 2022.
- [6] M. Granvogl and P. Schieberle, "The sensomics approach: A useful tool to unravel the genuine aroma blueprint of foods and aroma changes during food processing," in *Comprehensive Analytical Chemistry*, 1st ed., vol. 96, Elsevier B.V., 2022, pp. 41–68.
- [7] S. A. Mehlenbacher, D. C. Smith, and L. K. Brenner, "Variance Components and Heritability of Nut and Kernel Defects in Hazelnut," *Plant Breed.*, vol. 110, no. 2, pp. 144–152, 1993.
- [8] S. Squara *et al.*, "Validation of a high-throughput method for the accurate quantification of secondary products of lipid oxidation in high-quality hazelnuts (*Corylus avellana* L.): A robust tool for quality assessment," *J. Food Compos. Anal.*, vol. 114, p. 104766, Dec. 2022.
- [9] S. Squara *et al.*, "Artificial Intelligence decision-making tools based on comprehensive two-dimensional gas chromatography data: the challenge of quantitative volatilomics in food quality assessment," *J. Chromatogr. A*, vol. 1700, p. 464041, Jul. 2023.
- [10] C. Bicchi *et al.*, "Quantitative analysis of volatiles from solid matrices of vegetable origin by high concentration capacity headspace techniques: Determination of furan in roasted coffee," *J. Chromatogr. A*, vol. 1218, no. 6, pp. 753–762, 2011.
- [11] T. M. Amrein, H. Lukac, L. Andres, R. Perren, F. Escher, and R. Amadò, "Acrylamide in roasted almonds and hazelnuts," *J. Agric. Food Chem.*, vol. 53, no. 20, pp. 7819–7825, 2005.
- [12] C. Cagliero, T. D. Ho, C. Zhang, C. Bicchi, and J. L. Anderson, "Determination of acrylamide in brewed coffee and coffee powder using polymeric ionic liquid-based sorbent coatings in solid-phase microextraction coupled to gas chromatography-mass spectrometry," *J. Chromatogr. A*, vol. 1449, pp. 2–7, 2016.
- [13] G. Strocchi, P. Rubiolo, C. Cordero, C. Bicchi, and E. Liberto, "Acrylamide in coffee:

What is known and what still needs to be explored. A review," *Food Chem.*, vol. 393, no. June, p. 133406, Nov. 2022.

- [14] M. R. Ruosi *et al.*, "A Further Tool To Monitor the Coffee Roasting Process: Aroma Composition and Chemical Indices," *J. Agric. Food Chem.*, vol. 60, no. 45, pp. 11283–11291, Nov. 2012.
- [15] J. Kiefl and P. Schieberle, "Evaluation of process parameters governing the aroma generation in three hazelnut cultivars (*Corylus avellana* L.) by correlating quantitative key odorant profiling with sensory evaluation," *J. Agric. Food Chem.*, vol. 61, no. 22, pp. 5236–5244, 2013.
- [16] L. Nicolotti, C. Cordero, C. Bicchi, P. Rubiolo, B. Sgorbini, and E. Liberto, "Volatile profiling of high quality hazelnuts (*Corylus avellana* L.): Chemical indices of roasting," *Food Chem.*, vol. 138, no. 2–3, pp. 1723–1733, 2013.
- [17] E. Commission, "Commission Implementing Regulation (EU) 2021/808 of 22 March 2021 on the performance of analytical methods for residues of pharmacologically active substances used in food-producing animals and on the interpretation of results as well as on the methods to," *Off. J. Eur. Union*, vol. 180, no. 401, pp. 84–109, 2021.
- [18] G. Culot, G. Nassimbeni, G. Orzes, and M. Sartor, "Behind the definition of Industry 4.0: Analysis and open questions," *Int. J. Prod. Econ.*, 2020.
- [19] "<https://knepublishing.com/index.php/KnE-Life/article/view/2547/5464>." .
- [20] D.-W. Katja, *Practical Gas Chromatography: A Comprehensive Reference*. 2014. .
- [21] L. M. Blumberg and M. S. Klee, "Metrics of separation in chromatography," *J. Chromatogr. A*, vol. 933, no. 1–2, pp. 1–11, 2001.
- [22] R. L. Grob and E. F. Barry, *Modern Practice of Gas Chromatography*, vol. 29, no. 9. 1978.
- [23] M. S. Klee and L. M. Blumberg, "Theoretical and Practical Aspects of Fast Gas Chromatography and Method Translation," *J. Chromatogr. Sci.*, vol. 40, no. 5, pp. 234–247, 2002.
- [24] W. Jennings, E. Mittlefehldt, and P. Strempel, *Analytical Gas Chromatography*, Second. Folsom, California: J W scientific, 1997.
- [25] D. Bressanello *et al.*, "Chemometric Modeling of Coffee Sensory Notes through Their Chemical Signatures: Potential and Limits in Defining an Analytical Tool for Quality Control," *J. Agric. Food Chem.*, vol. 66, no. 27, pp. 7096–7109, 2018.

Experimental Section II

Maximizing the Information Power by Multidimensional Techniques and AI Aided Strategies

Chapter 9

Artificial Intelligence Decision-Making Tools Based on Comprehensive Two-Dimensional Gas Chromatography Data: The Challenge of Quantitative Volatilomics in Food Quality Assessment

Effective investigation of food volatilome by comprehensive two-dimensional gas chromatography with parallel detection by mass spectrometry and flame ionization detector (GC×GC-MS/FID) gives access to valuable information related to industrial quality. However, without accurate quantitative data, results transferability over time and across laboratories is prevented. The study applies quantitative volatilomics by multiple headspace solid phase microextraction (MHS-SPME) to a large selection of hazelnut samples (*Corylus avellana* L. n=207) representing the top-quality selection of interest for the confectionery industry. By untargeted and targeted fingerprinting, performant classification models validate the role of chemical patterns strongly correlated to quality parameters (i.e., botanical/geographical origin, post-harvest practices, storage time and conditions). By quantification of marker analytes, Artificial Intelligence (AI) tools are derived: the augmented smelling based on sensomics with blueprint related to key-aroma compounds and spoilage odorant; decision-makers for rancidity level and storage quality; origin tracers. By reliable quantification AI can be applied with confidence and could be the driver for industrial strategies.

This chapter was published as Original Research Paper in the Journal of Chromatography A: S. Squara, A. Caratti, A. Fina, E. Liberto, N. Spigolon, G. Genova, G. Castello, I. Cincera, C. Bicchi, C. Cordero, Artificial Intelligence decision-making tools based on comprehensive two-dimensional gas chromatography data: the challenge of quantitative volatilomics in food quality assessment, J. Chromatogr. A. 1700 (2023) 464041. DOI: 10.1016/j.chroma.2023.464041.

9.1 Introduction

Consumer acceptability of food is significantly impacted by its pleasant aroma shaped by the presence of key-aroma compounds and potent odorants within the complex fraction of volatile organic compounds [1]. The so-called aroma *blueprint* unequivocally evokes in humans the product's identity [2]. Although the majority of volatile compounds do not exert a direct effect on aroma perception, their potential in informing about product quality and authenticity/identity is relevant. Analytical strategies capable of effectively investigating the so called *food volatilome* (or *volatome*) which “contains all of the volatile metabolites as well as other volatile organic and inorganic compounds that originate from an organism” super-organism, or ecosystem [3], are of great interest for the food industry. By this fraction, primary materials, semi-finished and finished products can be monitored based on their botanical/geographical characteristics (*i.e.*, the phenotype), the impact of processing, or shelf life.

In hazelnuts (*Corylus avellana* L.), a primary material used by the confectionery industry, the volatilome encrypts information about many quality features. Native volatiles, such terpenes and terpenoids [4], are a direct expression of the plant phenotype, informing also for the presence of bacteria and moulds [5]. Monoterpenoids in raw hazelnuts are α -pinene, β -pinene, (*E*)-p-2-menthen-1-ol, camphene, δ -carene, α -thujene, γ -terpinene, sabinene, limonene, *cis*-sabinene hydrate, α -terpinolene, β -phellandrene, and *p*-cymene [6]–[8].

Another class of volatiles with great informative power on hazelnut quality and sensory profile is that of secondary products of lipid oxidation. They derive from lipids autoxidation by β -scission and hydroperoxide epi-dioxide decomposition of fatty acids hydroperoxides (*i.e.*, primary oxidation products) [9]. Within this class, linear saturated aldehydes, unsaturated aldehydes, and methyl-ketones might have sensory relevance conferring unpleasant aroma qualities (*e.g.*, *rancid*, *fatty*, and *waxy* notes). Moreover, although connoted by high odor thresholds (*i.e.*, lower odor potency), alcohols, and short-chain fatty acids [9], [10] are also formed during hydroperoxides degradation; they give a contribution to the informative potential of the volatilome as markers of lipid stability during storage of hazelnuts [6], [11], [12].

Post-harvest practices are known to have a role in kernel stability over time; high water activity (a_w), uncontrolled temperatures during drying, and substrates availability might promote enzymatic activation of the seed as well as the development/growth of bacteria and moulds further increasing the volatilome chemical complexity. Moreover, inappropriate post-harvesting is generally correlated to the off-flavors due to the presence of highly impacting odorants belonging to the classes of primary and secondary alcohols, carboxylic acids [6], lactones, furans, and some aromatics (benzaldehyde, phenyl ethyl alcohol, phenyl ethyl acetaldehyde, alkylated phenol derivatives) [5], [11], [13], [14].

Nowadays, raw hazelnuts are checked for their quality and the absence of sensory defects by visual inspection by trained staff and evaluation by a sensory panel. Representative samples from all incoming batches are tested before their acceptance or rejection. This approach, although highly specific since performed by a direct measure of the target quality feature (*i.e.*, morphological defects or presence of off-flavors), is strongly influenced by the operator (*e.g.*

level of experience, eyes sensibility, generically human variability), time-consuming [15], and hardly standardisable over time and across-panels.

This study aimed at developing and validating an accurate quantification method targeted to robust quality markers of raw hazelnuts to be used by the industry as an objective support tool for quality control (QC) purposes. Volatile makers within the detectable volatilome were selected based on their information potential: aroma blueprint odorants [11], [16], spoilage indicators [5], botanical/geographical tracers, secondary products of lipid oxidation [10], and storage time/conditions indicators [6].

To achieve suitable separation efficiency, resolution, sensitivity, and quantification accuracy, an analytical platform based on comprehensive two-dimensional gas chromatography with parallel detection by mass spectrometry and flame ionization detector (GC×GC-MS/FID) was designed. Moreover, intending the methodology suitable for industrial laboratories, fully automated sample preparation was by exhaustive extraction by multiple headspace solid-phase microextraction (MHS-SPME) to achieve accurate quantitative results [17]–[19] due to the solid nature of hazelnut kernels.

The method has the potential of the sensomics-based expert system (SEBES) for efficient and quick screening of key-odorants in food [20]. As an Artificial Intelligence (AI) smelling machine, the aroma blueprint of food can be monitored objectively. Moreover, by extending method capabilities to a larger set of quality markers, it might support AI decision-making tools once reference values are defined for each marker/pattern of markers [21].

To prove the method's suitability, a large set of hazelnut samples from industrial batches ($n=72$) were selected. Samples were from different botanical/geographical origins ($n=4$), undergone different post-harvesting (optimal or high moisture), and were stored in different conditions (5°C with air or under vacuum) for twelve months (time 0-4-6-9-12 months).

9.2 Materials and methods

9.2.1 Reference standards and solvents

Pure standards of n -alkanes (from $n\text{-C}_7$ to $n\text{-C}_{30}$) used for Linear Retention Indices (I^T) calibration; of α/β -thujone used as Internal Standards (ISs); of hexanal, heptanal, octanal, nonanal, decanal, 4-heptanone, 2-pentanol, 2-heptanol, γ -hexalactone used for external calibration; and solvents (cyclohexane and dibutyl phthalate – 99% of purity) were all from Merck (Milan, Italy).

9.2.2 Reference solutions and calibration standards

Standard stock solutions of reference analytes were prepared at a concentration of 10 g/L in cyclohexane and stored at -18°C for one week.

The n -alkanes solution for I^T calibration was prepared by mixing suitable amounts of standard stock solutions of pure standards to reach the concentration of 0.050 g/L using cyclohexane as solvent.

ISs working solution for standard-in fiber pre-loading was prepared at 0.100 g/L in dibutyl phthalate and stored at -18°C in sealed vials. ISs were used for validation purposes (method

precision and repeatability) and to normalize the analytes' absolute responses (i.e., % normalized response).

The reference working solution for analytes external calibration was prepared by mixing suitable amounts of standard stock solutions using dibutyl phthalate as solvent. Calibration curves were set to cover analyte absolute amounts within the range 1-2.5-5-10-25-50-100-250-500 ng. External standard calibration by MHS-SPME was conducted on FID while the MS trace was used to confirm analytes' identity and exclude co-elutions. Sampling conditions were those applied to hazelnut samples (50°C – 50 min sampling) with 4 consecutive extraction steps, achieving an almost exhaustive extraction for calibrated analytes [22].

9.2.3 Hazelnut samples

Raw hazelnuts from 2020 harvest year, were provided by Soremartec Italia Srl (Alba, Cuneo, Italy). Samples collected at each time point (see below) were stored at -18°C away from UV exposure until analysis.

The functional variables characterizing the sample set included four different cultivars/geographical origins (*Tonda Gentile Trilobata* – TGT, *Tonda Gentile Romana* – TGR, Akçakoca mix – AKC, Giresun mix – GIR) informing about the phenotype expression in the volatilome. Post-harvest conditions were tested on *ad hoc* samples from AKC and GIR, where hazelnuts were-dried on multiple layers with humidity kept between 15-20% in order to artificially simulate a rainy season (*Bad Post-Harvest* drying – GIRbPH and AKCbPH) and compared to an optimal in-field sun-drying on a single layer kept at humidity levels under 6% (*Good Post-Harvest* drying – GIRgPH and AKCgPH) [6].

Hazelnuts were then stored until 12 months after their harvest, and samples were collected and analyzed at 0,4,6,9, and 12 months. Storage, impacting sensory quality and off-flavor development, was at 5°C and 65% of equilibrium relative humidity with a standard atmosphere (SA) or under vacuum (UV).

A summary of samples characteristics and acronyms is provided in **Supplementary Table 1 – ST1**.

9.2.4 Headspace solid-phase microextraction: sample preparation, devices and conditions

Multiple HS-SPME was conducted by a multipurpose sampler, model MPS-2 (Gerstel, Mülheim a/d Ruhr, Germany), with a 2 cm Divinylbenzene/Carboxen/Polydimethylsiloxane (DVB/CAR/PDMS) *df* 50/30µm fiber from Merck (Milan, Italy), conditioned beforehand as recommended by the manufacturer. The standard-in-fiber procedure on 5 µL of IS placed in a 20 mL glass sealed vial was applied to pre-load the IS on the SPME; it was conducted at 50°C for 5 min.

MHS-SPME was performed on 100 ± 0.2 mg of finely ground hazelnuts (i.e., *granella*) prepared with a lab-scale grinder at Soremartec. The average dimensions of the powder were 500 µm (±50).

Sampling was at 50°C for 50 min. MHS-SPME was applied with 4 consecutive extraction steps on a 15% of samples randomly distributed over the analytical batch to define reliable decay constants (β) for accurate quantification [18], [23]–[25].

9.2.5 GC×GC-MS/FID: instrument setup and conditions

Analyses were performed on an Agilent 7890B unit coupled with an Agilent 5977B fast quadrupole MS detector (Agilent Technologies, Little Falls, DE, USA) with a high-efficiency source (HES) working in electron ionization mode at 70 eV. The scan range was between 40 and 250 m/z achieving an acquisition frequency of 28 Hz. The MS source was set at 230°C, the quadrupole at 150°C, and the transfer line temperature was at 280°C. Parallel detection was achieved with an FID set at 300°C base temperature, the H₂ flow was 40 mL/min, the air flow was 450 mL/min, and the sampling frequency was 200 Hz.

The split-splitless injector was set at 250°C, operating in pulsed-split mode at 250 kPa until 2.5 min with a 1:5 split ratio. The carrier gas was helium with a differential flow set at 0.4 mL/min for the first dimension (¹D) column and 10 mL/min in the second dimension (²D).

Modulation was by differential-flow modulator via the reverse-inject dynamics, it was realized on the capillary flow technology™ CFT (G4573A, Agilent Technologies). The modulation period (P_M) was 2.5s and pulse time was 0.25 s.

The column configuration was the following: ¹D DB-HeavyWax™ (100% polyethylene glycol - PEG; 20 m × 0.18 mm d_c × 0.18 μ m d_f) coupled with a ²D DB17 ((50%-phenyl)-methylpolysiloxane; 1.8 m × 0.18 mm d_c × 0.18 μ m d_f) both from Agilent Technologies. After the ²D column, the flow was split using a three-way unpurged capillary microfluidic splitter (G3181B, Agilent Technologies). The connections toward the MS and FID consisted of deactivated silica capillaries (Agilent Technologies) 0.5 m × 0.1 mm d_c and 1.1 m × 0.18 mm d_c respectively, resulting in a split ratio of 70:30 FID/MS. The bleeding capillary (5.81 m × 0.1 mm d_c) consisted of deactivated silica, dimensions were calculated using a validated calculator [22], [26].

The oven temperature program was the following: 40°C (2 min) to 130°C (0 min) @ 4°C/min, and to 260°C (10 min) @ 8°C/min.

9.2.6 Data acquisition and processing

Data were acquired using MassHunter WorkStation (Agilent Technologies) and processed with GC Image GC×GC Software ver. 2021r2 (GC Image, LLC, Lincoln, NE, USA). Data mining was performed using Matlab R2021a (The MathWorks, Inc., Natick, Massachusetts, United States) with the following packages: PCA toolbox (v1.5)[27] and Classification toolbox (v6.0)[28], and XLSTAT 2014 by Addinsoft (New York, USA).

9.3 Results and discussion

The analytical workflow employed in this study, based on the combined untargeted and targeted (UT) fingerprinting of the detectable volatilome of hazelnuts, has a great informative potential which can be exploited for answering many questions related to the samples properties. For instance, by UT fingerprinting and machine learning, informative patterns of UT features and their correlations to quality can be revealed and then used for quality

assessment and benchmarking. At the same time, thanks to the accurate quantification of an extended range of volatile components, robust data is collected for their adoption in long-term and/or inter-laboratory studies. The combination of both, the informative potential and the accurate quantification would be the basis for AI decision-making strategies [21].

Next sections will focus at first on the results related to the informative fingerprinting potential and its relation to hazelnut quality, and then on the robust and transferable outcomes of quantitative fingerprinting with an insight on the augmented smelling capability.

9.3.1 Raw hazelnut volatile fingerprint

Hazelnut volatilome was explored through the combined UT fingerprinting procedure [29]–[31] with previously validated processing parameters for the UT template construction [32], *i.e.*, S/N threshold of 50 datapoints (dp), distance threshold of 10 dp in both FID and MS channels, and additional direct match factor (DMF) threshold of 700 according to the NIST similarity algorithm for the MS channel only [33]. The resulting feature template was characterized by 44 reliable peaks and 164 peak regions for the MS channel and by 271 reliable peaks and 439 peak regions for the FID channel, of which 117 were putatively identified with the aid of I^T (± 15 tolerance) and spectrum similarity (> 900 DMF) for the MS channel. The resulting discrepancy in the number of reliable peaks between the two channels is attributed to both, the higher specificity of MS vs. FID that provides positive matches just for candidate peaks with a spectral similarity above the 700 DMF threshold, and the split ratio of 70:30 FID/MS which impacts on the absolute channel sensitivity.

The contour plot of the detected volatiles from a cumulative image resulting from all fresh (To) hazelnut samples is shown in **Figure 9.1**, while the list of targeted peaks is provided in **Supplementary Table 2 – ST2** together with their retention times and experimental and tabulated $^1D I^T$. The complexity of the matrix is shown in an exemplary heatmap in **Supplementary Figure 1 – SF1** where all UT features are represented in a color scale (from blue to red) according to their relative abundance (*i.e.*, normalized responses vs. internal standard). Hierarchical clustering of variables was by Pearson similarity.

The detectable volatilome was explored using chemometrics to verify the known correlations between biological properties/functional variables and chemical patterns, and possibly highlight new potential markers to be used for screenings at industrial level. Both unsupervised (Principal Component Analysis PCA) and supervised (Partial Least Squares Discriminant Analysis PLS-DA) algorithms were adopted.

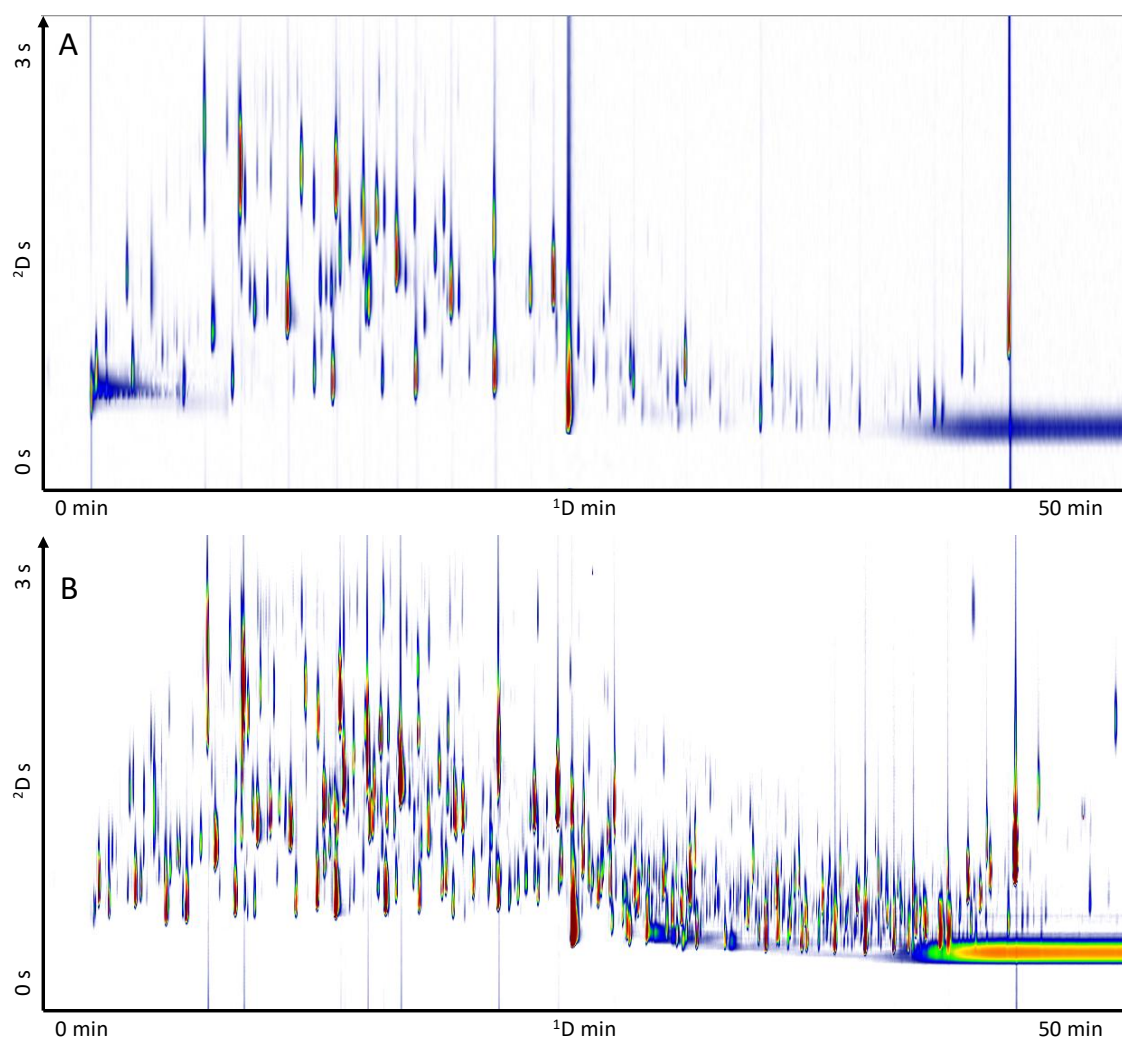


Figure 9.1: Contour plots corresponding to the detectable volatilome of fresh hazelnuts (0 months of storage) analysed by a polar $\times$ semi-polar column combination (i.e., HeavyWax $\times$ OV17) and resulting from the parallel detection by MS (1A) and FID (1B).

Figure 9.2A shows the PCA scores on the first two principal components (PC₁ and PC₂ for a total explained variance of 32%) obtained computing FID normalized responses of 439 UT peak regions collected from all samples (including 3 replicates each, $n=203$); samples are colorized according to storage time. Results confirm previous findings of large variability of the volatile fraction of raw hazelnuts as a function of botanical/geographical origin and post-harvest practices [6]. Fresh raw hazelnuts have a distinctive volatile signature which is mainly driven by the genetic and pedoclimatic characteristics of the harvest region (phenotype). However, along with storage, these distinctive traits are fuzzed by the impact of autooxidation which produces a large amount of volatile secondary products (saturated and mono-unsaturated linear aldehydes above all). Note that raw hazelnuts have a 50-60% of the gross composition represented by fat [34], [35].

Besides natural clustering shown by the PCA, supervised algorithms were used to understand the differential impact of functional variables left hidden by this preliminary approach. The classification of samples according to the storage months is improved with the supervised algorithm, as shown in **Figure 9.2B**. Scores plot for geographical origin macro-

regions (*i.e.*, Italy vs. Turkey) and storage conditions' (SA vs. UV) models are displayed in **Figure 9.2C** and **9.2D**, respectively. Classifications were validated according to the Montecarlo cross-validation algorithm which uses 20% of the total number of samples as the testing dataset and reiterates the validation procedure 1,000 times with a random choice of test samples [36]. The confusion matrix and ROC curves are shown in **Supplementary Figure 2 – SF2**. Moreover, variable importance in the projection (VIP) scores were examined to confirm the role of markers associated with the investigated variables and eventually highlight new volatiles with a potential marker role. Forty-eight targeted analytes were characterized by VIPs greater than 1 on LV1 and LV3 in the storage months' model, while 40 of them were discriminant in the storage conditions model. The complete list is reported in **Supplementary Table 3 – ST3** with an arrow indicating whether the analyte is positively or negatively contributing to the probability for an unknown sample to be classified as a member of a particular class. Note that the PLS-DA multi-class models were based on UT features % responses (*i.e.*, 2D peak response normalized over the IS and then over the total chromatographic response – MSTUS method). Therefore, an analyte could have positive/negative probability coefficients as a function of the specific class.

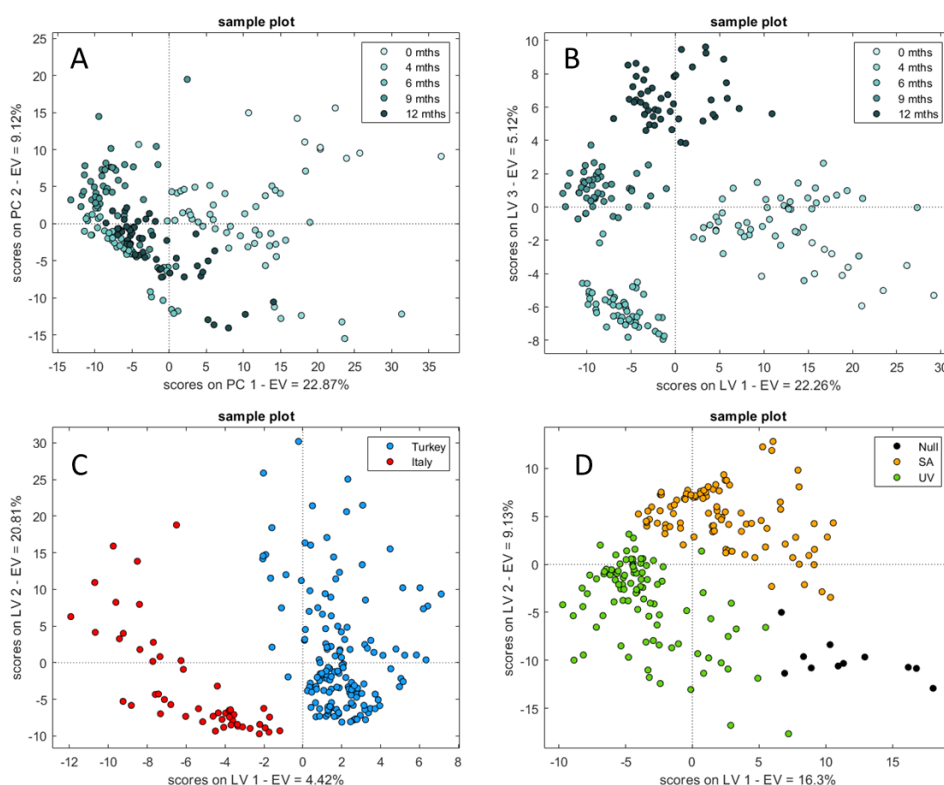


Figure 9.2: (2A) PCA score plot of untargeted and targeted (UT) reliable peaks (203 samples x 439 variables) based on normalized responses after Z-score normalization, displaying the storage months (*i.e.*, 0 months , 4 months, 6 months, 9 months, 12 months with lighter to darker shades of green). PLS-DA score plot of untargeted and targeted (UT) reliable peaks (203 samples x 439 variables) based on normalized responses after Z-score normalization, displaying the storage months (2B), the geographical area (*i.e.*, Turkey – blue, Italy – red) (2C), and the storage conditions (*i.e.*, Null – black, Standard Atmosphere (SA) – orange, Under Vacuum (UV) – green) (2D).

The combination of the information collected from classification models and variables interaction showed that short-chain saturated aldehydes registered higher normalized responses in samples undergone sub-optimal storage conditions (SA) with a positive correlation with storage months. Being secondary products of lipid autoxidation, their role as rancidity markers is validated by further supporting their use in the AI decision-making tool [37]. Moreover, at longer shelf-life times (9-12 months), the extent of oxidative processes converts carbonyl species into carboxylic acids, where an increase is registered for both octanoic and nonanoic homologs. Both chemical classes (carbonyls and carboxylic acids) are negatively impacting the storage classification model for the UV class. Other discriminant chemical species are primary alcohols (*e.g.*, 1-hexanol), which are also associated with SA storage, and are the final degradation products of linoleic acid [9], [38]–[40] [10], [41]–[50]. These findings are aligned with previous studies on storage stability by Kinderlerer *et al.* [10], on Turkish cultivars/blend quality by Alasalvar *et al.* [51], [52], by Cialiè Rosso *et al.* [6] and Cordero and co-workers [14], [53] when focusing on high-quality hazelnuts for confectionery.

Monoterpenes did not show significant evolution along shelf-life, although they have a weak influence on the SA storage model, most probably due to a differential impact of storage ambient to the seed metabolism.

9.3.2 Multiple markers quantitative analysis: accuracy and precision

Quantification is one of the most difficult tasks when HS sampling is applied to a solid matrix/sample. The physical form of the matrix prevents the adoption, at least for accurate quantification purposes, of any equilibrium-based approach since the analyte(s)' release from the condensed phase to the HS is governed by a combination of sorption/desorption and adsorption mechanisms [25], [54], the latter affected by displacement issues. Moreover, HS linearity, *i.e.*, the absence of the HS saturation, is compulsory especially if HS sampling is conducted in non-equilibrium conditions. To achieve a suitable accuracy level, external calibration should also take into account the matrix effect, as done by standard addition procedure with or without the use of stable isotope-labelled standards [18]. Note that standard addition provides accurate results in absence of HS saturation or for equilibrium sampling [17], [18], [55], [56].

In the case of non-equilibrium samplings, MHE approach is the elective route [57], [58]. It intrinsically overcomes the matrix effect by approaching/modeling exhaustive extraction of analytes from the sample HS. MHE was modelled by Kolb and Ettre [57] and successfully applied to HS-SPME sampling [18], [24], [25], [54], [59], [60] as MHS-SPME. Principles of operation and fundamental equations are out of the scope of this manuscript, information is provided in reference studies [18], [24], [25], [54], [59], [60].

In the present study, MHS-SPME was applied by external calibration on a selection of marker volatiles (hexanal, heptanal, octanal, nonanal, decanal, 4-heptanone, 2-pentanol, 2-heptanol, γ -hexalactone) selected for their informative role (lipid autoxidation secondary products, key-aromas, spoilage indicators) and also covering a suitable volatility/polarity range. External calibration was used in combination with FID-predicted relative response factors (RRFs) based on combustion enthalpies and molecular formula [61] to quantify additional volatiles not affected by HS saturation in the sampling conditions.

FID-predicted RRFs correlate the analyte response with its amount in the sample, relative to an internal standard (IS), taking into account the molecular weight and the structure (*e.g.*, number C-H-O and other heteroatoms, the number of aromatic rings) together with coefficients related to combustion enthalpy and IS identity. **Equation 9.1** allows the estimation of the RRF for analytes whose formula is known. By predicted RRF accurate quantification without calibration is possible as long as IS quantification is consistent and accurate.

Equation 9.1.
$$\text{RRF} = 10^3 * \left(\frac{MW_i}{MW_{IS}} \right) * (-61.3 + 88.8\eta_C + 18.7\eta_H - 41.3\eta_O + 6.4\eta_N + 64.0\eta_S - 20.2\eta_F - 23.5\eta_{Cl} - 10.2\eta_{Br} - 1.75\eta_I + 127\eta_{benz})^{-1}$$

In this study, the IS for RRFs application was hexanal, which amount was estimated by external calibration in all samples.

Table 9.1 lists all quantified analytes together with their retention times (t_R , ${}^2t_R$), experimental and tabulated ${}^1D$ I^T , odor qualities and odor threshold as listed in reference literature [62], calculated RRFs, β values (*i.e.*, constant representing the MHE exponential decay function [57]), and precision data on normalized responses over 3 analytical replicates from 12 different samples. For analytes included in the external calibration, linear function parameters (slope, intercept, and determination coefficient R^2) are also reported.

Due to the lack of certified reference materials, the quantification accuracy by MHS-SPME-GC×GC-MS/FID was verified using a previously validated and industrially certified method that quantifies linear saturated aldehydes (hexanal, heptanal, octanal, nonanal, decanal) by MHS-SPME-GC-MS [37]. The reference method is currently applied in QC industrial laboratories for rancidity assessment. To cover the full range of oxidation stages, therefore validating the current methodology within the entire dynamic range of application, results from nine samples were cross-compared and percent relative error (% RE) was calculated. Results indicate a quite good accuracy, RE was on average |11.6|% and ever below |20|% for all analytes which is considered acceptable by EU Commission indications for food safety analytical methods [63]. Results on accuracy are visualized as linear regression in **Figure 9.3A** where the amount (ng/g) of target aldehydes estimated by the certified method is considered as the independent variable (x) and results obtained by external calibration with the candidate method (*i.e.*, MHS-SPME-GC×GC-MS/FID) are treated as the dependent variable (y). The coefficient of determination indicates a quite good linear correlation between methods with just a few outliers in the higher range, reasonably due to the achievement of the limit of HS linearity.

Given the high complexity of the hazelnut volatilome and the existence of multiple markers informing about quality, the ability to expand the analytical method's quantitation capability to a wide range of analytes while maintaining acceptable accuracy was appealing. Therefore, FID-structure's specific response factors based on combustion enthalpies were estimated using hexanal (IS) for all analytes falling within HS linearity condition. **Table 9.1** lists RRFs calculated by **Equation 9.1** for analytes which showed an exponential decay function by MHE at the sampling conditions. β values above 0.95 were excluded, according to Kolb and Ettre [57] above that limit, HS saturation occurs.

The RRF quantification accuracy was tested vs. the results obtained from external calibrations. Five test analytes were used to cover the full spectrum of chemical classes in the hazelnut volatilome (*i.e.*, octanal – aldehydes, 4-heptanone – ketones, 2-pentanol and 2-heptanol – alcohols, γ -hexalactone – esters). The %RE was calculated on 48 samples randomly distributed within all analytical batches. Results referred for acceptable accuracy, with median % RE values of 18.7 for octanal (% RSD 15), 15.8 for 4-heptanone (% RSD 12), 20 for γ -hexalactone (% RSD 17), and 10.5 for 2-heptanol (% RSD 8). Results were tested for correlation and linear regression and shown in **Figure 9.3B**.

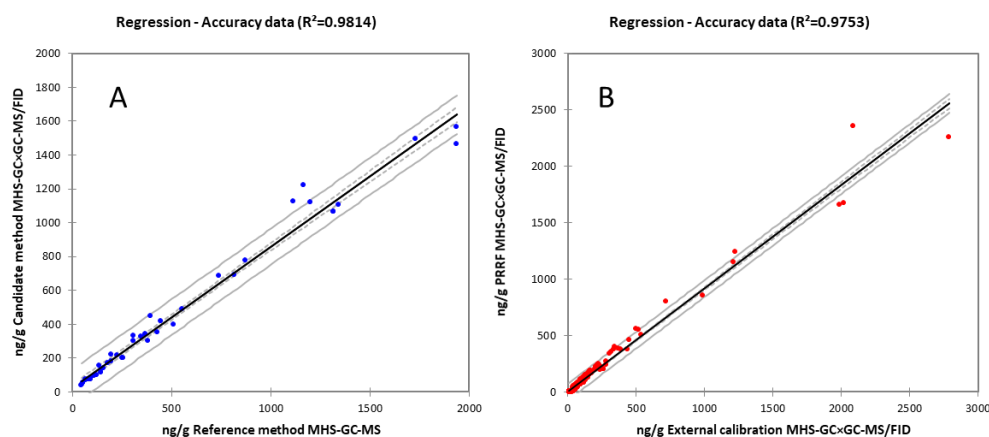


Figure 9.3: Regression analysis comparing quantification results between methods. (3A) Correlation between the candidate method by MHS-SPME-GC×GC-MS/FID and the reference method by MHS-GC-MS. Analytes compared are the linear saturated aldehydes from C6 to C10. (3B) Correlation between the estimated amounts by predicted RRFs and the external calibration. Analytes compared are hexanal, 4-heptanone, 2-pentanol, 2-heptanol, and γ -hexalactone.

To further confirm the inaccuracy of internal standardization if adopted for estimating analytes amount in HS-SPME non-equilibrium sampling [18], [19], [22], [55], results obtained by applying **Equation 9.2** were compared to those from external calibration.

Equation 9.2

$$C_x = \frac{A_x \cdot C_{IS}}{A_{IS}}$$

Where C_x is the amount or concentration of the analyte x in the analyzed sample, C_{IS} is the amount or concentration of the internal standard, A is the chromatographic area or response for the analyte (A_x) and the IS (A_{IS}). In practice, by internal standardization, the response factor of any target analyte is assumed to be equal to unity, and the matrix effect is neglected.

In the case of internal standardization, the median % RE values achieved 406 for octanal (% RSD 95), 646 for 4-heptanone (% RSD 108), 69 for γ -hexalactone (% RSD 36), and 633.5 for 2-heptanol (% RSD 108).

Once validated for its reliability, the quantitative results were used to develop an AI decision-making tool capable of augmented smelling with further potential to detect spoilage odors, rancidity while keeping its ability to correct classify samples according to origin and shelf-life.

Table 9.1: quantified analytes together with their retention times (t_{tr} , $^2t_{tr}$), experimental and tabulated ^1D I^T , odor qualities and odor threshold as listed in reference literature [46], calculated RRFs, β values (*i.e.*, constant representing the MHE exponential decay function [41]), and precision data on normalized responses over 3 analytical replicates from 12 different samples. For analytes included in the external calibration, linear function parameters (slope, intercept, and determination coefficient R^2) are also reported.

											External calibration			
Analyte	^t _{tr} min	^z _{tr} sec	Exp. I ^T	Tab I ^T	Odour quality	OT (ng/g)	RRFs	β ± SD		Uncertainty (% RSD)	Range (ng/g)	Slope	Intercept	R ²
Hexanal	11.5	0.93	1083	1080	green, grassy	276	0.967	0.662	0.026	11	1-50	0.030	0.038	0.998
											50-500	0.021	-0.079	0.986
Heptanal	15.3	0.98	1183	1186	oily, fatty	50	1.412	0.748	0.061	17	1-50	0.028	0.000	0.996
											50-500	0.024	0.129	0.983
Octanal	19.1	1.01	1288	1289	soapy	51	1.365	0.792	0.037	7	1-50	0.025	0.094	1.000
											50-500	0.021	0.037	0.983
Nonanal	22.8	1.03	1393	1390	fatty, waxy	610	1.329	0.892	0.022	11	1-50	0.031	0.227	0.932
											50-500	0.017	0.382	0.986
Butanoic acid	29.3	0.38	1621	1637	cheesy	135	2.357	0.928	0.007	10				
Decanal	26.1	0.91	1497	1496	orange peel like	75	1.301	0.837	0.064	14	1-50	0.035	0.073	1.000
											50-500	0.024	0.274	0.979
4-Heptanone	13.0	1.05	1124	1131	fruity, sweet	8.2	1.412	0.699	0.037	9	1-50	0.030	0.092	1.000
											50-500	0.035	-0.623	0.987
2-Pentanol	12.7	0.59	1115	1115	fruity	470	1.504	0.765	0.028	10	1-50	0.020	0.017	0.999
											50-500	0.033	-1.362	0.976
Nonanoic acid	38.0	0.38	2154	2144	fatty, waxy	2400	1.540	0.803	0.015	17				
α-pinene	9.3	1.67	1022	1022	terpenic	274	1.168	0.842	0.028	10				
3-Methyl-4-heptanone	14.0	1.17	1149	1178	nutty	0.86	1.365	0.875	0.045	9				
Octanoic acid	36.6	0.36	2048	2046	fatty waxy rancid	3000	1.608	0.781	0.021	16				
Pentanal	8.0	0.81	980	978	pungent, like bitter almond	150	1.574	0.684	0.057	13				
Hexanoic acid	33.4	0.38	1836	1839	pungent, musty	460	1.828	0.898	0.026	20				
1-Hexanol	21.1	0.60	1345	1340	fruity	400	1.425	0.768	0.046	10				
2,3-Butanediol	27.0	0.43	1530	1545	buttery, creamy	20000	2.184	0.918	0.035	10				
Decanoic acid	39.3	0.35	2258	2265	fatty	230000	1.487	0.861	0.035	9				
2-Heptanone	15.1	0.97	1180	1175	herbaceous	300	1.412	0.617	0.024	19				
4-Heptanol	18.7	0.68	1277	1281	alcoholic	410	1.371	0.685	0.053	11				
1-Pentanol	17.4	0.58	1242	1244	fruity	470	1.504	0.675	0.055	10				
γ-Butyrolactone	29.5	0.51	1630	1626	creamy, fatty, caramel	20000	2.569	0.922	0.013	9				
Heptanoic acid	35.1	0.38	1943	1946	rancid, fatty	100	1.700	0.797	0.066	19				
3-Methylbutanoic acid	30.2	0.38	1663	1666	sweaty	22	2.024	0.865	0.086	11				
Propanoic acid	27.1	0.41	1535	1534	Sour	720	3.047	0.865	0.055	11				

Table 9.1: continue

											External calibration			
Analyte	¹ t _R min	² t _R sec	Exp. I ^T	Tab I ^T	Odour quality	OT (ng/g)	RRFs	β ± SD		Uncertainty (% RSD)	Range (ng/g)	Slope	Intercept	R ²
Sabinene	12.8	1.52	1117	1120	woody, spicy, citrus	750	1.168	0.785	0.046	9				
Limonene	15.6	1.43	1193	1199	terpenic	14700	1.168	0.838	0.046	10				
2-Heptanol	19.9	0.65	1312	1318	fresh, lemon	263	1.371	0.725	0.039	10	1-50	0.031	0.031	0.995
											50-500	0.042	-1.218	0.990
Pentanoic acid	31.5	0.39	1729	1733	sweaty, fruity	400	2.024	0.906	0.035	12				
2-Hexanol	16.4	0.63	1214	1211	herbaceous	15	1.425	0.745	0.038	10				
p-Cymene	18.3	1.23	1268	1268	woody, terpenic	18000	1.190	0.828	0.025	10				
Benzaldehyde	26.9	0.64	1527	1530	sweet, oily	60	1.351	0.923	0.023	10				
δ-3-Carene	13.8	1.61	1144	1147	citrus, pine, herbal	770	1.050	0.829	0.069	10				
γ-Hexalactone	31.1	0.55	1704	1703	sweet, creamy	50	1.913	0.926	0.020	13	1-50	0.026	0.021	0.985
											50-500	0.033	-0.449	0.999
(E)-2-Heptenal	20.3	0.85	1322	1332	green, fatty, oily	14	1.457	0.800	0.032	13				
(E)-2-Hexenal					green, fatty, cheesy	0.42	1.535	0.812	0.023	9				
1-Octanol	27.5	0.55	1549	1546	waxy, green, citrus	900	1.331	0.888	0.009	13				
2-Decanone	25.9	0.76	1491	1495	orange, floral	8.3	1.301	0.778	0.127	18				
2-Nonanone	22.5	1.01	1387	1390	sweet green	100	1.329	0.825	0.035	8				
2-Octanone	18.8	1.12	1279	1281	herbal	510	1.365	0.742	0.011	13				
γ-Octalactone	34.8	0.51	1923	1916	sweet	120	1.657	0.928	0.004	21				
2-Decenal, (E)-	29.8	0.66	1642	1630	fatty	101	1.327	0.887	0.109	13				
(E)-2-Nonenal	27.2	0.77	1536	1543	fatty	4.1	1.359	0.933	0.010	12				
3-Methylbutanal	6.4	0.80	918	922	malty	5	1.574	0.848	0.012	12				

9.3.3 AI decision-making tools: augmented smelling and robust classification

Aligned with the SEBES concept, which acts as an AI smelling machine [20], the extended quantification results were used to evaluate the putative contribution of potent odorants to the overall aroma profile of samples. Analytes that reported a median for the Odor Activity Value (OAV) > 1 were computed and OAV values visualized in the log₁₀ scale as a spider diagrams. **Figure 9.4A** shows the blueprint of fresh hazelnuts belonging to the three cultivar/geographical regions investigated.

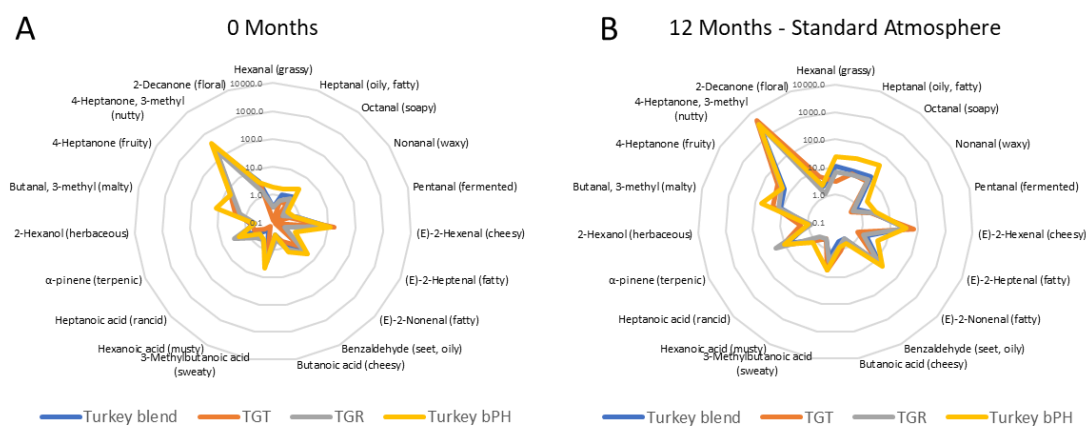


Figure 9.4: Spider-diagrams showing OAVs for odorants related to the hazelnut aroma blueprint, spoilage, and rancidity. OAVs are reported in log₁₀ scale. The shelf life changes are shown at 0 months, and 12 months with storage condition with standard atmosphere (SA); cultivar/geographical blend are the Tonda Gentile Trilobata (TGT), Tonda Gentile Romana (TGR), Turkey blend, and Turkey undergone bad post-harvest conditions (bPH).

Fresh samples (0 months) have similar profiles, with lower amounts of oxidative markers responsible for the *fatty* and *green* notes, and with the aroma blueprint dominated by 3-methyl-4-heptanone, which is responsible for the *nutty* quality. Along shelf life, samples have a different evolution, **Figure 9.4B** illustrates the blueprint at 12 months of storage in SA. TGT samples, considered as a gold standard for their superior sensory quality, tend to maintain the original characteristics for longer periods without major changes in the aroma-active compound patterns while also eliciting stronger resistance to rancidity (*i.e.*, lower amounts of saturated aldehydes). On the other hand, samples from the other cultivar/geographical origins have a different evolution with the Turkey selection showing a general increase in the amounts of short-chain fatty acids likely impacting *rancid* and *musty* odors. Moreover, TGT samples are the ones with generally higher values of key-aroma compounds. To note, the spider diagrams are represented in log₁₀ scale given the different order of magnitude of the OAV values. This means that minimal differences in the pictograms correspond to exponential variations; the complete list of average OAV values for all the samples at the different time-points is reported in **Supplementary Table 4 –ST4**.

Moreover, quantified data can be also processed with unsupervised and supervised chemometric tools. The latter are of particular interest given the possibility to generate models for quality benchmarking where diagnostic quantitative patterns of markers can be consistently used across many laboratories as AI decision-makers. PLS-DA models were created for two quality variables of interest for the industry: (a) discriminate the gold standard

from the other cultivars/origins (*i.e.*, TGT *vs.* rest); and (b) discriminate the critical in-filed post-harvesting versus the optimal drying (*i.e.*, bPH *vs.* rest).

The first model, which model score plot is illustrated in **Figure 9.5A**, was created to delineate the peculiar pattern of variables discriminating the gold standard TGT samples independently by the storage time. Odorous ketones, such as 4-heptanone, 2-decanone, and 3-methyl-4-heptanone are diagnostic markers for the TGT samples, while oxidation products as the carboxylic acids (C₅ to C₁₀), primary alcohols (C₆ and C₇), and both saturated and unsaturated aldehydes (C₅ to C₉) are the analytes discriminating all the other samples. These results are in keeping with previous observations on TGT cultivar and its distinctive volatile signature [14], [16], [53], [64], [65].

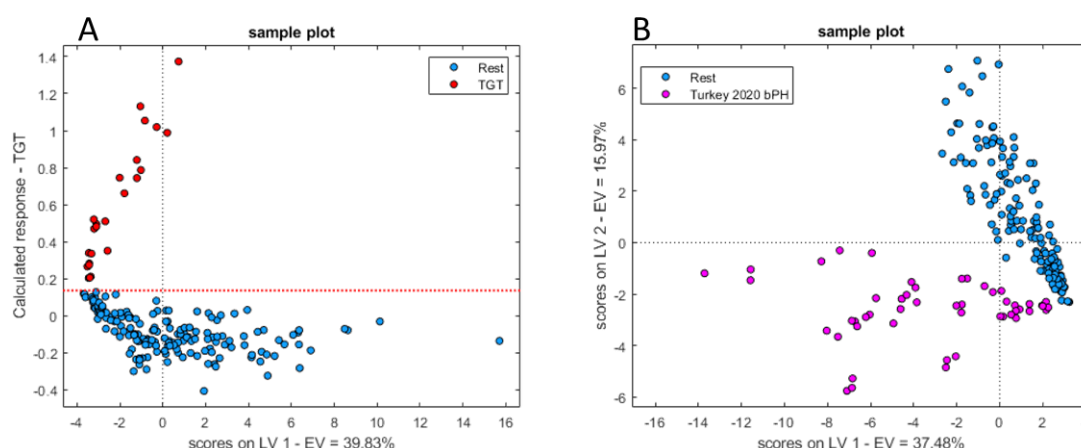


Figure 9.5: PLS-DA score plot on quantified analytes (203 samples x 43 variables), displaying the discrimination model results for TGT (Red) *vs.* all (Blue) (5A), and Turkey bPH (Purple) *vs.* all (Blue) (5B) .

The second classification model was conducted to assess critical threshold values for pattern variables correlated to bad post-harvest practices (**Figure 9.5B**). This is of particular importance for the industrial quality of kernels, as it was demonstrated to have a tremendous impact on their stability after long shelf-life periods [6]. Interestingly, eight variables were identified as the driving analytes for the classification algorithms: these are pentanoic and hexanoic acid (previously shown to be correlated with mouldy defect [5]), C₆ to C₈ saturated aldehydes, (*E*)-2-heptenal, 3-methylbutanal (correlated with higher oxidative levels probably due to the increased enzymatic activity correlated with increased moisture content), and two terpenoids, *i.e.*, p-cymene and sabinene (having a role in the plant defence against biotic stresses [66]).

9.4 Conclusions

The characterization of high-quality hazelnuts was done, for the first time, using a combination of chromatographic fingerprinting and multi-target quantification to achieve quantitative volatilomic as the basis of robust AI decision-making tools. Thanks to the improved separation power and chromatographic resolution compared to 1D-GC, fingerprinting is more specific and selective providing accurate quantification of several quality markers for product benchmarking and discrimination. Moreover, accurate

quantitative data can be transferred over time and across laboratories enabling the validation of AI decision-making tools over many harvest years.

The use of comprehensive two-dimensional gas chromatography and parallel detection by MS/FID implemented in a commercial platform with automated sample preparation and data interpretation, a step forward in the objective evaluation of industrial quality is made. Although non-mandatory by law, as for safety requirements, the industry has a great interest in defining chemical quality traits for all incoming materials, since these traits are those that guarantee consumers' preferences/choices while supporting the industry's competitive advantage in the market. By a single analysis, it is possible to objectively capture relevant quality traits named: aroma blueprint, storage quality and time, geographical area of harvest, and spoilage. All these traits can be turned into decision-making tools driving and orienting industry strategic investments and value chains. Since edible crops are living organisms, their complexity is preventing any strategy for quality assessment based on simple, single-marker measurement.

The analytical strategy proposed is also of great interest in a scenario where climate change, market fluctuations, and geopolitical stability could impact the availability of crops with suitable quality. Robust tools can guide strategic warehousing and industrial storage while guiding breeding strategies in new promising geographical areas.

Electronic Supplementary Material

Supplementary Table 1 – ST1: list of samples with characteristics and acronyms adopted in the text.

Country - Region	Harvest year	Technological processes	Storage type	Storage months	Acronym
Turkey - Akçakoca	2020	None	Null	0	AKC_To
Turkey - Giresun	2020	None	Null	0	GIR_To
Italy - Piedmont	2020	None	Null	0	TGT_To
Italy - Lazio	2020	None	Null	0	TGR_To
Turkey - Akçakoca	2020	Good Post-harvest	Null	0	AKCgPH_To
Turkey - Akçakoca	2020	Bad Post-harvest	Null	0	AKCbPH_To
Turkey - Giresun	2020	Good Post-harvest	Null	0	GIRgPH_To
Turkey - Giresun	2020	Bad Post-harvest	Null	0	GIRbPH_To
Turkey - Akçakoca	2020	None	5°C - UV	4	AKC_T4_Good
Turkey - Giresun	2020	None	5°C - UV	4	GIR_T4_Good
Italy - Piedmont	2020	None	5°C - UV	4	TGT_T4_Good
Italy - Lazio	2020	None	5°C - UV	4	TGR_T4_Good
Turkey - Akçakoca	2020	Good Post-harvest	5°C - UV	4	AKCgPH_T4_Good
Turkey - Akçakoca	2020	Bad Post-harvest	5°C - UV	4	AKCbPH_T4_Good
Turkey - Giresun	2020	Good Post-harvest	5°C - UV	4	GIRgPH_T4_Good
Turkey - Giresun	2020	Bad Post-harvest	5°C - UV	4	GIRbPH_T4_Good
Turkey - Akçakoca	2020	None	5°C - SA	4	AKC_T4_Bad
Turkey - Giresun	2020	None	5°C - SA	4	GIR_T4_Bad
Italy - Piedmont	2020	None	5°C - SA	4	TGT_T4_Bad
Italy - Lazio	2020	None	5°C - SA	4	TGR_T4_Bad
Turkey - Akçakoca	2020	Good Post-harvest	5°C - SA	4	AKCgPH_T4_Bad
Turkey - Akçakoca	2020	Bad Post-harvest	5°C - SA	4	AKCbPH_T4_Bad
Turkey - Giresun	2020	Good Post-harvest	5°C - SA	4	GIRgPH_T4_Bad
Turkey - Giresun	2020	Bad Post-harvest	5°C - SA	4	GIRbPH_T4_Bad
Turkey - Akçakoca	2020	None	5°C - UV	6	AKC_T6_Good
Turkey - Giresun	2020	None	5°C - UV	6	GIR_T6_Good
Italy - Piedmont	2020	None	5°C - UV	6	TGT_T6_Good
Italy - Lazio	2020	None	5°C - UV	6	TGR_T6_Good
Turkey - Akçakoca	2020	Good Post-harvest	5°C - UV	6	AKCgPH_T6_Good
Turkey - Akçakoca	2020	Bad Post-harvest	5°C - UV	6	AKCbPH_T6_Good
Turkey - Giresun	2020	Good Post-harvest	5°C - UV	6	GIRgPH_T6_Good
Turkey - Giresun	2020	Bad Post-harvest	5°C - UV	6	GIRbPH_T6_Good
Turkey - Akçakoca	2020	None	5°C - SA	6	AKC_T6_Bad
Turkey - Giresun	2020	None	5°C - SA	6	GIR_T6_Bad
Italy - Piedmont	2020	None	5°C - SA	6	TGT_T6_Bad
Italy - Lazio	2020	None	5°C - SA	6	TGR_T6_Bad
Turkey - Akçakoca	2020	Good Post-harvest	5°C - SA	6	AKCgPH_T6_Bad
Turkey - Akçakoca	2020	Bad Post-harvest	5°C - SA	6	AKCbPH_T6_Bad
Turkey - Giresun	2020	Good Post-harvest	5°C - SA	6	GIRgPH_T6_Bad
Turkey - Giresun	2020	Bad Post-harvest	5°C - SA	6	GIRbPH_T6_Bad
Turkey - Akçakoca	2020	None	5°C - UV	9	AKC_T9_Good
Turkey - Giresun	2020	None	5°C - UV	9	GIR_T9_Good
Italy - Piedmont	2020	None	5°C - UV	9	TGT_T9_Good
Italy - Lazio	2020	None	5°C - UV	9	TGR_T9_Good
Turkey - Akçakoca	2020	Good Post-harvest	5°C - UV	9	AKCgPH_T9_Good
Turkey - Akçakoca	2020	Bad Post-harvest	5°C - UV	9	AKCbPH_T9_Good
Turkey - Giresun	2020	Good Post-harvest	5°C - UV	9	GIRgPH_T9_Good
Turkey - Giresun	2020	Bad Post-harvest	5°C - UV	9	GIRbPH_T9_Good
Turkey - Akçakoca	2020	None	5°C - SA	9	AKC_T9_Bad
Turkey - Giresun	2020	None	5°C - SA	9	GIR_T9_Bad
Italy - Piedmont	2020	None	5°C - SA	9	TGT_T9_Bad
Italy - Lazio	2020	None	5°C - SA	9	TGR_T9_Bad
Turkey - Akçakoca	2020	Good Post-harvest	5°C - SA	9	AKCgPH_T9_Bad
Turkey - Akçakoca	2020	Bad Post-harvest	5°C - SA	9	AKCbPH_T9_Bad
Turkey - Giresun	2020	Good Post-harvest	5°C - SA	9	GIRgPH_T9_Bad
Turkey - Giresun	2020	Bad Post-harvest	5°C - SA	9	GIRbPH_T9_Bad
Turkey - Akçakoca	2020	None	5°C - UV	12	AKC_T12_Good
Turkey - Giresun	2020	None	5°C - UV	12	GIR_T12_Good

Supplementary Table 1 – ST1: continue

Country - Region	Harvest year	Technological processes	Storage type	Storage months	Acronym
Italy - Piedmont	2020	None	5°C - UV	12	TGT_T12_Good
Italy - Lazio	2020	None	5°C - UV	12	TGR_T12_Good
Turkey - Akçakoca	2020	Good Post-harvest	5°C - UV	12	AKCgPH_T12_Good
Turkey - Akçakoca	2020	Bad Post-harvest	5°C - UV	12	AKCbPH_T12_Good
Turkey - Giresun	2020	Good Post-harvest	5°C - UV	12	GIRgPH_T12_Good
Turkey - Giresun	2020	Bad Post-harvest	5°C - UV	12	GIRbPH_T12_Good
Turkey - Akçakoca	2020	None	5°C - SA	12	AKC_T12_Bad
Turkey - Giresun	2020	None	5°C - SA	12	GIR_T12_Bad
Italy - Piedmont	2020	None	5°C - SA	12	TGT_T12_Bad
Italy - Lazio	2020	None	5°C - SA	12	TGR_T12_Bad
Turkey - Akçakoca	2020	Good Post-harvest	5°C - SA	12	AKCgPH_T12_Bad
Turkey - Akçakoca	2020	Bad Post-harvest	5°C - SA	12	AKCbPH_T12_Bad
Turkey - Giresun	2020	Good Post-harvest	5°C - SA	12	GIRgPH_T12_Bad
Turkey - Giresun	2020	Bad Post-harvest	5°C - SA	12	GIRbPH_T12_Bad

Supplementary Table 2 – ST2: list of targeted peaks together with retention times and experimental and tabulated $'D I^T$.

Analyte	$'t_R \text{ min}$	$^2 t_R \text{ sec}$	Experimetal I^T	Literature I^T
Hexane	2.67	0.66	600	600
Octane	4.08	1.14	800	800
Allyl acetate	4.33	0.62	839	
1-Octene	4.75	1.14	855	837
Butanal	5.42	0.74	881	875
Ethyl cyclohexane	5.54	1.32	885	885
Nonane	5.92	1.57	900	900
2-Butanone	5.96	0.75	901	905
2-Methyl butanal	6.25	0.84	913	915
3-Methyl butanal	6.38	0.80	918	922
2,5-Dimethyl furan	7.33	0.86	955	952
n-Butyl ether	7.67	1.49	967	974
Pentanal	8.00	0.81	980	978
2,3-Butandione	8.21	0.64	988	982
Decane	8.71	1.91	1000	1000
4-Methyl-2-pentanone	8.92	0.89	1011	1010
3-Methyl-2-pentanone	9.25	0.95	1020	1016
α -Pinene	9.29	1.67	1022	1022
2-Butanol	9.38	0.58	1024	1022
α -Thujene	9.54	1.55	1028	1010
2-Ethyl-5-methyl furan	9.75	1.04	1034	1028
1-Propanol	9.83	0.55	1037	1037
Toluene	9.96	0.96	1040	1036
1-Decene	10.08	1.68	1043	1052
3-Hexanone	10.42	0.98	1053	1051
2,3,5-Trimethyl furan	10.54	1.00	1056	1051
Camphene	10.71	1.66	1061	1066
2,4-Dimethyl-3-pentanone	11.00	1.09	1069	
Butyl acetate	11.25	1.04	1076	1075
Hexanal	11.50	0.93	1083	1080
2-Methyl-1-propanol	11.75	0.56	1090	1086
Undecane	12.00	2.11	1100	1100
β -Pinene	12.17	1.69	1101	1110
2-Pentanol	12.71	0.59	1115	1115
Sabinene	12.75	1.52	1117	1120
4-Heptanone	13.04	1.05	1124	1131
3-Penten-2-one	13.04	0.76	1124	1131
Ethylbenzene	13.54	1.05	1138	1134
p-Xylene	13.54	1.05	1138	1134
1-Butanol	13.58	0.56	1139	1146

Supplementary Table 2 – ST2: continue

Analyte	¹ t _R min	² t _R sec	Experimenal I ^T	Literature I ^T
5-Methyl-2-hexanone	13.67	0.97	1141	1142
δ-Carene	13.79	1.61	1144	1147
5-Methyl undecane	13.96	1.97	1149	1157
3-Methyl-4-heptanone	13.96	1.17	1149	1178
β-Myrcene	14.42	1.32	1161	1159
α-Terpinene	15.04	1.37	1178	1178
2-Heptanone	15.13	0.97	1180	1175
Heptanal	15.25	0.98	1183	1186
2-Ethyl hexanal	15.33	1.14	1185	1197
3-Methyl-2-pentanol	15.54	0.65	1191	1202
Limonene	15.63	1.43	1193	1199
Dodecane	15.75	2.12	1200	1200
2-methyl-1-butanol	15.88	0.57	1201	1208
2,6-Dimethyl-4-heptanone	15.92	1.02	1201	1189
β-Phellandrene	16.00	1.47	1203	1205
2-Hexanol	16.38	0.63	1214	1211
Butanoic acid, butyl ester	16.54	1.18	1218	1221
2-Pentyl furan	16.96	1.12	1230	1230
6-Methyl-2-heptanone	17.21	1.02	1237	1228
γ-Terpinene	17.38	1.45	1241	1243
1-Pentanol	17.42	0.58	1242	1244
Styrene	17.83	0.99	1254	1251
p-Cymene	18.33	1.23	1268	1268
4-Heptanol	18.67	0.68	1277	1281
α-Terpinolene	18.75	1.46	1279	1280
1,2,4-Trimethyl benzene	18.75	1.12	1279	1275
2-Octanone	18.75	1.12	1279	1281
Acetoin	19.00	0.55	1286	1283
Octanal	19.08	1.01	1288	1289
Hexanenitrile	19.42	0.88	1297	1303
2-Ethyl-2-hexenal	19.42	1.01	1297	1310
2-Heptanol	19.92	0.65	1312	1318
(E)-2-heptenal	20.29	0.85	1322	1332
1-Hexanol	21.08	0.60	1345	1340
1,1'-oxybis hexane	21.63	1.68	1360	1367
2-Nonanone	22.54	1.01	1387	1390
Nonanal	22.75	1.03	1393	1390
Ethanol, 2-butoxy-	22.92	0.62	1397	1405
α-Thujone (IS)	23.79	1.06	1424	1429
2,3-Epoxyhexanol	23.96	0.74	1430	1428
β-Thujone (IS)	24.46	1.04	1445	1451
Acetic acid	24.50	0.41	1447	1449
cis-Sabinene hydrate	24.96	0.77	1461	1451
2-Ethyl-1-hexanol	25.67	0.61	1483	1484
2-Decanone	25.92	0.76	1491	1495
Decanal	26.13	0.91	1497	1496
2,6-dimethyl-4-heptanol	26.25	0.67	1502	1509
Formic acid	26.29	0.45	1503	1510
Benzaldehyde	26.92	0.64	1527	1530
2,3-Butanediol	27.00	0.43	1530	1545
Propanoic acid	27.13	0.41	1535	1534
(E)-2-nonenal	27.17	0.77	1536	1543
(E)-p-2-Menthen-1-ol	27.38	0.64	1544	1563
1-Octanol	27.50	0.55	1549	1546
2,3-Butanediol isomer	27.92	0.42	1565	1570
Terpinen-4-ol	28.67	0.65	1593	1600
Butanoic acid	29.29	0.38	1621	1637
Butyrolactone	29.50	0.51	1630	1626
Menthol	29.54	0.58	1632	1626
(E)-2-Decenal,	29.75	0.66	1642	1630
2-Octynoic acid, methyl ester	29.92	0.64	1650	
3-Methyl butanoic acid	30.21	0.38	1663	1666
γ-Hexalactone	31.08	0.55	1704	1703

Supplementary Table 2 – ST2: continue

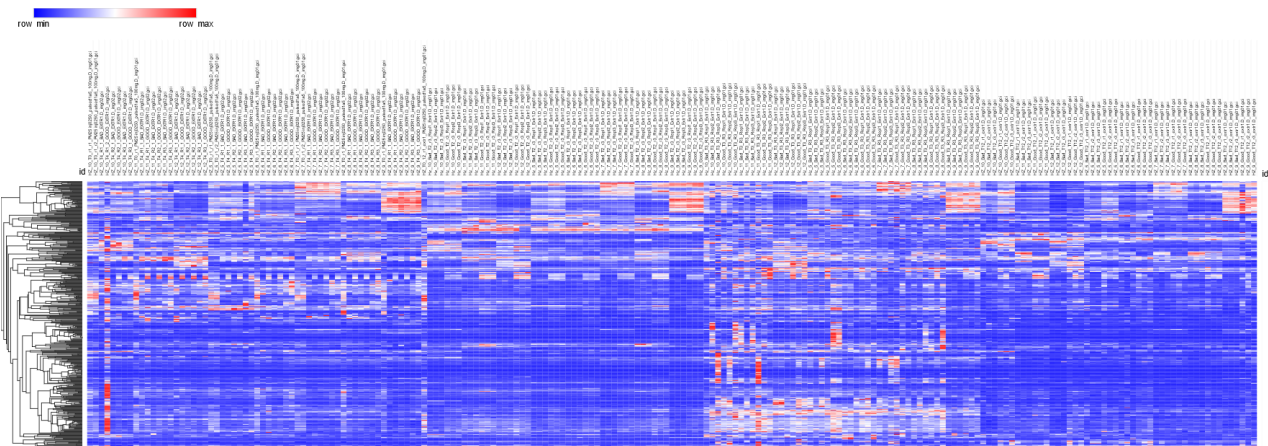
Analyte	¹ t _R min	² t _R sec	Experimenal I ^T	Literature I ^T
Pentanoic acid	31.54	0.39	1729	1733
2-(2-Butoxyethoxy)ethanol	32.58	0.50	1786	1786
Hexanoic acid	33.42	0.38	1836	1839
Butyl benzoate	33.92	0.59	1866	1856
2-Phenylethanol	34.54	0.45	1906	1915
Tetradecanal	34.75	0.68	1920	1919
γ-Octalactone	34.79	0.51	1923	1916
Heptanoic acid	35.08	0.38	1943	1946
γ-Nonalactone	36.21	0.42	2021	2028
Octanoic acid	36.58	0.36	2048	2046
Hexanedioic acid, bis(2-methylpropyl) ester	37.63	0.56	2125	2119
Nonanoic acid	38.00	0.38	2154	2144
Decanoic acid	39.29	0.35	2258	2265
Dibutyl phthalate (solvent for IS)	44.92	0.77	2727	2678

Supplementary Table 3 – ST3: list of targeted peaks characterized by VIPs greater than 1 on the classification model based on storage months and on storage type. Arrows indicate whether the analyte is positively (green) or negatively (red) contributing to the probability for an unknown sample to be classified as a member of a particular class. Note that a multi-class classification based on % responses was developed by PLS-DA therefore it could be that an analyte contributes with positive/negative probability coefficient in each class.

Downloadable at the link: <https://ars.els-cdn.com/content/image/1-s2.0-S0021967323002674-mm1.docx>

Supplementary Table 4 – ST4: Average OAVs for all quantified analytes in the samples under study. Quantification uncertainty (% RSD) is provided in Table 1.

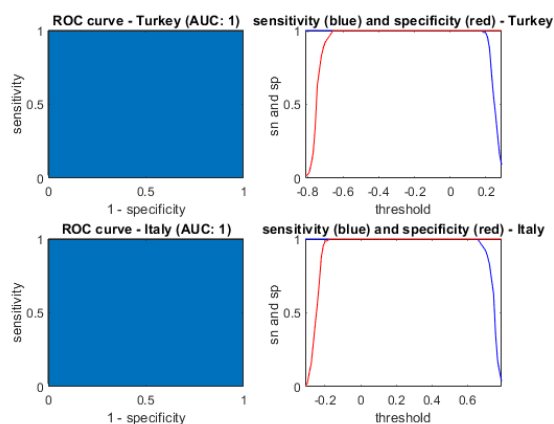
Downloadable at the link: <https://ars.els-cdn.com/content/image/1-s2.0-S0021967323002674-mm2.xlsx>



Supplementary Figure 1 – SF1: heat-map based on UT features normalized response from all analyzed samples including analytical replicates (n=203).

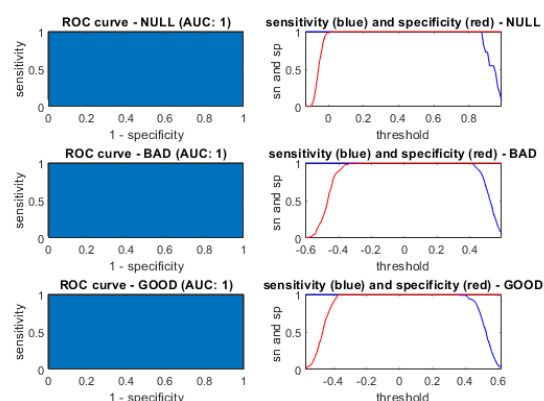
A

'real/predicted'	'Turkey'	'Italy'	'not assigned'
'Turkey'	'30276'	'535'	'0'
'Italy'	'94'	'10095'	'0'



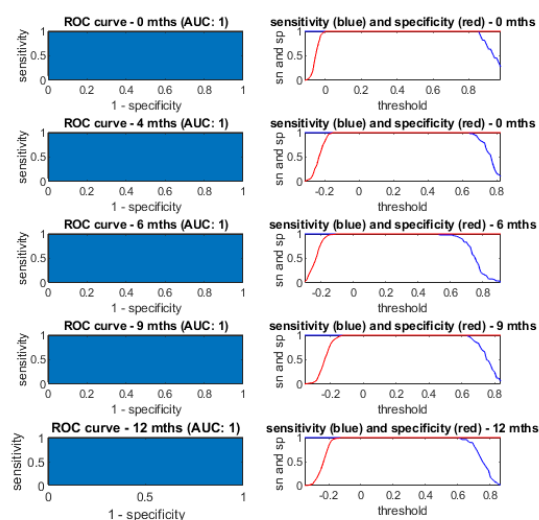
C

'real/predicted'	'NULL'	'BAD'	'GOOD'	'not assigned'
'NULL'	'1708'	'2'	'0'	'532'
'BAD'	'0'	'18875'	'396'	'183'
'GOOD'	'0'	'479'	'18695'	'130'



B

'real/predicted'	0 mths'	4 mths'	6 mths'	9 mths'	12 mths'	'not assigned'
0 mths'	'1970'	'0'	'0'	'0'	'0'	'217'
4 mths'	'0'	'9552'	'19'	'66'	'0'	'147'
6 mths'	'0'	'0'	'9626'	'0'	'0'	'0'
9 mths'	'0'	'0'	'43'	'9500'	'2'	'198'
12 mths'	'0'	'0'	'0'	'0'	'9660'	'0'



Supplementary Figure 2 – SF2: Confusion matrices and ROC curves deriving from PLS-DA modelling of untargeted and targeted (UT) reliable peaks (203 samples x 439 variables) on blob volumes after Z-score normalization using (a) country of origin, (b) storage months, and (c) storage condition as explanatory variables.

References

- [1] A. Dunkel *et al.*, “Nature’s chemical signatures in human olfaction: A foodborne perspective for future biotechnology,” *Angew. Chemie - Int. Ed.*, vol. 53, no. 28, pp. 7124–7143, 2014.
- [2] M. Granvogl and P. Schieberle, “The sensomics approach: A useful tool to unravel the genuine aroma blueprint of foods and aroma changes during food processing,” in *Comprehensive Analytical Chemistry*, 1st ed., vol. 96, Elsevier B.V., 2022, pp. 41–68.
- [3] A. Amann *et al.*, “The human volatilome: volatile organic compounds (VOCs) in exhaled breath, skin emanations, urine, feces and saliva,” *J. Breath Res.*, vol. 8, no. 3, p. 34001, Sep. 2014.
- [4] P. M. Dewick, “The Biochemistry of Plant Phenolics,” *Phytochemistry*, vol. 25, no. 8, pp. 2011–2012, Jul. 1986.
- [5] F. Stilo *et al.*, “An effective chromatographic fingerprinting workflow based on comprehensive two-dimensional gas chromatography – Mass spectrometry to establish volatiles patterns discriminative of spoiled hazelnuts (*Corylus avellana* L.),” *Food Chem.*, vol. 340, p. 128135, Mar. 2021.
- [6] M. Cialiè Rosso *et al.*, “Evolution of potent odorants within the volatile metabolome of high-quality hazelnuts (*Corylus avellana* L.): evaluation by comprehensive two-dimensional gas chromatography coupled with mass spectrometry,” *Anal. Bioanal. Chem.*, vol. 410, no. 15, pp. 3491–3506, 2018.
- [7] L. Nicolotti *et al.*, “Quantitative fingerprinting by headspace-Two-dimensional comprehensive gas chromatography-mass spectrometry of solid matrices: Some challenging aspects of the exhaustive assessment of food volatiles,” *Anal. Chim. Acta*, vol. 798, pp. 115–125, 2013.
- [8] J. Kiefl, *Differentiation of Hazelnut Cultivars (Corylus Avellana L.) by Metabolomics and Sensomics Approaches Using Comprehensive Two-dimensional Gas Chromatography Time-of-flight Mass Spectrometry (GCxGC/TOF-MS)*. Verlag Deutsche Forschungsanstalt für Lebensmittelchemie (DFA), 2013.
- [9] H. D. Belitz, W. Grosch, and P. Schieberle, *Food chemistry, 4th edition*. Springer, Berlin, Heidelberg, 2009.
- [10] J. L. Kinderlerer and S. Johnson, “Rancidity in hazelnuts due to volatile aliphatic aldehydes,” *J. Sci. Food Agric.*, vol. 58, no. 1, pp. 89–93, 1992.
- [11] S. Squara *et al.*, “*Corylus avellana* L. Aroma Blueprint: Potent Odorants Signatures in the Volatilome of High Quality Hazelnuts,” *Front. Plant Sci.*, vol. 13, no. March, pp. 1–25, Mar. 2022.
- [12] F. Ortega-Gavilán, S. Squara, C. Cordero, L. Cuadros-Rodríguez, and M. G. Bagur-González, “Application of chemometric tools combined with instrument-agnostic GC-fingerprinting for hazelnut quality assessment,” *J. Food Compos. Anal.*, vol. 115, no. September 2022, p. 104904, Jan. 2023.
- [13] V. Cristofori, S. Ferramondo, G. Bertazza, and C. Bignami, “Nut quality and sensory

evaluation of hazelnut cultivars,” *Acta Hortic.*, vol. 845, pp. 657–664, 2009.

- [14] C. Cordero *et al.*, “Profiling food volatiles by comprehensive two-dimensional gas chromatography coupled with mass spectrometry: Advanced fingerprinting approaches for comparative analysis of the volatile fraction of roasted hazelnuts (*Corylus avellana* L.) from different ori,” *J. Chromatogr. A*, vol. 1217, no. 37, pp. 5848–5858, 2010.
- [15] A. Giraudo, R. Calvini, G. Orlandi, A. Ulrici, F. Geobaldo, and F. Savorani, “Development of an automated method for the identification of defective hazelnuts based on RGB image analysis and colourgrams,” *Food Control*, vol. 94, no. July, pp. 233–240, 2018.
- [16] J. Kiefl, G. Pollner, and P. Schieberle, “Sensomics analysis of key hazelnut odorants (*Corylus avellana* L. ‘Tonda Gentile’) using comprehensive two-dimensional gas chromatography in combination with time-of-flight mass spectrometry (GC×GC-TOF-MS),” *J. Agric. Food Chem.*, vol. 61, no. 22, pp. 5226–5235, 2013.
- [17] C. Cordero *et al.*, “Odorants quantitation in high-quality cocoa by multiple headspace solid phase micro-extraction: Adoption of FID-predicted response factors to extend method capabilities and information potential,” *Anal. Chim. Acta*, vol. 1052, pp. 190–201, 2019.
- [18] B. Sgorbini, C. Cagliero, E. Liberto, P. Rubiolo, C. Bicchi, and C. Cordero, “Strategies for Accurate Quantitation of Volatiles from Foods and Plant-Origin Materials: A Challenging Task,” *J. Agric. Food Chem.*, vol. 67, no. 6, pp. 1619–1630, Feb. 2019.
- [19] H. H. Jeleń and M. N. Wiczorek, “Commentary: ‘Quantitative’ vs quantitative Headspace Solid-Phase Microextraction (HS-SPME) in food volatile and flavor compounds analysis,” *J. Food Compos. Anal.*, vol. 115, no. September 2022, p. 104955, Jan. 2023.
- [20] L. Nicolotti, V. Mall, and P. Schieberle, “Characterization of Key Aroma Compounds in a Commercial Rum and an Australian Red Wine by Means of a New Sensomics-Based Expert System (SEBES) - An Approach to Use Artificial Intelligence in Determining Food Odor Codes,” *J. Agric. Food Chem.*, vol. 67, no. 14, pp. 4011–4022, 2019.
- [21] A. Aleta *et al.*, “A Need for a Paradigm Shift in Healthy Nutrition Research,” *Front. Nutr.*, vol. 9, no. April, pp. 1–9, Apr. 2022.
- [22] F. STILO *et al.*, “Delineating the extra-virgin olive oil aroma blueprint by multiple headspace solid phase microextraction and differential-flow modulated comprehensive two-dimensional gas chromatography,” *J. Chromatogr. A*, vol. 1650, p. 462232, Aug. 2021.
- [23] B. Kolb and L. S. Ettre, *Static headspace-gas chromatography : theory and practice*. New York: Wiley-VCH, 2006.
- [24] B. Sgorbini, C. Bicchi, C. Cagliero, C. Cordero, E. Liberto, and P. Rubiolo, “Herbs and spices: Characterization and quantitation of biologically-active markers for routine quality control by multiple headspace solid-phase microextraction combined with separative or non-separative analysis,” *J. Chromatogr. A*, vol. 1376, pp. 9–17, Jan. 2015.
- [25] R. Costa, L. Tedone, S. De Grazia, P. Dugo, and L. Mondello, “Multiple headspace-solid-phase microextraction: An application to quantification of mushroom volatiles,” *Anal.*

Chim. Acta, vol. 770, pp. 1–6, 2013.

- [26] M. Giardina, J. D. McCurry, P. Cardinael, G. Semard-Jousset, C. Cordero, and C. Bicchi, “Development and validation of a pneumatic model for the reversed-flow differential flow modulator for comprehensive two-dimensional gas chromatography,” *J. Chromatogr. A*, vol. 1577, pp. 72–81, 2018.
- [27] D. Ballabio, “A MATLAB toolbox for Principal Component Analysis and unsupervised exploration of data structure,” *Chemom. Intell. Lab. Syst.*, vol. 149, pp. 1–9, 2015.
- [28] D. Ballabio and V. Consonni, “Classification tools in chemistry. Part 1: Linear models. PLS-DA,” *Analytical Methods*, vol. 5, no. 16, pp. 3790–3798, 2013.
- [29] C. Cordero *et al.*, “Comprehensive two-dimensional gas chromatography coupled with time of flight mass spectrometry featuring tandem ionization: Challenges and opportunities for accurate fingerprinting studies,” *J. Chromatogr. A*, vol. 1597, pp. 132–141, Jul. 2019.
- [30] M. Cialiè Rosso *et al.*, “Adding extra-dimensions to hazelnuts primary metabolome fingerprinting by comprehensive two-dimensional gas chromatography combined with time-of-flight mass spectrometry featuring tandem ionization: Insights on the aroma potential,” *J. Chromatogr. A*, vol. 1614, no. 460739, pp. 1–11, 2020.
- [31] F. Magagna *et al.*, “Combined untargeted and targeted fingerprinting with comprehensive two-dimensional chromatography for volatiles and ripening indicators in olive oil,” *Anal. Chim. Acta*, vol. 936, pp. 245–258, Sep. 2016.
- [32] S. Squara *et al.*, “Extending the breadth of saliva metabolome fingerprinting by smart template strategies and effective pattern realignment on comprehensive two-dimensional gas chromatographic data,” *Anal. Bioanal. Chem.*, no. 0123456789, Jan. 2023.
- [33] N. I. of S. and Technology, *NIST/EPA/NIH Mass Spectral Library with Search Program: (Data Version: NIST 08, Software Version 2.0f)*, vol. 2010, no. 1/22. Gaithersburg MD: National Institute of Standards and Technology (NIST), 2005.
- [34] C. Alasalvar and F. Shahidi, “Tree Nuts: Composition, Phytochemicals, and Health Effects,” 2008.
- [35] M. Cialiè Rosso, F. Stilo, S. Mascrez, C. Bicchi, G. Purcaro, and C. Cordero, “Shelf-Life Evolution of the Fatty Acid Fingerprint in High-Quality Hazelnuts (*Corylus avellana* L.) Harvested in Different Geographical Regions,” *Foods*, vol. 10, no. 3, p. 685, Mar. 2021.
- [36] Q. S. Xu and Y. Z. Liang, “Monte Carlo cross validation,” *Chemom. Intell. Lab. Syst.*, 2001.
- [37] S. Squara *et al.*, “Validation of a high-throughput method for the accurate quantification of secondary products of lipid oxidation in high-quality hazelnuts (*Corylus avellana* L.): A robust tool for quality assessment,” *J. Food Compos. Anal.*, vol. 114, p. 104766, Dec. 2022.
- [38] L. Q. Zepka *et al.*, *Study of the Volatile Compounds Useful for the Characterization of Frozen Anchoita (Engraulis anchoita) by SPME-GC-MS*. Elsevier Inc., 2014.
- [39] S. Kunyaboon, K. Thumanu, J. W. Park, C. Khongla, and J. Yongsawatdigul, “Evaluation

of lipid oxidation, volatile compounds and vibrational spectroscopy of silver carp (*Hypophthalmichthys molitrix*) during ice storage as related to the quality of its washed mince,” *Foods*, vol. 10, no. 3, 2021.

- [40] R. Hung, S. Lee, and J. W. Bennett, “The effects of low concentrations of the enantiomers of mushroom alcohol (1-octen-3-ol) on *Arabidopsis thaliana*,” *Mycology*, vol. 5, no. 2, pp. 73–80, 2014.
- [41] D. Ghirardello *et al.*, “Phenolic composition, antioxidant capacity and hexanal content of hazelnuts (*Corylus avellana* L.) as affected by different storage conditions,” *Postharvest Biol. Technol.*, vol. 112, pp. 95–104, 2016.
- [42] M. Pedrotti *et al.*, “Quality control of raw hazelnuts by rapid and non-invasive fingerprinting of volatile compound release,” *LWT*, vol. 143, no. February, p. 111089, 2021.
- [43] S. Belviso *et al.*, “Chemical, mechanical and sensory monitoring of hot air- and infrared-roasted hazelnuts (*Corylus avellana* L.) during nine months of storage,” *Food Chem.*, vol. 217, pp. 398–408, 2017.
- [44] M. Locatelli, J. D. Coisson, F. Travaglia, M. Bordiga, and M. Arlorio, “Impact of Roasting on Identification of Hazelnut (*Corylus avellana* L.) Origin: A Chemometric Approach,” *J. Agric. Food Chem.*, vol. 63, no. 32, pp. 7294–7303, 2015.
- [45] M. Özdemir, F. Açıktur, M. Yildiz, G. Biringen, T. Gürçan, and M. Löker, “Effect of roasting on some nutrients of hazelnuts (*Corylus Avellana* L.),” *Food Chem.*, vol. 73, no. 2, pp. 185–190, 2001.
- [46] M. Özdemir, “FACTORS INFLUENCING SHELF LIFE OF HAZELNUT,” *Basımı-Published Gıda Teknol.*, vol. 3, no. 3, pp. 66–71, 1998.
- [47] M. A. Koyuncu, A. Islam, and M. Küçük, “Fat and fatty acid composition of hazelnut kernels in vacuum packages during storage,” *Grasas y Aceites*, vol. 56, no. 4. pp. 263–266, 2005.
- [48] A. Turan, “Effect of drying methods on fatty acid profile and oil oxidation of hazelnut oil during storage,” *Eur. Food Res. Technol.*, vol. 244, no. 12, pp. 2181–2190, 2018.
- [49] S. M. Momchilova, S. P. Taneva, D. Zlatanov, G. A. Antova, M. J. Angelova-Romova, and E. Blagoeva, “Fatty acids, tocopherols and oxidative stability of hazelnuts during storage,” *Bulg. Chem. Commun.*, vol. 49, no. August, pp. 65–70, 2017.
- [50] A. Memoli *et al.*, “Effect of bug damage and mold contamination on fatty acids and sterols of hazelnut oil,” *Eur. Food Res. Technol.*, vol. 243, no. 4, pp. 651–658, Apr. 2017.
- [51] C. Alasalvar, E. Pelvan, B. Bahar, F. Korel, and H. Ölmez, “Flavour of natural and roasted Turkish hazelnut varieties (*Corylus avellana* L.) by descriptive sensory analysis, electronic nose and chemometrics,” *Int. J. Food Sci. Technol.*, vol. 47, no. 1, pp. 122–131, 2012.
- [52] C. Alasalvar, A. Z. Odabasi, N. Demir, M. Ö. Balaban, F. Shahidi, and K. R. Cadwallader, “Volatiles and flavor of five Turkish Hazelnut varieties as evaluated by descriptive sensory analysis, electronic nose, and dynamic headspace analysis/gas chromatography-mass spectrometry,” *J. Food Sci.*, 2004.

- [53] L. Nicolotti, C. Cordero, C. Bicchi, P. Rubiolo, B. Sgorbini, and E. Liberto, "Volatile profiling of high quality hazelnuts (*Corylus avellana* L.): Chemical indices of roasting," *Food Chem.*, vol. 138, no. 2–3, pp. 1723–1733, 2013.
- [54] C. Bicchi *et al.*, "Quantitative analysis of volatiles from solid matrices of vegetable origin by high concentration capacity headspace techniques: Determination of furan in roasted coffee," *J. Chromatogr. A*, vol. 1218, no. 6, pp. 753–762, 2011.
- [55] F. Stilo, C. Cordero, B. Sgorbini, C. Bicchi, and E. Liberto, "Highly Informative Fingerprinting of Extra-Virgin Olive Oil Volatiles: The Role of High Concentration-Capacity Sampling in Combination with Comprehensive Two-Dimensional Gas Chromatography," *Separations*, vol. 6, no. 3, p. 34, Jul. 2019.
- [56] F. Stilo, E. Liberto, S. E. Reichenbach, Q. Tao, C. Bicchi, and C. Cordero, "Exploring the Extra-Virgin Olive Oil Volatilome by Adding Extra Dimensions to Comprehensive Two-Dimensional Gas Chromatography and Time-of-Flight Mass Spectrometry Featuring Tandem Ionization: Validation of Ripening Markers in Headspace Linearity Condition," *J. AOAC Int.*, vol. 104, no. 2, pp. 274–287, 2021.
- [57] B. Kolb and L. S. Ettre, "Theory and practice of multiple headspace extraction," *Chromatographia*, vol. 32, no. 11–12, pp. 505–513, 1991.
- [58] J. Pawliszyn and C. F. Ross, "Headspace Analysis," in *Comprehensive Sampling and Sample Preparation*, vol. 2, Elsevier, 2012, pp. 27–50.
- [59] Ó. Ezquerro, G. Ortiz, B. Pons, and M. T. Tena, "Determination of benzene, toluene, ethylbenzene and xylenes in soils by multiple headspace solid-phase microextraction," *J. Chromatogr. A*, vol. 1035, no. 1, pp. 17–22, Apr. 2004.
- [60] W.-Q. Xie, Y.-X. Gong, and K.-X. Yu, "Enhancing the Sensitivity of Full Evaporation Technique Using Multiple Headspace Extraction Analysis," *Chromatographia*, vol. 80, no. 8, pp. 1263–1268, 2017.
- [61] J. Y. De Saint Laumer, E. Cicchetti, P. Merle, J. Egger, and A. Chaintreau, "Quantification in gas chromatography: Prediction of flame ionization detector response factors from combustion enthalpies and molecular structures," *Anal. Chem.*, vol. 82, no. 15, pp. 6457–6462, 2010.
- [62] F. Stilo, M. Cialliè Rosso, S. Squara, C. Bicchi, C. Cordero, and C. Cagliero, "*Corylus avellana* L. Natural Signature: Chiral Recognition of Selected Informative Components in the Volatilome of High-Quality Hazelnuts," *Front. Plant Sci.*, vol. 13, no. April, pp. 1–15, Apr. 2022.
- [63] E. Commission, "Commission Implementing Regulation (EU) 2021/808 of 22 March 2021 on the performance of analytical methods for residues of pharmacologically active substances used in food-producing animals and on the interpretation of results as well as on the methods to," *Off. J. Eur. Union*, vol. 180, no. 401, pp. 84–109, 2021.
- [64] J. Kiefl and P. Schieberle, "Evaluation of process parameters governing the aroma generation in three hazelnut cultivars (*Corylus avellana* L.) by correlating quantitative key odorant profiling with sensory evaluation," *J. Agric. Food Chem.*, vol. 61, no. 22, pp. 5236–5244, 2013.

- [65] M. Cialie Rosso *et al.*, “Combined Untargeted and Targeted Fingerprinting by Comprehensive Two-Dimensional Gas Chromatography to Track Compositional Changes on Hazelnut Primary Metabolome during Roasting,” *Appl. Sci.*, vol. 11, no. 2, p. 525, Jan. 2021.
- [66] B. Singh and R. A. Sharma, “Plant terpenes: defense responses, phylogenetic analysis, regulation and clinical applications,” *3 Biotech*, vol. 5, no. 2, pp. 129–151, Apr. 2015.

Chapter 10

Augmented Visualization by Computer Vision and Chromatographic Fingerprinting on Comprehensive Two-dimensional Gas Chromatographic Patterns: Unraveling Diagnostic Signatures in Food Volatilome

Computer Vision is an approach of Artificial Intelligence (AI) that conceptually enables “*computers and systems to derive useful information from digital images*” giving access to higher-level information and “*take actions or make recommendations based on that information*”. Comprehensive two-dimensional chromatography gives access to highly detailed, accurate, yet unstructured information on the sample's chemical composition, and makes it possible to exploit the AI concepts at the data processing level (*e.g.*, by Computer Vision) to rationalize raw data explorations. The goal is the understanding of the biological phenomena interrelated to a specific/diagnostic chemical signature. This study introduces a novel workflow for Computer Vision based on pattern recognition algorithms (*i.e.*, combined untargeted and targeted UT fingerprinting) which includes the generation of composite Class Images for representative samples' classes, their effective re-alignment and registration against a comprehensive feature template followed by Augmented Visualization by comparative visual analysis. As an illustrative application, a sample set originated from a Research Project on artisanal butter (from raw sweet cream to ripened butter) is explored, capturing the evolution of volatile components along the production chain and the impact of different microbial cultures on the finished product volatilome. The workflow has significant advantages compared to the classical one-step pairwise comparison process given the ability to realign and pairwise compare both targeted and untargeted chromatographic features belonging to Class Images resembling chemical patterns from many different samples with intrinsic biological variability.

This chapter was published as Original Research Paper in the Journal of Chromatography A:

A. Caratti, S. Squara, C. Bicchi, Q. Tao, D. Geschwender, S.E. Reichenbach, F. Ferrero, G. Borreani, C. Cordero, Augmented visualization by computer vision and chromatographic fingerprinting on comprehensive two-dimensional gas chromatographic patterns: Unraveling diagnostic signatures in food volatilome, J. Chromatogr. A. 1699 (2023) 464010. DOI: 10.1016/j.chroma.2023.464010.

10.1 Introduction

The process of chromatographic fingerprinting on patterns obtained by comprehensive two-dimensional chromatography (C2DC) aims at detecting, re-aligning, and comparing features extracted from 2D peaks across a series of analyzed samples [1]. The goal is to have access to higher-level information related to unique and/or distinctive composition and its relation to the phenomenon under study.

The representation of C2D chromatograms in digital images is obtained by *rasterization*, i.e., ordering detector data values from single modulation periods (or cycles) as a column of pixels (picture elements). Note that, when mass spectrometry (MS) with suitable analyzers is used, each pixel corresponds to a spectrum, thus information on feature identity is left available for further investigation. Rasterized pixel columns are then sequenced along the X-axis (from left to right) according to the first dimension (¹D) retention, while the second dimension (²D) data are presented in a right-handed Cartesian coordinate system, and the ²D retention time is plotted on the Y-axis (from bottom-to-top) [1]. Chromatographic images can be processed in various ways, adopting different feature types and related tools [1], [2].

The peak feature approaches, the most used tools implemented in all commercial data processing software, collect and sum data with associated metadata for each analyte peak. When multiple chromatograms are examined, peak features are then matched across multiple samples enabling detailed profiling of their components. If they correspond to targeted analytes (known constituents), their peak is explicitly matched across chromatograms by their target name. If they are not *a priori* targeted (i.e., untargeted analytes), the cross-comparative analysis is more challenging and unidentified peaks across many chromatograms must be matched by the software in a process usually referred to as peak matching or peak tracking [3]. An almost comprehensive review of the peak features approaches for C2DC has been recently proposed by Stilo *et al.* [1].

To compensate for peak matching inconsistencies occurring for large data sets, region and peak region feature concepts have been introduced. Within region feature approaches, the tile-based Fisher ratio analysis [4]–[6] enables effective cross-comparative analysis by adopting rectangular tiles to generate features over the chromatographic space and inspecting tile's detector response (including single *m/z* channels) for meaningful variations across samples (e.g., Fisher discriminant analysis). By peak region features, tiles are generated around 2D peaks footprint with some advantages when chromatographic misalignment occurs. Further insights on this feature type are provided below.

Independent of the feature type adopted, the process that re-aligns and compares 2D peak patterns across many chromatograms is referred to as chromatographic fingerprinting [1]. However, if it can provide visual evidence of compositional differences between samples (e.g., by comparative visualization [7]) it can also be considered as a Computer Vision (CV) approach. CV is defined as “*a field of artificial intelligence (AI) that enables computers and systems to derive meaningful information from digital images... and take actions or make recommendations based on that information. If AI enables computers to think, computer vision enables them to see, observe and understand*” [8].

Comparative visualization, a type of CV, was one of the earliest approaches applied to reveal compositional differences between pairs of 2D chromatograms. However, data point features which comprehensively cover all detected/untargeted analytes with the highest precision on 2D fingerprint, result in greater computational complexity since analytes are represented by multiple features (*i.e.*, multiple spectra per analyte), and/or might be imprecise due to retention-times inconsistencies which confound feature matching across chromatograms.

In this study, a novel workflow for effective comparative visualization of multiple chromatograms is proposed. In particular, CV is performed by combined untargeted and targeted (UT) fingerprinting [9], [10] with a fully automated procedure while Augmented Visualization (AV) [11] is made possible by 2D chromatograms alignment on reliable peaks and peak regions, information (*i.e.*, feature metadata) which is available at each stage.

As a challenging test bench, butter volatilome evolution along the production chain is studied. The complexity of the analyzed mixture relates to the large dynamic range of concentrations for characteristic chemical classes (*e.g.*, short-chain fatty acids, alcohols, aldehydes, esters and lactones, etc.) associated to a limited volatility range. A CV approach with molecular resolution is of great interest for food technology, it prompts information on compositional differences induced by a specific process, or technological conditions, supports decision-making strategies, and facilitates process optimization and product *identification* [12].

10.2 Materials and methods

10.2.1 Reference standards and solvents

Pure standards of *n*-alkanes (from *n*-C₇ to *n*-C₃₀) used for Linear Retention Indices (I^T) calibration, of α/β -thujone and 2-methyloctynoate used as Internal Standards (ISTDs), and solvents (cyclohexane and dibutyl phthalate – 99% of purity) were from Merck (Milan, Italy).

For key-aroma compounds and potent odorants investigated in the study, identity confirmation was by authentic standards available from Merck (Milan, Italy).

10.2.2 Reference solutions and calibration mixtures

Standard stock solutions of reference analytes were prepared at a concentration of 10 g/L in cyclohexane and stored at -18°C for one week.

The Reference Solution for analytes identity confirmation was prepared by mixing suitable amounts of standard stock solutions to reach the concentration of 0.100 g/L using cyclohexane as solvent.

The *n*-alkanes solution for I^T calibration was prepared by mixing suitable amounts of standard stock solutions of pure standards to reach the concentration of 0.050 g/L using cyclohexane as solvent.

ISTDs working solution for standard-in fiber pre-loading [13] was prepared at 0.100 g/L in dibutyl phthalate and stored at -18°C in sealed vials. ISs were used for validation purposes (method precision and repeatability) and to normalize the analytes' absolute responses (*i.e.*, % normalized response).

10.2.3 Hazelnut samples

Sweet cream from raw cow milk and butter samples were provided by Dr. Fabio Bruno and were part of an experimental production of artisanal butter by Beppino Occelli Agrinatura S.r.l (Farigliano, Cuneo, Italy). Samples consisted of raw sweet cream from cow milk from two different farms located in Piedmont, Italy (Farm #1 and Farm #2) which were then mixed in a ratio 1:1 before pasteurization (95°C/10 s) and inoculum with selected cultures (Standard - STD and TEST) for butter production. After 18 hours from the inoculum, samples were cooled to allow butterfat crystallization, which facilitates the churning and improves butter texture (process step generally referred to as *aging*). Once ready, samples were transferred to a continuous butter-maker via a plate heat exchanger. The butter obtained, after draining and washing, was then ripened until 40 days.

Quality control samples (QCs) were obtained from a commercial butter which was gently melted and homogenized before storage at -80°C until analyzed. The melting process, expected to have an impact on the volatile fraction, was performed to obtain a truly homogeneous QC for analytical checks. The butter matrix is known to have a certain inhomogeneity due to the nonuniform bacterial growth within the crystallized fat.

Details on the sample set under study are provided in **Table 10.1** together with the acronyms adopted in the text; a schematic diagram of available samples together with biological and technical replicates is shown in **Supplementary Figure 1**.

Table 10.1: List of samples together with processing details, acronyms adopted in the text, and corresponding class image generated for the Computer Vision (CV).

Production stage	Samples details	Analyzed samples		Class Images
		Farm #1	Farm #2	
Sweet cream - raw		SCR #1 (n=6)	SCR #2 (n=6)	SCR
		Mixed 1:1 ratio		
Pasteurized cream	Standard pasteurization 95°C/15s Inoculum "standard" for cream A or "test" for cream B - time 0	PC STD -T ₀ (n=6)	PC TEST -T ₀ (n=6)	PCT ₀
Inoculated cream	Inoculum "standard" for cream A or "test" for cream B - aging for 18 h	PC STD -T ₁₈ (n=6)	PC TEST -T ₁₈ (n=6)	PCT ₁₈
Butter	Butter at time 0	B STD-T ₀	B TEST-T ₀	BT ₀
	Butter after 8 days of ripening	B STD-T ₈	B TEST-T ₈	BT ₈
	Butter after 20 days of ripening	B STD-T ₂₀	B TEST-T ₂₀	BT ₂₀
	Butter after 40 days of ripening	B STD-T ₄₀	B TEST-T ₄₀	BT ₄₀

10.2.4 Headspace solid-phase microextraction: devices and conditions

Volatiles from cream and butter samples were extracted by HS-SPME with a divinylbenzene/carboxen/polydimethyl siloxane (DVB/CAR/PDMS) fiber (d_f 50/30 μ m; 2 cm length) from Merck Lifesciences - Supelco (Bellefonte, PA, USA) chosen on the basis of previous studies [14]. The SPME fiber was conditioned before use as recommended by the manufacturer.

The ISTDs for response normalization and quality control were preloaded onto the SPME device [15] by sampling 5.0 μ L of α/β -thujone and methyl 2-octynoate ISs solution (0.100 g/L)

placed in a 20 mL headspace vial. ISTDs pre-loading was performed by exposing the SPME device in the HS at a temperature of 50 °C for 5 min.

Sampling was carried out on 0.100 ± 0.005 g of sample, precisely weighed, in 20 mL headspace vials, kept at 50 °C for 50 minutes under constant agitation. The very low amount of sample was chosen to match the HS linearity conditions for most of the characteristic analytes of the dairy fat volatilome. After extraction, the SPME device was automatically transferred to the split/splitless injection port of the GC×GC system, kept at 250 °C, and submitted to thermal desorption for 5 min. Samples were analyzed in three replicates randomly distributed over a two-week time frame.

10.2.5 GC×GC-TOFMS: instrument set-up and conditions

Comprehensive two-dimensional GC analyses were carried out with an Agilent 7890B GC chromatograph (Agilent Technologies, Wilmington, DE, USA) coupled with a Markes BenchTOF Select™ mass spectrometer featuring Tandem Ionization™ (Markes International, Llantrisant, UK). The GC transfer line was set at 270 °C. The TOFMS was tuned for single ionization at 70 eV and the scan range was set between 35–350 m/z with a spectrum acquisition frequency of 100 Hz. The thermal modulator was a loop-type, two-stage KT 2004 (Zoex Corporation, Houston, TX) cooled with liquid nitrogen and controlled by Optimode, v2.0 (SRA Instruments, Cernusco sul Naviglio, Milan, Italy). The modulation period (P_M) was set at 2.5 s, while the hot-jet pulse duration was set at 250 ms. The cold-jet stream at the mass flow controller (MFC) was programmed to linearly reduce the total flow (*i.e.*, 20 L/min) from 40% to 8% along the analytical run.

The column set consisted of a ¹D HeavyWax™ column (100% polyethylene glycol – PEG; 30 m × 0.25 mm d_c × 0.25 μm d_f) coupled with a ²D DB17 column (50% phenyl-methylpolysiloxane; 1.0 m × 0.10 mm d_c × 0.10 μm d_f), both supplied by Agilent Technologies (Wilmington, DE, USA). A fused silica capillary loop (1.0 m × 0.1 mm d_c) was used in the modulator slit. SilTite™ μ-unions (Trajan Scientific and Medical, Melbourne, Australia) were used to connect the columns with the capillaries.

The GC split/splitless injector port was set at 250 °C and operated in pulsed-split mode (250 kPa overpressure applied to the injection port until 2 min) with a 1:20 split ratio. A special design liner for SPME thermal desorption (Merck Lifesciences) was used to improve the transfer of the analytes to the ¹D column and to limit band broadening in space. Helium was used as the carrier gas at a nominal flow of 1.3 mL/min. The oven temperature program was set as follows: from 50 °C (1 min) to 260 °C (5 min) at 4.0 °C min⁻¹, to 270 °C (10 min) at 10 °C min⁻¹.

The *n*-alkanes solution for I^T 's determination was analyzed under the following conditions: split/splitless injector in split mode, split ratio: 1:50, injector temperature: 250 °C, and injection volume: 1 μL.

10.2.6 Combined Untargeted and Targeted (UT) fingerprinting by smart templates

A template is a pattern of 2D peaks and/or graphic objects (*i.e.*, features) generated over a reference chromatogram or image (single or cumulative image), which is used to identify

similar patterns of 2D peaks in a set of analyzed chromatograms or images [10], [16]. This procedure is guided by specific matching functions that establish correspondences between features across many chromatograms; specificity is achieved by defining confidence thresholds on retention times and MS spectral similarity to compensate for variable peak detection, and through suitable transform functions that successfully compensate for retention time inconsistencies between runs [17], [18]. Once the correspondence between features is established, the template object metadata (*i.e.*, chemical name, retention times, mass spectra, informative ions and their relative ratios) are transferred to candidate peaks and/or graphic objects (*i.e.*, peak regions) in the analyzed chromatogram.

The feature template including untargeted reliable peaks and peak regions is obtained with a fully automated workflow in GC Image Investigator™ (GC Image™, GC Image LLC).

The workflow includes the following steps:

- Peak patterns are matched between all chromatogram pairs from a set of representative samples. Matching constraints were set for this study based on optimized parameters validated by Squara *et al.* [18], *i.e.*, template peaks inclusion criterion S/N threshold of 50 or higher; peak reference spectra from *peak spectrum*; MS constraint set at 700 for the direct match factor (DMF) and 700 for the reverse match factor (RMF) based on the NIST similarity algorithm [19].
- Reliable peaks across the set of analyzed chromatograms are then selected. The template of reliable peaks is used at the successive step for chromatogram re-alignment in the temporal domain. The criterion for inclusion was *relaxed* reliability [20] which selects 2D peaks that match at least 50% +1 of the chromatograms.
- Chromatograms of representative samples are aligned consistently with the average retention times of the reliable peaks and then summed/fused into a composite chromatogram.
- From the composite chromatogram, a comprehensive untargeted feature template is generated [20] including reliable peaks and peak regions. These last are defined over the footprint of all detected peaks in the composite chromatogram.

This process was applied to generate composite class images from samples with common characteristics, *e.g.* same processing step. In particular, class images were built for raw sweet cream (*SCR*); pasteurized sweet cream inoculated with selected cultures and analyzed at time 0 and after 18 hours (*PCT₀* and *PCT₁₈*); butter samples analyzed along ripening at 0-8-20-40 days (*BT₀*, *BT₈*, *BT₂₀*, *BT₄₀*). Acronyms and samples' characteristics are reported in **Table 10.1** while a schematic diagram of butter making process is illustrated in **Supplementary Figure 10.1**. The data processing workflow is illustrated in **Figure 10.1** and further details are reported in **Supplementary Figure 2**. In the *Results and Discussion* section insights on the novel workflow are provided.

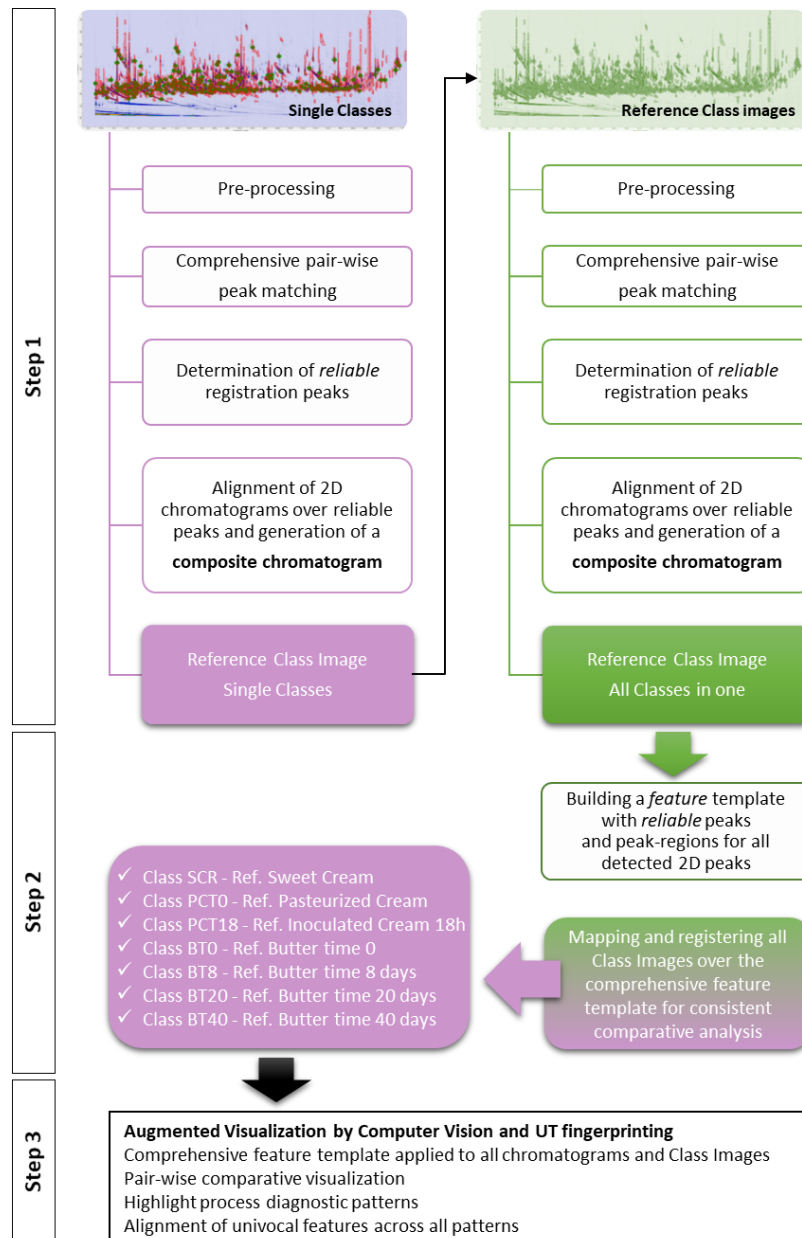


Figure 10.1: Multistep CV data processing workflow: step 1 includes all the data pre-processing operations needed to create composite class images for representative samples' classes. Subsequently, the class images are effectively re-alignment and registered against a comprehensive feature template. Subsequently, by comparative visual analysis, samples are pairwise compared to highlight process diagnostic patterns after the alignment of univocal features across all patterns.

Identification of targeted features was by matching EI-MS fragmentation patterns collected from the peak apex (NIST MS Search algorithm, version 2.0 [19]) with those of commercial and in-house databases. Identification thresholds were 900 for DMF and 950 for RMF. A further constraint was applied for reliable identification with ± 20 units of tolerance for experimental vs. tabulated linear retention indices (I^T) along the ^1D .

The list of targeted features is reported in **Table 10.2** together with chemical names, ^1D and ^2D retention times (^1t_R ; ^2t_R), experimental I^T , tabulated I^T , and odor quality as from reference literature.

Table 10.2: List of targeted features together with ^1D and ^2D retention times (^1t_R ; ^2t_R), experimental I^T and tabulated I^T , odor quality as from reference literature. Identification criteria: (a) by authentic reference compounds or (b) tentative identification combining $I^T \pm 10$ and spectral similarity direct match factor ≥ 900 .

Compound Name	Identification	^1t_R (min)	^2t_R (sec)	Experim. I^T	Tabulated I^T	Odor quality
Hexane	a	4.38	0.54	600	/	/
Heptane	a	4.78	0.7	700	/	/
Dimethyl sulfide	b	4.81	0.5	729	754	
Cyclohexane	a	4.9	0.72	811	766	/
Triethylamine	a	5.13	0.76	821	780	Fishy
Octane	a	5.48	1	835	/	/
Acetone	a	5.72	0.54	845	826	Solvent
1-Chlorobutane	b	6.07	0.72	860	842	/
4-Methyloctane	b	6.24	1.28	868	850	/
Butanal	a	6.59	0.64	883	867	Cocoa
Acetic acid ethenyl ester	b	6.72	0.65	887	890	
Ethyl acetate	a	6.83	0.66	893	884	Ethereal
Nonane	a	6.88	1.46	895	/	/
2-Butanone	b	7.05	0.56	907	907	
3-Methylbutanal	a	7.35	0.76	915	917	Aldehydic, fruity
Ethanol	a	7.58	0.5	925	926	Alcoholic
1-Nonene	b	7.82	1.42	935	931	/
2,2,4,6,6-Pentamethylheptane	b	8.17	1.9	950	954	/
Butyl ether	a	8.34	1.46	957	963	Ethereal
3-Methylnonane	b	8.34	1.82	957	966	/
Diacetyl	a	8.87	0.6	979	980	Buttery
Decane	a	9.16	2.02	992	/	/
Isobutyl acetate	a	9.74	0.92	1012	1013	Fruity
α -Pinene	a	9.92	1.66	1018	1023	Herbal, woody
Tetrachloroethylene	b	9.98	1.12	1020	1024	/
Chloroform	a	10.03	0.6	1021	1024	/
1-Propanol	a	10.33	0.52	1030	1031	Alcoholic
Ethyl butanoate	a	10.44	1.08	1034	1042	Fruity
1-Decene	b	10.44	1.88	1034	1044	/
Camphene	a	11.26	1.82	1060	1060	Woody, camphor
Isopentyl acetate	b	11.67	1.12	1080	1080	Fruity
Hexanal	a	11.9	0.98	1083	1083	Grassy, apple
Isobutyl alcohol	a	12.02	0.54	1091	1085	Ethereal
Undecane	a	12.25	2.44	1091	/	/
β -Pinene	a	12.66	1.9	1103	1109	Herbal
2-Pentanol	a	13.01	0.62	1112	1113	Mild, musty
Sabinene	a	13.18	1.68	1117	1121	Woody
1-Methoxy-2-propanol	b	13.36	0.62	1122	1126	/
(E)-2-Pentenal	a	13.53	0.84	1126	1127	Green, apple
4-Methyl-3-penten-2-one	b	13.65	0.92	1129	1131	Pungent, potato
p-Xylene	a	13.71	1.2	1131	1136	/
1-Butanol	a	13.88	0.58	1136	1138	Fermented, fruity
m-Xylene	a	13.94	1.14	1137	1141	/
δ -3-Carene	a	14.18	1.86	1143	1147	Citrus
3-Heptanone	a	14.41	1.18	1150	1148	Green, ketonic
3-Methylundecane	b	14.7	2.62	1157	/	/
2-Methylpropyl butanoate	b	14.7	1.4	1157	1158	Fruity

Table 10.2: continue

Compound Name	Identification	¹ t _R (min)	² t _R (sec)	Experim. I ^T	Tabulated I ^T	Odor quality
α-Phellandrene	a	14.76	1.74	1159	1160	Terpenic
β-Myrcene	a	14.76	1.52	1159	1162	Peppery, woody
2-Methoxy-1-propanol	b	14.76	0.68	1159	/	/
Pyridine	a	15.17	0.8	1170	1175	Fishy
α-Terpinene	a	15.34	1.76	1174	1177	/
2-Heptanone	a	15.46	1.1	1178	1178	Fruity, sweet
Heptanal	a	15.63	1.18	1182	1181	Oily, fatty
2-Ethyl hexanal	a	15.69	1.34	1184	1188	/
Methyl hexanoate	a	15.75	1.24	1185	1188	Fruity
Dodecane	a	15.93	2.74	1190	1200	/
Limonene	a	16.04	1.74	1193	1202	Terpenic, herbal
3-Methyl-1-butanol	a	16.22	0.6	1198	1207	Fermented, fusel
1,8-Cineole	a	16.39	1.84	1202	1211	Herbal, minty
2-Hexanol	a	16.74	0.68	1211	1220	Chemical, winey
(E)-2-Hexenal	a	16.86	0.94	1215	1223	Bitter almond, green
2-Pentylfuran	a	17.38	1.3	1228	1235	Fruity, green
β-Ocimene	a	17.44	1.58	1230	1238	Floral, green
Ethyl hexanoate	a	17.5	1.44	1231	1240	Sweet, fruity
1-Dodecene	b	17.62	2.28	1234	1242	/
γ-Terpinene	a	17.85	1.78	1240	1249	/
1-Butoxy-2-ethylhexane	b	18.14	2.62	1248	/	/
3-Octanone	a	18.2	1.3	1249	1250	Fresh, mushroom
Styrene	a	18.32	1.04	1252	1257	Balsamic, gasoline
Isoamyl butyrate	b	18.73	1.58	1263	1267	Fruity, waxy
Hexyl acetate	a	19.02	1.3	1271	1274	Fruity
4-Heptanol	b	19.13	0.78	1274	1281	Alcoholic
α-Terpinolene	a	19.25	1.82	1277	1283	Fresh, woody
1,3,5-Trimethylbenzene	b	19.25	1.34	1277	1285	/
3-Hydroxy-2-butanone (acetoin)	a	19.54	0.58	1284	1287	Buttery, creamy
Octanal	a	19.6	1.24	1286	1291	Fatty, sharp
Tridecane	a	19.72	2.86	1289	1300	/
2-Ethyl-2-hexenal	a	19.95	1.24	1295	1314	/
Hexanenitrile	b	20.01	0.9	1296	1305	/
2-Heptanol	a	20.48	0.74	1309	1318	Fresh, fruity
3-Methyl-2-buten-1-ol	b	20.59	0.6	1312	1321	Fruity, sweet
N,N-Dimethyl formamide	b	20.65	0.78	1313	1326	/
(Z)-2-Hexenal	a	20.88	1.02	1319	1326	Leaves, cut grass
α-Methyl styrene	b	21.12	1.18	1325	1327	/
6-Methyl-5-hepten-2-one	a	21.41	1.12	1333	1339	Fatty, green
1-Hexanol	a	21.76	0.68	1342	1340	Fruity, banana
4-Hydroxy-4-methyl-2-pentanone	b	22.34	0.68	1358	1366	/
Dipropyl disulfide	b	22.87	1.64	1371	1372	Sulfury, earthy
2-Nonanone	a	23.33	1.32	1384	1390	Fruity, cheesy
Nonanal	a	23.51	1.3	1388	1397	Fatty, waxy
Tetradecane	a	23.54	2.86	1388	1400	Mild, waxy
2-Butoxyethanol	a	23.74	0.72	1395	1403	/
α-Thujone	a	24.68	1.4	1420	1421	Thujonic
(E)-2-Octenal	a	24.79	1.08	1423	1428	Green, nut
Benzene, 1-ethenyl-3-ethyl-	b	25.01	1.01	1427	1424	
p-Cymenene	b	25.08	1.28	1431	1430	Phenolic, spicy
Ethyl octanoate	a	25.08	1.58	1431	1431	Fruity, waxy
β-Thujone	a	25.38	1.3	1439	1440	Thujonic
1-Heptanol	a	25.55	0.72	1443	1444	Herb
1-Octen-3-ol	a	25.78	0.82	1449	1450	Mushroom, herbal
(E,E)-2,4-Heptadienal	a	26.08	0.9	1457	1465	Fatty, rancid
Furfural	b	26.13	0.66	1459	1466	Bready, brown
Butyl-2-ethylhexanoate	b	26.25	2	1462	1471	/
Acetic acid	a	26.72	0.4	1475	1479	Sour, vinegary
Decanal	a	27.36	1.42	1492	1490	Sweet, waxy
2-Nonanol	a	27.88	0.86	1507	1508	Waxy
(E,E)-3,5-Octadien-2-one	b	28	1	1510	1518	Geranium-like
Camphor	a	28.06	1.44	1512	1512	Camphoreous
Benzaldehyde	a	28.23	0.86	1517	1521	Fruity, sweet

Table 10.2: continue

Compound Name	Identification	¹ t _R (min)	² t _R (sec)	Experim. I ^T	Tabulated I ^T	Odor quality
1,3-Butanediol	a	28.53	0.5	1525	1540	/
(E)-2-Nonenal	a	28.64	1.14	1528	1532	Paper-like, fatty
Linalool	a	28.93	0.86	1537	1537	Floral, citrus
1-Octanol	a	29.23	0.78	1545	1547	Nut, mushroom
5-Methyl-2-furancarboxaldehyde	a	29.51	1.05	1548	1570	Caramel-like
2,3-Butanediol	b	29.75	0.5	1560	1569	Creamy
2-Undecanone	a	30.8	1.42	1590	1592	Fruity, waxy
Undecanal	a	31.03	1.48	1597	1606	Waxy, aldehydic
Benzonitrile	b	31.15	0.8	1600	1597	/
2-(2-ethoxyethoxy)-ethanol	b	31.38	0.7	1607	1599	/
Butanoic acid, 4-chloro-	b	31.85	0.74	1621	/	/
Menthol	a	32.08	0.96	1628	1632	Menthol, cooling
Butyrolactone	b	32.15	1.05	1629	1632	
(E)-2-Decenal	a	32.32	1.22	1635	1642	Fishy, fatty
Acetophenone	a	32.55	0.92	1642	1645	Floral, powdery
Methyl-2-octynoate	a	32.61	1.16	1644	/	/
Butanoic acid	a	32.84	0.44	1650	1647	/
3-Methylbutanoic acid	a	33.75	0.58	1668	1666	Cheese, sweat
α-Terpineol	a	34.01	0.98	1685	1692	Terpenic, citrus
γ-Hexalactone	a	34.3	0.88	1694	1703	Herbal, coconut
Dodecanal	a	34.53	1.52	1701	1709	Soapy, waxy, citrus
Naphthalene	a	35.47	1.14	1730	1740	Pungent
(E)-2-Undecenal	a	35.88	1.26	1743	1750	Fruity, waxy
α,α-Dimethylbenzenemethanol	b	35.99	0.76	1747	1766	/
Methoxy phenyl oxime	b	35.99	0.44	1747	/	/
δ-Hexalactone	a	37.1	0.9	1781	1893	Creamy, chocolate
1,3-Propanediol	b	38.21	0.6	1792	1789	
Butyl benzoate	a	39.43	1.28	1857	1870	Balsamic, fruity
Hexanoic acid	a	40.02	0.48	1876	1880	Cheesy, fatty
Dimethyl sulfone	b	40.37	0.56	1887	1895	/
Benzene ethanol	b	40.66	0.7	1897	1906	Floral
γ-Octalactone	a	40.95	0.96	1907	1925	Coconut, lactonic
Diethylene glycol	a	42.29	0.52	1952	1968	/
δ-Octalactone	a	42.35	1.04	1954	1946	Coconut, creamy
1-Dodecanol	b	42.58	1.28	1960	1966	
Biphenyl	b	43.05	1.22	1978	1986	Green, floral
Heptanoic acid	a	43.52	0.48	1994	1997	Rancid, cheesy
γ-Nonalactone	a	44.22	1.02	2019	2038	Fatty, coconut
Octanoic acid	a	46.55	0.5	2105	2088	Rancid, soapy
2-Phenoxyethanol	b	47.25	0.68	2130	2110	Floral
δ-Decalactone	a	48.71	1.12	2181	2160	Coconut, peach
Nonanoic acid	a	49.58	0.52	2213	2197	Fatty, waxy, cheesy
Farnesol	a	51.16	1.66	2276	2293	Floral, fresh
Decanoic acid	a	52.44	0.54	2328	2309	Soapy, waxy
δ-Undecalactone	b	52.25	1.58	2340	2341	Sweet, coconut
1(3H)-Isobenzofuranone	b	52.85	0.84	2344	2356	/
γ-Dodecalactone	b	53.06	1.2	2364	2365	Creamy, sweet
Diethyl Phthalate	a	53.32	1.12	2363	2378	/
δ-Dodecalactone	a	54.6	1.2	2415	2420	Creamy, sweet
Dodecanoic acid	a	57.87	0.58	2554	2537	Fatty, soapy, coconut

10.2.7 Comparative visualization by visual features

Visual features fingerprinting, with pair-wise image comparison, is a comparative visualization tool that computes the difference at each data point (*i.e.*, the output of the detector at a point in time) between pairs of 2D chromatograms. Absolute or relative response differences are mapped into hue-intensity-saturation (HIS) color space to generate an image for visualizing the compositional differences between image pairs in the retention times plane. The approach enables comparative visualization with compositional differences rendered with an array of colorization modes.

In practice, the algorithm represents GC×GC raw data as an $a [m, n]$ matrix, where a is the *analyzed* chromatogram with indexed pixels by ¹D retention time, m , and ²D retention time, n . Each pixel corresponds to a detector data package and its colorization reveals differential detector response between image pairs (*analyzed* vs. *reference*) [7], [16]. In this study, comparative visualization is by the grayscale *fuzzy difference* rendering. It subtracts pixel-by-pixel the *analyzed* image minus the *reference* image. The range of pixel values is mapped linearly to grayscale from black to white (0 to 255), with dark values (0-127) for negative differences, gray (value 128) for zero difference, and bright values (128-255) for positive differences [21]. This difference is non-zero only if the pixel in one image is outside the range of pixel values in a small neighbourhood in the other image.

10.2.8 Data acquisition and 2D data processing

Data were acquired by TOF-DS™ (Markes International, Llantrisant, UK) and processed by GC Image® suite, Release 2021r1 (GC Image, LLC, Lincoln NE, USA).

Data Processing was by a laptop with a standard processor Intel(R) Core™ i7-7500CPU@2.70GHz, 2904 Mha, 2 core, 4 logic processors; RAM 16Gb; NVIDIA GeForce 940MX.

Statistical analysis and chemometrics were performed using Matlab R2021a (The MathWorks, Inc., Natick, Massachusetts, United States) with the following packages: PCA toolbox (v1.5)[22] and Classification toolbox (v6.0)[23], and XLSTAT statistical and data analysis solution software (Addinsoft 2020, New York, USA).

10.3 Results and discussion

The intriguing concept of Computer Vision to highlight compositional differences between sample pairs is not new in the scenario of data processing tools for C2DC [1], [24]. It can be achieved by combining comparative visualization [7] with a comprehensive mapping of features, *i.e.*, *UT* fingerprinting [9]. However, even for uniform batch analyses, such as those acquired in a limited time frame without any change in the system hardware (*e.g.*, column set, modulation loop alignment, MS calibration and tuning), retention times misalignment might occur preventing consistent comparisons. An example of comparative visualization by colorized fuzzy difference of two slightly misaligned images is represented in **Figure 10.2A-C**. In **Figure 10.2A** the misalignment is along the ¹D in an analysis with a retention offset of 4.05% in the γ -terpinene region, **Figure 10.2B** shows a misalignment in the ²D with an offset of 15.71% in the 3-methyl butanoic acid elution region, and **Figure 10.2C** shows a misalignment in both dimensions. Comparative visualization, if supervised by an analyst could lead to coherent conclusions due to experience and knowledge. If Computer Vision is systematized in a decision-making process, misinterpretation could occur deriving wrong conclusions about the actual compositional differences between samples pair.

In this scenario, approaches that examine the response recorded for peak and/or peak region features are preferable. By applying retention times tolerance windows for features alignment and matching, correct correspondences can be established. Additional specificity is also achieved by including MS similarity constraints above a pre-defined threshold [25].

However, while peak and peak region patterns can be re-aligned and rationalized in peak tables for data mining/data analysis, chromatogram images cannot.

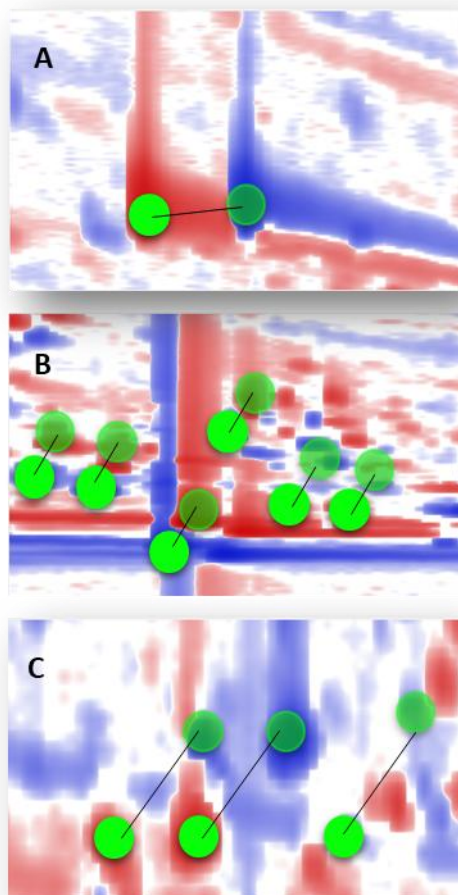


Figure 10.2: (A-C) Comparative visualization as colorized fuzzy ratio of three regions of misaligned chromatograms with ^1D retention shift (A), ^2D retention shifts (B) and both (C). Colored circles indicate the 2D peaks centroids and their relative position in the two chromatograms. Black lines connect peaks of the same feature in the misaligned images and ideally illustrate how template transformation operates to re-align patterns.

To solve this issue and enable an effective and reliable CV, chromatogram alignment and registration in the temporal domain are mandatory. Reichenbach and co-workers [10], [26] and Schmarr *et al.* [27] described similar approaches for the geometrical alignment of features across many chromatograms. Both methodologies adopt peak region features obtained by the alignment and combination of source chromatograms into a single composite chromatogram (*e.g.*, simply by addition or other fusion operations [27]). The feature template, generated by outlining footprint regions around each detected peak in the composite chromatogram, is then used to geometrically remap each source chromatogram. To note, the transform operation to align a chromatogram for compositing is bidirectional and can be used to map the feature template back to the source chromatogram. An example of the transform operation from the template to the chromatogram image is illustrated in **Figure 10.2D-F**

where for the different misalignments described above, the template is matched to the correct peak feature by its transform.

The alignment consistency is improved if the feature template matching is guided by appropriate transform models and additional constraints related to spectral similarity when the MS channel is available. Available transformation models include the global affine transform [28], capable of compensating for linear shifts in both chromatographic dimensions, and the global, low-degree polynomial alignment transformations which have better performances with non-linear and/or severe misalignments [29]. Both these algorithms are available in commercial software (GC Image™, GC Image LLC) as those proposed by Pierce *et al.* [30] and Zhang *et al.* [31]. Other approaches, not yet available in commercial data processing platforms for C2DC, include the piecewise linear interpolation/extrapolation for ¹D shifts [32]; natural-neighbor interpolation for the ²D [33]; semi-parametric warping [34]; and BARCHAN that uses a non-rigid transformation in the first dimension and a rigid transformation in the second dimension [35].

The next sections include an insight into the CV challenges faced approaching a large dataset of chromatograms as those collected along the butter-making process, followed by the description of the novel workflow for Augmented Visualization with molecular resolution. Results are presented and validated against a conventional peak feature approach.

10.3.1 Computer Vision by datapoint features: chromatograms transform and registration

The dataset adopted to develop and validate the CV workflow included 84 chromatograms (7 butter processing steps × 2 inocula × 2 biological replicates × 3 analytical replicates) plus additional quality control samples (QCs) from a commercial butter sample analyzed to check for analytical system consistency over time (n=12). Once the alignment of QCs in terms of peak features response was examined, they were excluded by the CV workflow.

Peaks and peak regions in the comprehensive feature template (see *section 3.2* for template construction) were 346, of them 160 were putatively targeted according to criteria described in *section 2.6* and 186 were left untargeted by assigning them a unique identifier. The feature template was matched, transformed, and aligned over the 84 source chromatograms. Features metadata were collected in a spreadsheet for data evaluation and machine learning (*section 3.3*).

Retention time fluctuations along the two chromatographic dimensions were at first examined since they would prevent a direct image comparison (*cf.* **Figure 10.2A-C**). Along the ¹D, retention times accounted for a % relative standard deviation (%RSD) computed on their absolute values of 0.43 mean/0.25 median with a minimum value of 0.07 for γ -dodecalactone and a maximum of 5.82 for γ -terpinene. The non-normal distribution of t_R values was connected to the retention variability/inconsistency of higher volatiles likely impacted by extra-chromatographic phenomena related to SPME fiber desorption kinetic into the GC injector. On the other hand, ²D misalignment showed %RSD values of 8.77 mean/8.27 median with a minimum value of 0.1 for *n*-C₁₄ (tetradecane) and a maximum of 16.7 for 3-methyl butanoic acid. This variability, as expected, was higher because of the simplified system

configuration adopted that did not include a secondary oven isolated by the loop-type thermal modulator operating with liquid nitrogen as a coolant.

Figure 10.3A shows an enlarged area of a sample's chromatogram (*B STD-T20 R2_rep #3*) with the superimposed comprehensive feature template, containing peak features objects (red areas). By matching and transforming with polynomial second-order function (**Figure 10.3B**) the template compensates for retention time shifts. To note, the enlarged area corresponds to the elution region of butanoic acid (larger peak) and 3-methylbutanoic acid (black arrow) that was affected by a severe misalignment across the chromatograms set.

Comparative visualization is highly dependent on chromatogram registration by feature template matching, otherwise it would fail in precisely informing about compositional differences. As shown in **Figure 10.3C** without registration, several peak regions would be misaligned (yellow arrows); after registering, they were effectively aligned as shown in **Figure 10.3D**.

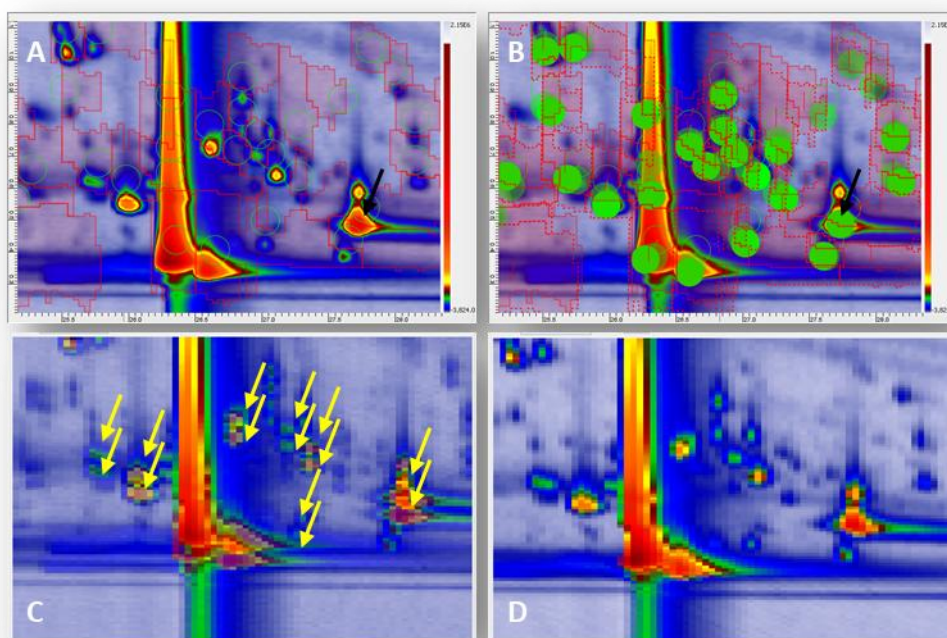


Figure 10.3: Enlarged section of the *B STD-T20 R2_rep n°3* 2D chromatogram, between 25 and 28min (t^1D), 0.2 and 1.1s (t^2D). In figure 3A the feature template is over imposed and (3B) the retention time shifts corrected operated by the template matching algorithm. Comparative analysis between two misaligned chromatograms visualized in transparency before (3C) and after (3D) image registration.

After the temporal misalignment issue was solved, the possibility of achieving an Augmented Visualization with access to chromatogram metadata, *i.e.*, information about features' identity, was considered. The concept of AV refers to “*computational techniques for visualizing what cannot be seen with raw image input*” [11]. With the software platform here adopted, comparative visualization is linked to feature (data points and peaks) metadata. Therefore, once the information contained in the feature template is linked to the

chromatographic image, comparative visualization is *augmented* by molecular information giving access to higher-level information about the phenomenon under study.

10.3.2 Augmented Visualization by UT fingerprinting on composite class images: principle and workflow

The novel workflow was based on the UT fingerprinting automated process, a routine available in GC Image Investigator™ (GC Image™, GC Image LLC). The procedure is described in detail in the experimental section. In practice, samples were grouped in *classes* as a function of the processing stage: raw sweet cream (*SCR*); pasteurized sweet cream inoculated with selected cultures and analyzed at time 0 and after 18 hours (*PCT0* and *PCT18*); butter samples analyzed along ripening at 0-8-20-40 days (*BT0*, *BT8*, *BT20*, *BT40*).

The process was carried out with the primary aim of samples' chromatogram re-alignment and combination into a composite class image representing all detected features in that class. The number of source chromatograms for each class was homogeneous, 6 chromatograms (2 biological replicates $\times$ 3 analytical replicates) were automatically pre-processed (file import, rasterization, colorization, baseline correction, 2D peaks detection and integration) with method parameters optimized by the analyst according to the sample compositional complexity, chromatographic resolution and absolute sensitivity. S/N threshold was set at 50 and baseline corrected according to Reichenbach *et al.* using the *default* algorithm [36]. Then, a comprehensive pair-wise peak matching was run for a total of 30 cross-matches [*i.e.*, (6 $\times$ 6)-6] and a computational time of about 30 minutes. The determination of reliable registration peaks was by most relaxed constraint and included 2D peaks that matched in at least 50% +1 of the chromatograms. This last operation took less than one minute and selected on average 1,000 peaks per class. Note that the higher the number of anchor points for template matching, the better the image registration. The 2D chromatograms were then aligned according to the set of reliable registration peaks, transformed and registered before saving data, and making them available for combination into a composite image. For each chromatogram, this process took about 2 minutes. As the last step, the composite chromatogram (Class Image) was generated by the combination of the registered images in a single one, which was then processed for baseline correction and 2D peaks detection. The reliable template was matched and graphic objects (*i.e.*, areas corresponding to peak region features, shown with the red contour in **Figure 10.3A-B**) were generated around all detected 2D peaks before their inclusion in the final feature template. The total computational time from raw data pre-processing until the final step with feature template composition was about 50 minutes per Class.

A schematic diagram of the workflow is illustrated in **Figure 10.1** (left side) while an in-depth scheme is provided in **Supplementary Figure 2**.

The seven Class Images were then used to feed a new UT fingerprinting workflow (**Figure 10.1**, right side) at this stage to re-align them against a comprehensive feature and vice-versa, in a way to obtain transformed and realigned Class Images for effective pair-wise comparisons. Moreover, the comprehensive feature template obtained at this stage was targeted by adding putative identifications for known components (see *section 2.6* for identification criteria) and matched back to all single chromatograms (*i.e.*, single samples $n=84$ and Class Images $n=7$).

This global re-alignment enabled AV with access to the higher-level information and traditional data mining by machine learning on the complete data set (*i.e.*, 84 chromatograms × 346 features).

The next section illustrates as proof of concept some investigation possibilities offered by AV with registered Class Images. The flexibility of the workflow is so high that the analyst can answer many different questions just by properly designing the Class Images composition and consequent pair-wise matches.

10.3.3 Molecular patterns in butter processing

The volatile fraction of butter and butter oil has been the object of several studies since the mid-1950s; one of the most complete lists of known components (*i.e.*, targeted volatiles) accounts for 287 volatiles [37]. Within them, the sub-group of potent odorants and key-aroma compounds play a crucial role in consumers' acceptability and preferences [38]. However, many other volatiles contribute to the definition of butter quality. For example, terpenes and terpenoids, transferred from cow feeding to milk and sweet cream, inform about feeding systems (*e.g.*, pasture or indoor feeding [39]), esters naturally formed during fermentation and cream acidification, are characteristic of the inocula, since precursors (*i.e.*, primary and secondary alcohols and acids) are formed by specific metabolic pathways [40]. Lactones, formed by the intra-esterification of hydroxyl acids [41], leave a diagnostic signature on many dairy products.

The evolution of the detectable volatilome along the butter production chain was examined by sequential pair-wise comparisons of Class Images. AV captured the compositional differences of products against the primary material, *i.e.*, the raw sweet cream. By these approaches, volatiles with increasing or decreasing trends are easily mapped over the chromatographic space thanks to specific colorizations, while metadata inspection enables access to higher-level information. **Figure 10.4** shows as a heat map the fold change for a selection of features that showed the largest variations within the processing steps. Fold change is calculated through the ratio of the target analytes' absolute responses from Class Images and taking the raw sweet cream Class Image as a reference. Comparisons are between *PCT*₀ vs. *SCR* (pasteurized cream at T₀ vs. raw sweet cream – S₁), *PCT*₁₈ vs. *SCR* (pasteurized cream at T₁₈ vs. raw sweet cream – S₂), *BT*₀ vs. *SCR* (butter at T₀ vs. raw sweet cream – S₃), *BT*₈ vs. *SCR* (butter at T₈ vs. raw sweet cream – S₄), *BT*₂₀ vs. *SCR* (butter at T₂₀ vs. raw sweet cream – S₅), *BT*₄₀ vs. *SCR* (butter at T₄₀ vs. raw sweet cream – S₆).

Analytes that showed the largest variations, with a progressive increment during processing are volatile acids (acetic acid, butanoic acid, and hexanoic acid) reaching a relative abundance in the ripened butter after 40 days, which corresponds to one order of magnitude higher than in the raw sweet cream. They are formed by both fermentation of milk primary metabolites and by enzymatic hydrolysis of triglycerides. Another analyte of interest for butter aroma identity is acetoin (3-hydroxy-2-butanone) [42], [43]; it is formed from sugars through the pyruvate pathway by the lactic acid bacteria metabolism [44]. It reaches a maximum after inoculation of ferments in pasteurized raw cream (S₂ comparison – 15.3 fold-change) and then stays stable along butter ripening. Interestingly, its reduction product, 2,3-butanediol, has a similar trend. Another group of analytes showing meaningful variations are lactones, they

contribute to the *creamy* and *sweet* aroma of milk and dairy products [45]; γ -nonalactone, δ -decalactone and δ -dodecalactone have an incremental trend along processing.

Acetic acid	5.52	7.25	8.64	11.88	10.15	11.13
Butanoic acid	2.60	4.45	4.68	6.60	5.61	6.39
Hexanoic acid	1.56	2.10	2.05	2.33	2.17	2.28
Octanoic acid	1.27	1.47	0.78	1.17	1.16	1.11
Nonanoic acid	0.60	0.76	0.69	0.75	0.72	0.69
Acetone	1.10	1.08	0.85	0.51	0.91	0.50
3-Hydroxy-2-butanone (Acetoin)	4.05	15.28	9.82	12.27	9.75	9.83
2-Butanone	0.54	0.40	0.26	0.36	0.34	0.39
2-Heptanone	3.55	1.73	3.11	1.99	2.94	3.14
Hexanal	0.58	0.86	0.34	0.38	0.48	0.50
Octanal	1.13	1.82	1.14	1.88	1.23	1.02
Nonanal	1.31	1.33	0.77	1.32	0.81	0.71
Ethanol	1.66	1.59	1.46	1.75	1.12	2.01
1,3-Propanediol	12.27	15.29	5.63	4.69	13.03	13.08
2,3-Butanediol Isomer	25.66	33.03	17.41	22.05	22.78	23.13
1-Hexanol	4.60	6.77	1.88	2.21	1.74	1.77
1-Dodecanol	0.65	0.66	0.94	0.26	1.00	0.98
Acetic acid ethenyl ester	2.40	4.53	4.27	4.56	3.50	3.63
Butyrolactone	1.28	1.61	1.29	1.79	2.11	2.89
γ -Nonalactone	0.05	0.05	0.50	0.80	0.90	0.81
δ -Decalactone	1.44	1.82	1.92	2.24	2.24	2.37
γ -Hexalactone	0.31	0.13	0.18	0.30	0.41	0.50
δ -Dodecalactone	1.70	1.63	2.11	6.41	6.84	5.03
Dimethyl sulfide	0.69	1.08	0.63	0.45	0.53	0.95
Dimethyl sulfone	1.00	0.01	0.03	0.15	0.79	0.73
	S1	S2	S3	S4	S5	S6

Figure 10.4: Heatmap of target analytes showing the largest variations (indicated as fold change) relative to raw sweet cream. The heatmap colorization (light blue to dark blue) is scaled on the row values. Comparisons are between PCTo vs. SCR – S₁, PCT₁₈ vs. SCR – S₂, BTo vs. SCR – S₃, BT₈ vs. SCR – S₄, BT₂₀ vs. – S₅, BT₄₀ vs. SCR – S₆.

AV facilitates the prompt capture of these information, **Figure 10.5A** shows the comparative visualization rendered as *grayscale fuzzy ratio* between PCTo vs. SCR (pasteurized cream at To vs. raw sweet cream – S₁). The yellow boxes over imposed on the chromatographic plane (zoomed area with green lines) highlight the peak region of acetoin (3-hydroxy-2-butanone). The access to relative and absolute responses, bottom tabular area, is facilitated and eventually guided by analysts with a mouse click on the chromatogram image. In **Figure 10.5B**, another possibility is a comparative side-by-side visualization. The red box indicates the butanoic acid peak regions in the comparison between Class Images BT₄₀ vs. SCR (butter at T₄₀ vs. raw sweet cream – S₆).

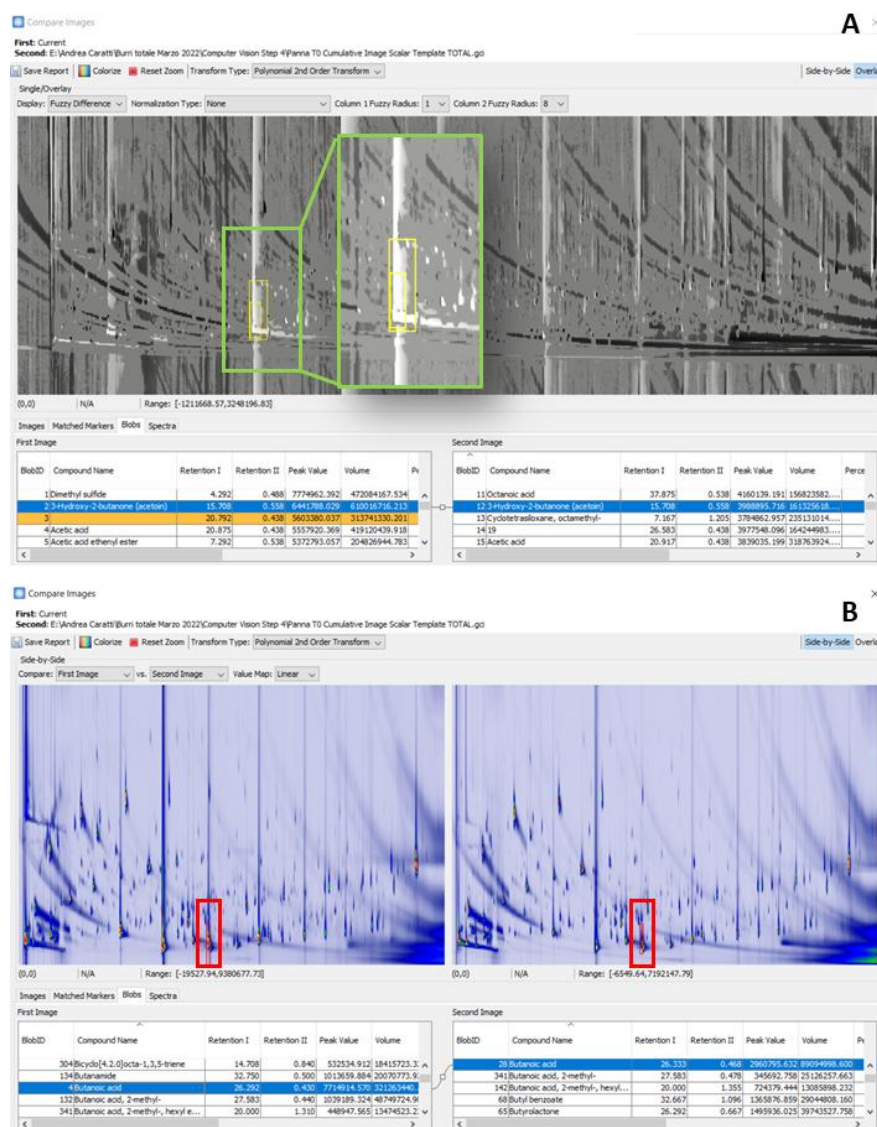


Figure 10.5: (5A) Comparative overlaid visualization rendered as grayscale fuzzy ratio between PCTo vs. SCR – S1. The peak region of 3-hydroxy-2-butanone (Acetoin – yellow boxes) is highlighted and zoomed within green areas, and its metadata in the two chromatograms are reported in the tabular area. (5B) Comparative side-by-side visualization between BT4o vs. SCR – S6. Butanoic acid peak regions with its metadata in both chromatograms is highlighted by red boxes.

Validation of the CV results was by conventional UT fingerprinting on the complete data matrix including 84 chromatograms $\times$ 28 features (having Fisher ratio values > 8). Peak region features absolute responses for the three randomized replicated analyses were examined by Partial Least Square Discriminant Analysis (PLS-DA) after range scaling to highlight variables capable of classes discrimination (*i.e.*, SCR, PCTo, PCT18, BTo, BT8, BT2o, and BT4o). All PLS-DA models were cross-validated via Montecarlo validation excluding 20% of the dataset out of the model construction and using it as a validation set; this validation process was reiterated 1,000 times to satisfy robustness requirements. PLS-DA results are visualized in **Figure 10.6A**; from left to right processing steps are distributed over the Cartesian space delineated by the first two components. Sweet raw cream and pasteurized cream at to form clear clusters, due to an almost similar composition of their volatile fraction; after 18 hours from the inoculum (PCT18 – yellow indicators) the volatilome is

evolving/changing and samples dispersion increases coherently with the logarithmic development of lactic bacteria. Then, after fat crystallization and butter ripening (*BTo* → *BT40*), sample clusters show some overlapping. Of the most informative variables, according to Variable Importance for the Projections (VIPs), capable of effectively discriminating all classes (i.e., $VIPs \pm SD > 1$), acetic acid ranked second after an untargeted feature (#115 not identified). They are then followed with decreasing VIPs values by: 1-heptanol; butanoic acid; propanoic acid; 1-octen-3-ol; feature #48; acetic acid ethenyl ester; feature #285; 3-hydroxy-2-butanone (acetoin); δ -decalactone; 3-methylbutanoic acid; hexanoic acid; γ -nonalactone; and 2,3-butanediol isomers. Most of them were promptly revealed by CV by Class Images comparison. PLS-DA in **Figure 10.6A** was built with 7 latent variables (LVs), and achieved good sensitivity, specificity, and precision results; in particular, the best classification performances are achieved for the raw and pasteurized sweet cream, while worse discriminating capabilities are achieved when trying to discriminate butter samples according to shelf life. As reported in the confusion matrix and in the ROC curves (**Supplementary Figure 3**), the greatest mistakes happen within successive time points, this may be recollected to the fact that different samples in different physical areas of the butter are at different ripening stages due to inhomogeneous distribution of microbial in the matrix.

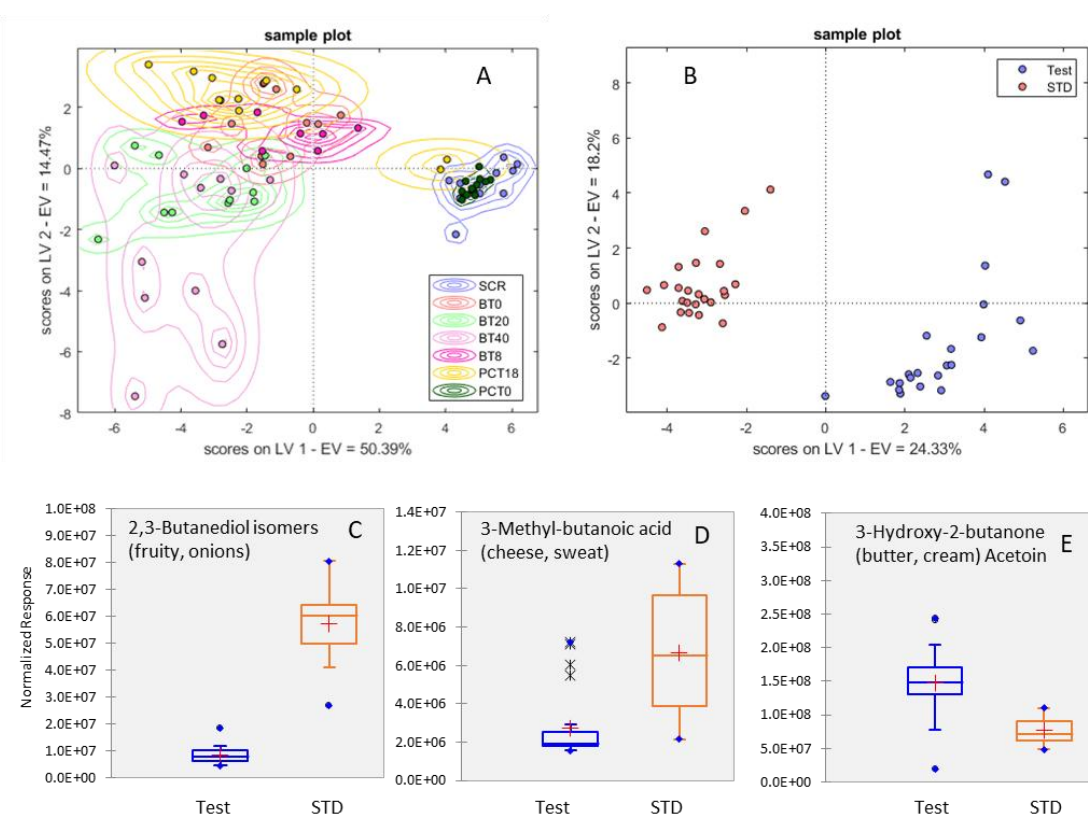


Figure 10.6: (6A) PLS-DA (82 samples x 28 variables with Fisher ratio > 8) score plot on the first two latent variables of the various steps of artisanal butter production and ripening from raw sweet cream after range scaling as pre-processing step. Montecarlo validation 20% out was conducted for model validation and performance evaluation. (6B) PLS-DA (48 samples x 49 variables with Fisher ratio > 4) score plot on the first two latent variables of the two butter samples deriving from different microbial inocula (Test vs STD) after range scaling as pre-processing step. Montecarlo validation 20% out was conducted for model validation and performance evaluation. (6C, 6D, 6E) boxplots showing the absolute response distribution of the most impacting key odorants in Test vs STD butter samples.

By further investigating the differential impact of the two inocula (i.e., Test vs. STD), a PLS-DA was conducted on UT peak region features normalized responses from butter samples (from 0 to 40 days of ripening). Results are visualized in Figure 10.6B where the sample classes are represented by pink and purple indicators. The model is guided by 28 UT features with a VIPs $\pm$ SD >1, of them, those with the highest odor impact are 2-heptanol; 2,3-butanediol isomers; 1,3-propanediol; octanal; 3-methylbutanoic acid; acetic acid; 3-hydroxy-2-butanone (acetoin); 1-octen-3-ol; 2-heptanone; 2-butanone; 5-methyl-2-furancarboxaldehyde; δ -undecalactone; and γ -butyrolactone. Key-odorants differential distribution is rendered in box-plots of **Figures 10.6C-E** [2,3-butanediol isomers; 3-methylbutanoic acid; and 3-hydroxy-2-butanone (acetoin)]. The second model created on two butter samples deriving by different microbial inocula (STD vs Test) shows optimal classification parameters as reported in **Supplementary Figure 3**. This confirms how the volatile fraction is heavily influenced by the microbial culture.

Augmented visualization results on Class Images were validated by conventional data mining (PLS-DA) on UT fingerprinting on single chromatogram Images. AV by combining data from different samples minimizes the effect of confounding/secondary variables (e.g., sweet cream suppliers, inoculum bacteria, and timing) while facilitating the prompt capture of common features. Analytes with a clear increment along processing are also those with a classification role in the class model; they are acetic acid; butanoic acid; hexanoic acid; 3-hydroxy-2-butanone (acetoin); 2,3-butanediol isomers; acetic acid ethenyl ester; γ -nonalactone; and δ -decalactone. Those which were not cross-validated have a secondary information role most related to the secondary variables' impact.

10.4 Conclusions

This study demonstrates how CV combined with chromatographic fingerprinting can effectively enable Augmented Visualization facilitating access to samples' compositional data and interpretation of the phenomena related to it. The capture of diagnostic patterns within the complex fraction of volatiles from food (food volatilome) is complicated by the concurrent effect of many biological phenomena which could hide informative patterns.

The multi-step workflow proposed is successful in realigning and combining chromatograms from samples belonging to specific classes, pre-defined based on the investigation questions. Once generated, Class Images can be pairwise compared with easy access to peak feature metadata. As a bench test, the butter volatilome was investigated for its evolution along the production chain. Besides the presence of confounding variables, step-wise comparisons promptly revealed analytes and analyte patterns with increasing trends from primary material (i.e. sweet cream) to the final product (i.e., ripened butter at 40 days).

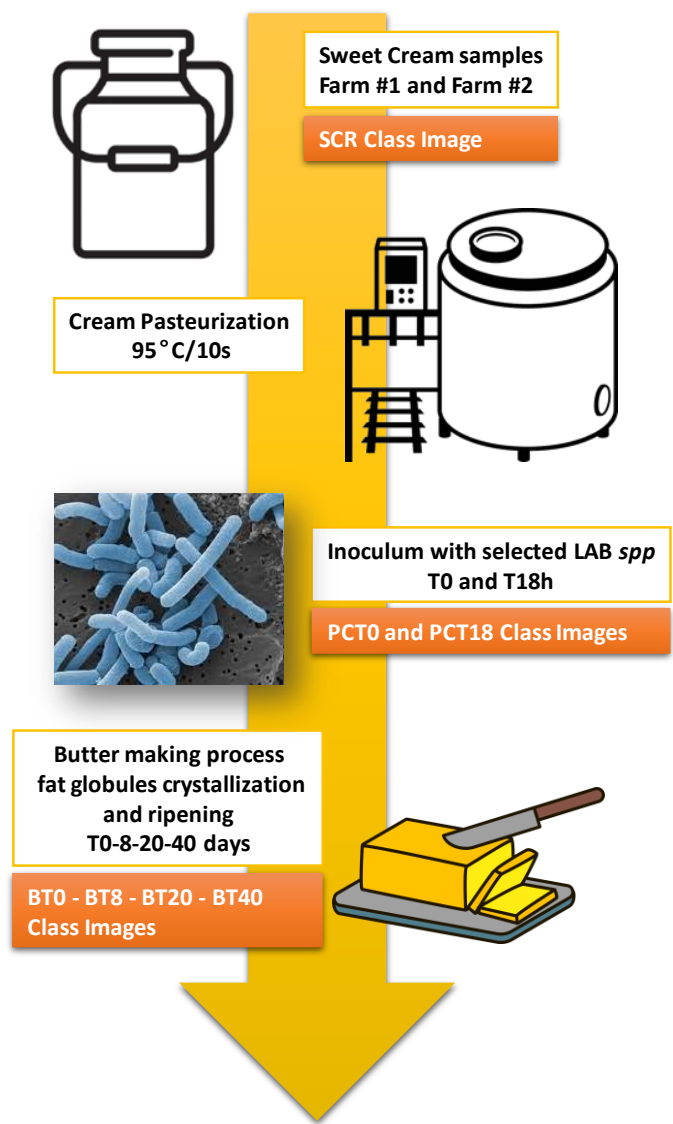
The workflow is highly flexible, reliable, and robust by compensating for temporal misalignments occurring in large time-frame studies. Validation against a conventional approach based on UT fingerprinting on peak and peak region features confirms its consistency, supports its adoption as decision-making tool in food quality applications, and suggests its application in many other fields.

In a scenario where AI algorithms can support decisions in many fields, the high-resolution separation and multi-level information provided by C2DC are ideally suited to the Augmented Visualization processes.

Acknowledgments

This work was carried out within the “Milk Zero” Project: Effetti della gestione delle produzioni nell'azienda zootecnica sulla presenza di contaminanti nei prodotti lattiero-caseari: Studio di una filiera a residuo zero in Piemonte. Bando di ricerca: PSR 2014-2020 - 16.2.1 - Az. 2 - Progetti pilota - Piattaforma tecnologica bioeconomia.

Supplementary Information

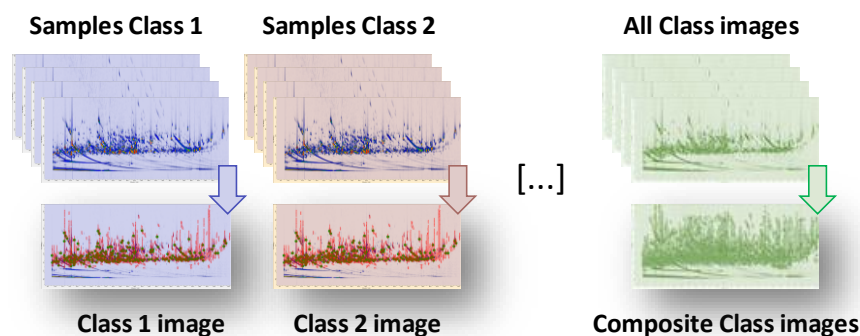


Supplementary Figure 1: schematic diagram of butter making process is illustrated in

Step 1 - Generation of Class Images

Beginning with **84** chromatograms from real samples
(7 classes × 2 inocula × 2 biological replicates × 3 analytical replicates)

- Pre-processing: file import , rasterization, colorization
baseline correction
2D peaks detection and integration
- Comprehensive pair-wise peak matching
- Determination of *reliable* registration peaks (most relaxed constraint)
- Alignment of 2D chromatograms
- Generation of composite chromatograms (Class Images)



Step 2 - Building a *comprehensive feature template*

Comprehensive feature template is generated on the reference
Composite Class Image generated by cross-alignment and matching of
all class images (*i.e.*, Class 1, Class 2, etc.)

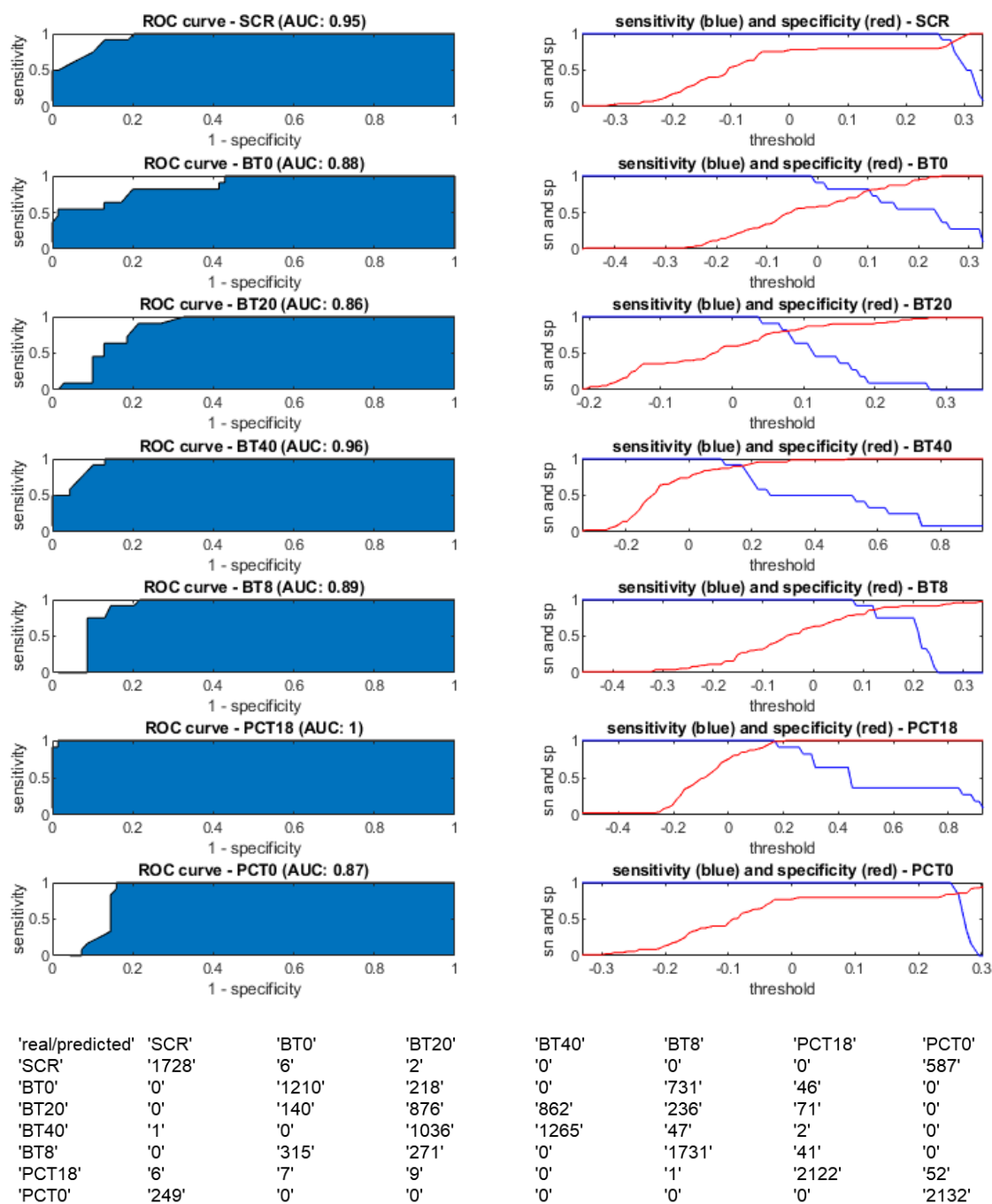
- definition of pattern of peak-regions for all detected 2D peaks
- building of a template with reliable peaks and peak-regions
(including targeted and untargeted features) - UT template
- Targeting by Linear Retention Index (I') calibration and MS spectral
similarity (NIST Algorithm)

Step 3 - Computer Vision

Pattern recognition and comparative visualization

- Matching the *comprehensive feature template* on all single
chromatogram images and on all Class Images
- Perform comparative visualization to highlight pattern differences
between image pairs
- Extract metadata from aligned features (^1t_R , ^2t_R , % response, MS
spectrum etc.) in image pairs
- Access to higher level information

Supplementary Figure 2: details on the Computer Vision processing workflow (see also **Figure 1**).



Supplementary Figure 3: Performance parameters expressed as ROC curves and confusion matrix of the classification model of the different steps in the butter production chain and further ripening stage.

References

- [1] F. Stilo, C. Bicchi, A. M. Jimenez-Carvelo, L. Cuadros-Rodriguez, S. E. Reichenbach, and C. Cordero, "Chromatographic fingerprinting by comprehensive two-dimensional chromatography: Fundamentals and tools," *TrAC Trends Anal. Chem.*, vol. 134, p. 116133, Jan. 2021.
- [2] S. E. Reichenbach, X. Tian, C. Cordero, and Q. Tao, "Features for non-targeted cross-sample analysis with comprehensive two-dimensional chromatography," *J. Chromatogr. A*, vol. 1226, pp. 140–148, Feb. 2012.
- [3] F. Stilo, C. Bicchi, A. Robbat, S. E. Reichenbach, and C. Cordero, "Untargeted approaches in food-omics: The potential of comprehensive two-dimensional gas chromatography/mass spectrometry," *TrAC Trends Anal. Chem.*, vol. 135, p. 116162, Feb. 2021.
- [4] B. A. Parsons, L. C. Marney, W. C. Siegler, J. C. Hoggard, B. W. Wright, and R. E. Synovec, "Tile-Based Fisher Ratio Analysis of Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry (GC $\times$ GC-TOFMS) Data Using a Null Distribution Approach," *Anal. Chem.*, vol. 87, no. 7, pp. 3812–3819, Apr. 2015.
- [5] L. C. Marney, W. Christopher Siegler, B. A. Parsons, J. C. Hoggard, B. W. Wright, and R. E. Synovec, "Tile-based Fisher-ratio software for improved feature selection analysis of comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry data," *Talanta*, 2013.
- [6] P. E. Sudol, G. S. Ochoa, and R. E. Synovec, "Investigation of the limit of discovery using tile-based Fisher ratio analysis with comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry," *J. Chromatogr. A*, 2021.
- [7] B. V. Hollingsworth, S. E. Reichenbach, Q. Tao, and A. Visvanathan, "Comparative visualization for comprehensive two-dimensional gas chromatography," in *Journal of Chromatography A*, 2006, vol. 1105, no. 1–2 SPEC. ISS., pp. 51–58.
- [8] "What is Computer Vision?" [Online]. Available: <https://www.ibm.com/topics/computer-vision>. [Accessed: 25-Nov-2022].
- [9] F. Magagna *et al.*, "Combined untargeted and targeted fingerprinting with comprehensive two-dimensional chromatography for volatiles and ripening indicators in olive oil," *Anal. Chim. Acta*, vol. 936, pp. 245–258, Sep. 2016.
- [10] S. E. Reichenbach, C. A. Zini, K. P. Nicolli, J. E. Welke, C. Cordero, and Q. Tao, "Benchmarking machine learning methods for comprehensive chemical fingerprinting and pattern recognition," *J. Chromatogr. A*, vol. 1595, pp. 158–167, Jun. 2019.
- [11] S. Mori and H. Saito, "An overview of augmented visualization: observing the real world as desired," *APSIPA Trans. Signal Inf. Process.*, vol. 7, no. 1, 2018.
- [12] L. Cuadros-Rodríguez, C. Ruiz-Samblás, L. Valverde-Som, E. Pérez-Castaño, and A. González-Casado, "Chromatographic fingerprinting: An innovative approach for food 'identification' and food authentication – A tutorial," *Anal. Chim. Acta*, vol. 909, pp. 9–23, Feb. 2016.

- [13] Y. Wang, J. O'Reilly, Y. Chen, and J. Pawliszyn, "Equilibrium in-fibre standardisation technique for solid-phase microextraction," *J. Chromatogr. A*, vol. 1072, no. 1, pp. 13–17, Apr. 2005.
- [14] C. Cordero *et al.*, "High concentration capacity sample preparation techniques to improve the informative potential of two-dimensional comprehensive gas chromatography–mass spectrometry: Application to sensomics," *J. Chromatogr. A*, vol. 1318, pp. 1–11, Nov. 2013.
- [15] L. Nicolotti *et al.*, "Quantitative fingerprinting by headspace-Two-dimensional comprehensive gas chromatography-mass spectrometry of solid matrices: Some challenging aspects of the exhaustive assessment of food volatiles," *Anal. Chim. Acta*, vol. 798, pp. 115–125, 2013.
- [16] D. Bressanello *et al.*, "Combined untargeted and targeted fingerprinting by comprehensive two-dimensional gas chromatography: revealing fructose-induced changes in mice urinary metabolic signatures," *Anal. Bioanal. Chem.*, vol. 410, no. 11, pp. 2723–2737, 2018.
- [17] F. Stilo, E. Liberto, S. E. Reichenbach, Q. Tao, C. Bicchi, and C. Cordero, "Untargeted and Targeted Fingerprinting of Extra Virgin Olive Oil Volatiles by Comprehensive Two-Dimensional Gas Chromatography with Mass Spectrometry: Challenges in Long-Term Studies," *J. Agric. Food Chem.*, vol. 67, no. 18, pp. 5289–5302, May 2019.
- [18] S. Squara *et al.*, "Extending the breadth of saliva metabolome fingerprinting by smart template strategies and effective pattern realignment on comprehensive two-dimensional gas chromatographic data," *Anal. Bioanal. Chem.*, no. 0123456789, Jan. 2023.
- [19] N. I. of S. and Technology, *NIST/EPA/NIH Mass Spectral Library with Search Program: (Data Version: NIST 08, Software Version 2.0f)*, vol. 2010, no. 1/22. Gaithersburg MD: National Institute of Standards and Technology (NIST), 2005.
- [20] S. E. Reichenbach, X. Tian, A. A. Boateng, C. A. Mullen, C. Cordero, and Q. Tao, "Reliable peak selection for multisample analysis with comprehensive two-dimensional chromatography," *Anal. Chem.*, vol. 85, no. 10, pp. 4974–4981, 2013.
- [21] GC Image™, "GC Image GCxGC Edition Users' Guide," 2017.
- [22] D. Ballabio, "A MATLAB toolbox for Principal Component Analysis and unsupervised exploration of data structure," *Chemom. Intell. Lab. Syst.*, vol. 149, pp. 1–9, 2015.
- [23] D. Ballabio and V. Consonni, "Classification tools in chemistry. Part 1: Linear models. PLS-DA," *Analytical Methods*, vol. 5, no. 16, pp. 3790–3798, 2013.
- [24] F. Stilo *et al.*, "An effective chromatographic fingerprinting workflow based on comprehensive two-dimensional gas chromatography – Mass spectrometry to establish volatiles patterns discriminative of spoiled hazelnuts (*Corylus avellana* L.)," *Food Chem.*, vol. 340, p. 128135, Mar. 2021.
- [25] S. E. Reichenbach, P. W. Carr, D. R. Stoll, and Q. Tao, "Smart Templates for peak pattern matching with comprehensive two-dimensional liquid chromatography," *J. Chromatogr. A*, vol. 1216, no. 16, pp. 3458–3466, 2009.

- [26] C. Cordero *et al.*, "Targeted and non-targeted approaches for complex natural sample profiling by GC×GC-qMS," *J. Chromatogr. Sci.*, vol. 48, no. 4, pp. 251–261, 2010.
- [27] H.-G. Schmarr and J. Bernhardt, "Profiling analysis of volatile compounds from fruits using comprehensive two-dimensional gas chromatography and image processing techniques," *J. Chromatogr. A*, vol. 1217, no. 4, pp. 565–574, 2010.
- [28] S. E. Reichenbach *et al.*, "Alignment for Comprehensive Two-Dimensional Gas Chromatography with Dual Secondary Columns and Detectors," *Anal. Chem.*, vol. 87, no. 19, pp. 10056–10063, 2015.
- [29] D. W. Rempe, S. E. Reichenbach, Q. Tao, C. Cordero, W. E. Rathbun, and C. A. Zini, "Effectiveness of Global, Low-Degree Polynomial Transformations for GC×GC Data Alignment," *Anal. Chem.*, vol. 88, no. 20, pp. 10028–10035, 2016.
- [30] K. M. Pierce, L. F. Wood, B. W. Wright, and R. E. Synovec, "A comprehensive two-dimensional retention time alignment algorithm to enhance chemometric analysis of comprehensive two-dimensional separation data," *Anal. Chem.*, vol. 77, no. 23, pp. 7735–7743, 2005.
- [31] D. Zhang, X. Huang, F. E. Regnier, and M. Zhang, "Two-Dimensional Correlation Optimized Warping Algorithm for Aligning GC×GC–MS Data," *Anal. Chem.*, vol. 80, no. 8, pp. 2664–2671, Apr. 2008.
- [32] J. Gros, D. Nabi, P. Dimitriou-Christidis, R. Rutler, and J. S. Arey, "Robust algorithm for aligning two-dimensional chromatograms," *Anal. Chem.*, vol. 84, no. 21, pp. 9033–9040, 2012.
- [33] Y. Zushi, J. Gros, Q. Tao, S. E. Reichenbach, S. Hashimoto, and J. S. Arey, "Pixel-by-pixel correction of retention time shifts in chromatograms from comprehensive two-dimensional gas chromatography coupled to high resolution time-of-flight mass spectrometry," *J. Chromatogr. A*, vol. 1508, pp. 121–129, 2017.
- [34] W. P. H. De Boer and J. Lankelma, "Two-dimensional semi-parametric alignment of chromatograms," *J. Chromatogr. A*, vol. 1345, pp. 193–199, 2014.
- [35] C. Couprie, L. Duval, M. Moreaud, S. Hénou, M. Tebib, and V. Souchon, "BARCHAN: Blob Alignment for Robust CHromatographic ANALysis," *J. Chromatogr. A*, vol. 1484, pp. 65–72, 2017.
- [36] S. E. Reichenbach, M. Ni, D. Zhang, and E. B. Ledford, "Image background removal in comprehensive two-dimensional gas chromatography," *J. Chromatogr. A*, vol. 985, no. 1–2, pp. 47–56, 2003.
- [37] H. Maarse, *Volatile Compounds in Foods and Beverages*. Routledge, 2017.
- [38] G. Reineccius, "Front Matter," in *Flavor Chemistry and Technology, Second Edition*, CRC Press, 2005, p. 489.
- [39] T. F. O'Callaghan *et al.*, "Quality characteristics, chemical composition, and sensory properties of butter from cows on pasture versus indoor feeding systems," *J. Dairy Sci.*, vol. 99, no. 12, pp. 9441–9460, Dec. 2016.
- [40] S. Mallia, F. Escher, and H. Schlichtherle-Cerny, "Aroma-active compounds of butter: A

review,” *Eur. Food Res. Technol.*, vol. 226, no. 3, pp. 315–325, 2008.

- [41] S. R. Lee, C. Macku, and T. Shibamoto, “Isolation and identification of headspace volatiles formed in heated butter,” *J. Agric. Food Chem.*, vol. 39, no. 11, pp. 1972–1975, Nov. 1991.
- [42] P. Schieberle, K. Gassenmeier, H. Guth, A. Sen, and W. Grosch, “Character Impact Odour Compounds of Different Kinds of Butter,” *LWT - Food Sci. Technol.*, vol. 26, no. 4, pp. 347–356, Aug. 1993.
- [43] D. G. Peterson and G. A. Reineccius, “Characterization of the volatile compounds that constitute fresh sweet cream butter aroma,” *Flavour Fragr. J.*, vol. 18, no. 3, pp. 215–220, May 2003.
- [44] V. M. Dillon, “NATURAL ANTI-MICROBIAL SYSTEMS | Preservative Effects During Storage,” in *Encyclopedia of Food Microbiology*, Elsevier, 2014, pp. 941–947.
- [45] M. Preininger and W. Grosch, “Evaluation of Key Odorants of the Neutral Volatiles of Emmentaler Cheese by the Calculation of Odour Activity Values,” *Leb. und-Technologie*, vol. 27, pp. 237–244, 1994.

Chapter 11

Translating 2D-Chromatographic Fingerprinting to Quantitative Volatilomics: Unrevealing Compositional Changes in Maize Silage Volatilome for Robust Markers Discovery

This study examines the complex volatilome of maize silage, both with and without commercial heterolactic strains inoculation, conserved for 100 days, using quantitative volatilomics. Chemical classes linked to microbial metabolism were analyzed across a concentration range from 10 $\mu\text{g g}^{-1}$ to 1 ng g^{-1} . A reference method using comprehensive two-dimensional gas chromatography (GC $\times$ GC) and time-of-flight mass spectrometry (TOF MS) with loop-type thermal modulation (TM) was translated to a differential-flow modulation (FM) platform with parallel MS and flame ionization detector (FID) detection. With translation the original method analytes elution order and resolution are preserved. The new method allowed accurate quantification using multiple headspace solid phase microextraction (MHS-SPME) and FID-predicted relative response factors (RRF). Both methods showed comparable discriminatory power, with FM GC $\times$ GC-MS/FID achieving satisfactory quantification accuracy without external calibration. Analysis of 98 volatiles provided insights into silage fermentation, supporting marker discovery and correlations with silage quality and stability.

This chapter was published as Original Research Paper in the Journal Agricultural and Food Chemistry: [A. Caratti](#), F. Ferrero, E. Tabacco, F. Gerbaldo, E. Liberto, G. Borreani, C. Cordero, Translating 2D-Chromatographic Fingerprinting to Quantitative Volatilomics: Unrevealing Compositional Changes in Maize Silage Volatilome for Robust Marker Discovery, J. Agric. Food Chem. 72 (2024) 23616–23630. DOI: 10.1021/acs.jafc.4c05877.

11.1 Introduction

The use of bacterial inoculants to improve silage fermentation and aerobic stability has been widely studied, with the primary goal being to reduce aerobic deterioration during the feed-out phase.[1], [2] Among various lactic acid bacteria (LAB), heterofermentative *Lentilactobacillus buchneri* group strains are notably effective due to their ability to convert lactic acid into acetic acid and 1,2-propandiol, thus enhancing aerobic stability.[1], [3], [4] However, acetic acid alone does not fully explain yeast reduction over conservation and the improvements in aerobic stability .[3], [5], [6] Silage fermentation generates a complex array of volatile organic compounds (VOCs), including esters, aldehydes, alcohols, and sulfur derivatives, which have been linked to affect livestock intake and atmospheric VOC emissions.[7]–[9] Silage volatiles greatly influence animal performances psychogenic effects that can impact both on short- and long-term intake.[7] These effects would be mainly attributable to fermentation end products produced during silage conservation.[10]

Advanced analytical techniques as comprehensive two-dimensional gas chromatography (GC×GC) coupled with time-of-flight mass spectrometry (TOFMS) have significantly improved the understanding of silage VOCs.[11] This method offers superior separation and identification capabilities compared to traditional one-dimensional gas chromatography (1D-GC), making it suitable for exploring the complex composition of silage volatilomes.[12], [13] In a previous study¹⁰, GC×GC-TOFMS combined with automated headspace solid phase microextraction (HS-SPME) was used, for the first time, to detect the volatilome of maize silage inoculated or not with commercial strains of *Lentilactobacillus buchneri* and a new strain of *Lacticaseibacillus paracasei*. Of the 452 features (*i.e.*, volatile components) reliably aligned over the set of 2D chromatograms ($n=35$), 269 were putatively identified by combining linear retention index (I^T) in the first dimension (^1D) and electron ionization (EI) fragmentation spectra with authentic reference standards or database spectra (NIST similarity algorithm[14] direct match factor DMF > 900). Within the group of targeted VOCs, 72 were esters, 41 alcohols, 37 aldehydes, 31 aromatic derivatives, 25 terpenes/terpenoids, 22 ketones, and 12 were carboxylic acids resulting in a great improvement of the existing knowledge on silage VOCs composition.[7], [9], [15] By applying multivariate statistics in relation to known functional variables (*i.e.*, treatments and functional properties characterizing samples), a series of potential new markers responsible for aerobic stability were highlighted¹⁰, including aromatic esters (methyl benzoate and methyl salicylate) and alcohols (isoeugenol, guaiacol, 2,3-dimethylphenol) which are characterized by antifungal and antimicrobial activity.[16]–[18]

In the present study, we made a step ahead in the characterization of the silage volatilome by translating the methodology realized with GC×GC-TOFMS with a loop-type thermal modulator (TM) to a new platform, suitable for high sample throughput, equipped with a cryogen-free device realized with a reverse-inject differential-flow modulator (FM) and configured with a parallel detection by fast-scanning quadrupole MS (qMS) and flame ionization detection (FID).[19] The translated method on the FM GC×GC-qMS/FID preserves the original method resolution and elution order while enabling accurate quantification of analytes by FID-predicted relative response factors (RRF) without external calibration.[20]–[22] At the same time, thanks to a careful tuning of the chromatographic conditions, the chromatographic fingerprinting attitudes of the original method are preserved with the

possibility of keeping good discrimination and classification attitudes.[21] Moreover, the accurate quantification supports the discovery of potential markers by providing quantitative data that quickly indicates the actual impact of bio-active volatiles if they are present above a certain activity threshold. Current protocols for VOC monitoring are typically semi-quantitative or, even when they report data in concentration units, often use inconsistent quantification methods based on internal standardization.[15], [23], [24] The possibility of multi-VOCs quantitative profiling is attractive and could represent a turning point in the research field. VOCs are strongly connected to livestock welfare and meat and dairy product quality; a better understanding of the role of microbiota in silage, and of the effects of specific volatiles on animal intake and health, would support feeding strategies that positively impact on milk and cheese quality [2], [25], [26].

The sample set considered included maize silage samples obtained from an ensiling trial under different conditions (use of LAB inocula and delay of silage sealing) to stress the volatilome variability. The additional role of the microbial count of silages was included as a functional variable impacting the VOCs' dynamic range of quali-quantitative variation.

11.2 Materials and methods

11.2.1 Reference chemicals and solvents

Pure standards of *n*-alkanes (from *n*-C₉ to *n*-C₂₅) for Linear Retention Indices (*I*^T) calibration, α/β -thujone and 2-methyloctynoate used as Internal Standards (ISs), and solvents (cyclohexane and dibutyl phthalate – 99% of purity) were purchased from Merck (Milan, Italy).

The following marker volatiles adopted for external calibration and validation of the method accuracy were from Sigma Aldrich or Merck (Milan, Italy): (Z)-2-penten-1-ol (CAS 1576-95-0), 1-pentanol (CAS 71-41-0), 2-pentanol (CAS 6032-29-7), 2-heptanol (CAS 543-49-7), ethyl acetate (CAS 141-78-6), (*E*)-2-hexenal (CAS 6728-26-3), (Z)-3-hexen-1-ol (CAS 928-96-1), hexanal (CAS 66-25-1), octanal (CAS 124-13-0), nonanal (CAS 124-19-6), (*E*)-2-octenal (CAS 2548-87-0), linalool (CAS 78-70-6), phenol (CAS 108-95-2), and phenylethyl alcohol (CAS 60-12-8).

11.2.2 Reference solutions and calibration mixtures

Standard stock solutions of reference analytes were prepared at a concentration of 10 g L⁻¹ in cyclohexane and stored at -18°C for one-week.

ISs working solution for standard-in fiber pre-loading[27] was prepared at 0.100 g L⁻¹ in dibutyl phthalate and stored at -18°C in sealed vials.

The Reference Working Solution (RWS) for analytes subjected to external calibration was prepared by mixing suitable amounts of standard stock solutions to reach the concentration of 0.250 g L⁻¹ using cyclohexane as solvent.

Calibration Solutions (CS) for linearity assessment and accuracy verification of FID calibration vs. predicted RRFs were prepared by diluting suitable amounts of RWS in dibutyl phthalate. CSs were prepared to match the following absolute amounts: 1, 2.5, 5, 12.5, 25, 50,

75, 100, 125, 250, 300, 500 ng covering the actual dynamic range of concentrations for the target analytes of interest.

11.2.3 Fermented maize silage samples

Samples were collected in an experiment performed at the Experimental Centre of the University of Turin in the western Po plain, northern Italy (44°53'N, 7°41'E, altitude 232 m a.s.l.) on corn ensiled in laboratory silos conserved under standard conditions (Ferrero et al., 2021).[6] Corn was harvested with a precision forage harvester (Claas Jaguar 970 equipped with a New Holland 350W forage harvester head) as chopped whole crop to a theoretical chopping length of 12 mm, at around the 50% milk-line stage. The experiment was organized as randomized block design with five blocks to obtain replications per treatment (one in each block), and silos were prepared to be opened at 100 days (d) of anaerobic conservation. The factors evaluated were: a) delay of silo sealing, that were immediately sealed (NOT DELAYED), or sealed after 24 hours (DELAYED), to simulate aerobic stress which often characterize silage preparation and silo filling in commercial farms; b) inoculation or not with 2 commercial inocula containing *L. buchneri* group strains, that were MP1 (Magniva Platinum 1) and MPWC (Magniva Platinum whole crop - Lallemand Animal Nutrition, France) applied at the dose recommended on the label. The untreated and treated forages were then ensiled (about 10 kg of wet forage) in 20 L plastic silos equipped with a lid that only enabled gas release and conserved under controlled temperature ($20 \pm 1^\circ\text{C}$). The forages were packed by hand and final packing densities, on a wet basis, were $511 \pm 2.6 \text{ kg FM m}^{-3}$. The samples resulting for the dataset used in the study were 30. Analyzed samples together with the in-field trial coding and acronyms adopted in the manuscript are reported in the **Electronic Supplementary Material, Supplementary Table 1**.

At opening, each silo was weighed, and the content was mixed thoroughly and sub-sampled to determine the DM content, chemical composition, fermentation profile, microbial counts, and aerobic stability as reported in Ferrero *et al.*[6] After sampling, the silages were subjected to an aerobic stability test. Aerobic stability was determined by monitoring the temperature increases due to the microbial activity of the samples exposed to air. About three kilograms of each silo were allowed to aerobically deteriorate at room temperature ($20 \pm 1^\circ\text{C}$) in 17 L polystyrene boxes (290 mm diameter and 260 mm height), *i.e.* reaching a density of circa 350 kg FM/m^3 with half of the box filled. A single layer of aluminum cooking foil was placed over each box to prevent drying and dust contamination, but also to allow air penetration. The temperature of the room and the silage was measured each hour by a data logger. Aerobic stability was defined as the number of hours the silage remained stable before rising more than 2°C above room temperature.

11.2.4 Chemical and microbial analyses on silage samples

Each collected silage sample was split into five subsamples. One sub-sample was analyzed immediately for the DM content by oven drying at 80°C for 24 h. Dry matter was corrected according to Canale *et al.* (1983)[28], in order to consider the losses of volatile compounds that can take place at 80°C . The losses of volatile compounds were calculated as $\text{lactic acid} \times 0.106 + \text{acetic acid} \times 0.782 + \text{ethanol}$. A second fresh sub-sample was used to determinate the pH using a specific electrode (DL21 Titrator, Mettler Toledo, Milan, IT, pH accuracy $\pm$

0.002). A third sub-sample was extracted, using a Stomacher blender, for 4 min in H_2SO_4 0.05 mol L^{-1} at a 5:1 acid-to-sample material (fresh weight) ratio. An aliquot of 40 mL of silage acid extract was filtered with a 0.20- μm syringe filter and used for quantification of the fermentation products. The lactic and monocarboxylic acids (acetic, propionic, and butyric acids) were determined, by means of high-performance liquid chromatography (HPLC), in the acid extract. Ethanol and 1,2-propanediol were determined by means of HPLC, coupled to a refractive index detector, on a Aminex HPX-87H column (Bio-Rad Laboratories, Richmond, CA). The fourth subsample was used for the microbial analyses. For the microbial counts, 30 g of sample was transferred into a sterile homogenization bag, suspended 1:10 w/v in a peptone salt solution (1 g of bacteriological peptone and 9 g of sodium chloride per liter) and homogenized for 4 min in a laboratory Stomacher blender (Seward Ltd, London, UK). Serial dilutions were prepared and the mold and yeast numbers were determined using the pour plate technique with 40.0 g L^{-1} of Yeast Extract Glucose Chloramphenicol Agar (YGC agar, DIFCO, West Molesey, Surrey, UK) after incubation at 25°C for 3 and 5 days for yeast and mold, respectively. Yeast and mold colony forming units (cfu) were enumerated separately, according to their macromorphological features, on plates that yielded 1 to 100 cfu. The LAB were determined on MRS agar (Merck, Whitehouse Station, NY) with added natamycin (0.25 g/L), by incubating Petri plates at 30°C for 3 days under anaerobic conditions, according to Ferrero *et al.*[6] Since LAB are facultative anaerobe, the choice of anaerobic incubation was made to improve the selectivity of the media against *Bacillus spp.* Acetic acid bacteria (AAB) were counted following Spoelstra *et al.*[29] with minor modifications. The fifth subsample, used for volatilome analysis, was stored in a plastic, phthalate-free container, immediately frozen at -80 °C, and kept refrigerated until analysis.

11.2.4 VOCs sampling and analysis by HS-SPME-GC×GC-TOF MS and TM with the reference method

Headspace solid-phase microextraction: devices and conditions

The volatile organic compounds from the silage samples were sampled by means of HS-SPME using a SPR Auto sampler for GC (SepSolve-Analytical, Llantrisant, UK). A divinylbenzene/carboxen/polydimethyl siloxane (DVB/CAR/PDMS) fiber (d_f 50/30 μm ; 2 cm length) from Merck (Bellefonte, PA, USA) was used in accordance to previous studies.[11] The SPME fiber was conditioned before use, as recommended by the manufacturer. The ISs (α/β -thujone and methyl 2-octynoate) were preloaded onto the SPME device[27], [30] by sampling from a 20 mL headspace vial containing a 5.0 μL of the ISs solution (100 mg L^{-1}) prepared in dibutyl phthalate as a solvent. The ISs were equilibrated at 40 °C and the SPME device was exposed to the ISs HS for 5 min. ISs were used for validation purposes (method precision and repeatability) and to normalize the analytes' absolute responses (i.e., % *normalized response*).

Sampling was carried out on finely ground silage, 1.00 ± 0.03 g for the TM GC×GC-TOF MS method and on 0.100 ± 0.003 g for the FM GC×GC-MS/FID method, precisely weighed, placed in 20 mL headspace vials, and kept at 40°C for 50 minutes under constant agitation. The amount of the sample in the FM GC×GC-MS/FID method was reduced to a factor of 10 compared to the previous study to match for headspace linearity for a large set of volatiles.[31] In this condition, accurate quantification is possible by multiple headspace extraction even from solid samples where multiple equilibria (sorption and adsorption) might interfere with

analytes release on the headspace.[23] After extraction, the SPME device was automatically transferred to the split/splitless injection port of the GC×GC system, kept at 250 °C, and thermal desorption was then run for 5 min. The samples were analyzed in three replicates and randomly distributed over two weeks of measurements. Quality control samples (QCs) were prepared by mixing maize silage from a previous in-field trial; QCs consisted of 1.00 ± 0.03 g for the TM GC×GC-TOF MS method and on 0.100 ± 0.003 g for the FM GC×GC-MS/FID method of finely ground material subjected to the same sampling conditions applied to experimental samples. QCs (n=12) were analyzed each day to monitor system stability and response consistency.

TM GC×GC-TOF MS: instrument set-up and conditions

Comprehensive two-dimensional GC analyses were carried out with an Agilent 7890B GC chromatograph (Agilent Technologies, Wilmington, DE, USA) coupled with a Markes BenchTOF Select™ mass spectrometer featuring tandem ionization (Markes International, Llantrisant, UK). The GC transfer line was set at 270 °C. The TOF MS was tuned for single ionization at 70 eV and the scan range was set between 35–350 m/z with a spectrum acquisition frequency of 100 Hz. The thermal modulator was a loop-type, two-stage KT 2004 (Zoex Corporation, Houston, TX) cooled with liquid nitrogen and controlled by Optimode, v2.0 (SRA Instruments, Cernusco sul Naviglio, Milan, Italy). The modulation period (P_M) was set at 3.0 s, while the hot-jet pulse duration was set at 250 ms. The cold-jet stream at the mass flow controller (MFC) was programmed to linearly reduce the total flow (i.e., 20 L min⁻¹) from 40% to 8% along the analytical run.

The column set consisted of a ¹D HeavyWax™ column (100% polyethylene glycol – PEG; 30 m × 0.25 mm d_c × 0.25 μm d_f) coupled with a ²D OV1701 column (86% polydimethylsiloxane, 7% phenyl, 7% cyanopropyl; 1.0 m × 0.10 mm d_c × 0.10 μm d_f), both supplied by Agilent Technologies (Wilmington, DE, USA). A fused silica capillary loop (1.0 m × 0.1 mm d_c) was used in the modulator slit. SilTite™ μ-unions (Trajan Scientific and Medical, Melbourne, Australia) were used to connect the columns with the capillaries.

The GC split/splitless injector port was set at 250 °C and operated in split mode with a 1:20 split ratio. A special design liner for SPME thermal-desorption (Merck) was used to improve the transfer of the analytes to the ¹D column and to limit band broadening in-space. Helium was used as the carrier gas at a nominal flow of 1.3 mL/min. The oven temperature program was set as follows: from 50 °C (1 min) to 270 °C (5 min) at 4.0 °C min⁻¹.

The *n*-alkanes solution for I^T s determination was analyzed under the following conditions: split/splitless injector in split mode, 1:50 split ratio, 250 °C injector temperature, and 1 μL injection volume.

11.2.5 VOCs sampling and analysis by HS-SPME-GC×GC-qMS/FID and FM in translated conditions

Multiple headspace solid phase microextraction: devices and conditions

For quantitative volatilomics, silage samples were extracted by HS-SPME with a DVB/CAR/PDMS fiber (d_f 50/30 μm; 2 cm length) from Supelco installed on a multipurpose sampler, model MPS-2 (Gerstel, Mülheim a/d Ruhr, Germany). The ISs for response

normalization and quality control were preloaded onto the SPME device by sampling 5.0 μL of ISs solution (0.100 g L^{-1}) placed in a 20 mL headspace vial. ISTDs pre-loading was performed by exposing the SPME device in the HS at a temperature of $40\text{ }^{\circ}\text{C}$ for 5 min.

Sampling was carried out on $0.100 \pm 0.003\text{ g}$ of silage, precisely weighed, in 20 mL headspace vials, kept at 40°C for 50 minutes under constant stirring. The very low amount of sample was chosen to match HS linearity conditions for most of the characteristic volatile metabolites of the silage. After extraction, the SPME device was automatically transferred to the split/splitless injection port of the GC $\times$ GC system, kept at $250\text{ }^{\circ}\text{C}$, and thermal desorption was for 5 min. Samples were analyzed in two or three replicates randomly distributed over a two-week time frame.

MHS-SPME from samples and calibration solutions was conducted by applying the above indicated conditions, and the number of consecutive extraction steps was set to eight, achieving an almost exhaustive extraction for the analytes under study.[32] Multiple extractions were conducted on all calibration solutions/points and on a representative subset of samples to estimate the decay function (β) and validate its precision and consistency over the large dynamic range of concentrations of samples VOCs. Further details on calibration functions, quantification accuracy and method performance parameters are reported in the Results and Discussion section.

FM GC $\times$ GC-qMS/FID: instrument set-up and conditions

Automated MHS-SPME was performed by a multipurpose sampler, model MPS-2 (Gerstel), installed on a GC $\times$ GC system equipped with a reverse-inject differential-flow modulator based on capillary flow technology[™] (Agilent Technologies). The Agilent 7890B GC unit was coupled to an Agilent 5977B HES (high-efficiency source) fast quadrupole MS detector (Agilent Technologies) operating in electron ionization mode at 70 eV. Ion source and transfer-line temperatures were set at $280\text{ }^{\circ}\text{C}$, and the quadrupole temperature was set at 240°C . The scan range was set between 40 and 250 m/z , achieving an actual data acquisition frequency of 30 Hz. Parallel detection was by a fast FID with base temperature $280\text{ }^{\circ}\text{C}$; H_2 flow 40 mL/min, air flow 350 mL/min, and sampling frequency 200 Hz.

The column set was configured as: ^1D HeavyWax[™] column (100% polyethylene glycol - PEG; $20\text{ m} \times 0.18\text{ mm dc} \times 0.18\text{ }\mu\text{m df}$) coupled with ^2D OV1701 column (86% polydimethylsiloxane, 7% phenyl, 7% cyanopropyl; $1.8\text{ m} \times 0.18\text{ mm dc} \times 0.18\text{ }\mu\text{m df}$), both from Agilent Technologies. The connection between the ^2D column and deactivated silica capillaries toward MS ($0.5\text{ m} \times 0.1\text{ mm dc}$) and FID ($1.1\text{ m} \times 0.18\text{ mm dc}$) for parallel detection was by a three-way unpurged capillary microfluidic splitter (G3181B, Agilent). The resulting split ratio was 70:30 FID/MS. The bleeding capillary of deactivated fused silica ($6.06\text{ m} \times 0.1\text{ mm dc}$) installed on the modulator plate was dimensioned according to a previously validated flow calculator.[33]

The GC split/splitless injector port was set at $250\text{ }^{\circ}\text{C}$ and operated in split mode with a split ratio 1:20. The carrier gas was helium at a nominal flow of 0.4 mL min^{-1} along the ^1D column and 10 mL min^{-1} along the ^2D column. The oven temperature program was set: from $50\text{ }^{\circ}\text{C}$ (1.08 min) to 270°C (5.38') at $3.72\text{ }^{\circ}\text{C min}^{-1}$ and determined by method translation of the reference method previously validated in the authors' laboratory.[11], [34]

The workflow to translate chromatographic parameters is visualized in the **Supplementary Material** in **Supplementary Figure 1**. In practice, estimated operative pressures at the inlet (p_i) and outlet (p_{out} or p_{mid}) of the 1D column in the reference TM method are input in the calculator and used by the model to translate conditions for the FM method. For the FM configuration, the p_i and p_{out} or p_{aux} are set based on the *a priori* fixed flow conditions in both dimensions. The model calculates the oven temperature programming for the FM method by normalizing it according to the system void time (t_m).

The translated method modulation period (P_M) was defined after a scouting run and evaluating the 1D baseline peak-width (w_b) to obtain a comparable modulation ratio (M_R).[35] In translated conditions, P_M was set at 3 s and pulse time at 250 ms.

The *n*-alkanes liquid sample solution for I^T determination was analyzed under the following conditions: split/splitless injector in split mode, split ratio 1:50, injector temperature 250 °C, and injection volume 1 µL.

11.2.6 External standard calibration by MHS-SPME-GC×GC-qMS/FID and FID predicted RRFs

Calibration curves were built to cover analyte absolute amounts within the range 1-500 ng. External standard calibration was conducted on both MS and FID traces. For the MS detection channel, extracted target ions (Ti) were selected for each analyte; m/z values are reported in **Table 11.1**. Up to three additional qualifier ions (Q1, Q2, Q3) were also monitored for quality evaluation. For the FID channel, external calibration was based on 2D peak normalized volumes; analytes affected by co-elution issues were only quantified by MS. **Table 11.1** lists target analytes subjected to external calibration together with their experimental I^T , Ti and Qs m/z , regression functions parameters including calibration range, determination coefficients (R^2), precision, and limit of detection and quantification (LOD and LOQ).

Table 11.1: Calibrated analytes for accuracy validation. For the MS detection channel are reported target ions (Ti) and Qualifiers (Q#n) m/z values; calibration ranges, regression equation slope (m) and intercept (q) accompanied by the determination coefficient (R^2) and precision on response data collected over two-weeks of replicates. For the FID channel limit of detection and quantification (LOD and LOQ) were estimated according to the EU protocol for food safety analytical methods performance parameters by extrapolation from calibration curves.[36] Accuracy between FID and MS external calibration is expressed as % RE on calibration solutions (25 and 100 ng) and on a test sample; accuracy between predicted RRFs and FID external calibration is expressed as % RE on estimated amounts from the two approaches.

				MHS-SPME calibration MS				MHS-SPME calibration FID						Accuracy FID vs.MS			Predicted RRFs	
Compound name	Ti	Q#n	Range	Regression equation			Precision	Regression equation			Precision	LOD	LOQ	25 ng	100 ng	silage	Accuracy vs. FID	
	(m/z)	(m/z)	(ng)	m	q	R²	RSD%	m	q	R²	RSD%	(ng/g)	(ng/g)	%RE	%RE	%RE	RRF	%RE
1-Pentanol (IS)	70	87;55	1-5	1.35	0.03	0.995	2.40	0.21	0.02	0.999	4.20	0.39	1.30	98.2	97.7	95.2	1.00	-
			5-125	1.16	0.20	0.997		1.25	0.29	0.996								
2-Pentanol	73	87;55;45	1-100	1.61	0.79	0.997	7.00	0.98	0.68	0.996	9.70	0.16	0.53	105.2	115.2	108.3	1.00	94.9
(Z)-2-Penten-1-ol	68	86;57	1-125	1.60	0.09	0.997	2.60	1.27	0.32	0.996	6.30	0.29	0.97	111.2	109.1	115.2	1.05	111.3
(Z)-3-Hexen-1-ol	67	100;82;55	1-125	1.62	0.36	0.995	2.70	1.51	1.30	0.995	3.60	0.32	1.07	108.60	112.24	116.24	0.98	105.0
2-Heptanol	45	83;101	1-25	1.56	0.12	0.998	5.10	0.32	0.01	1.000	6.21	0.17	0.56	108.24	113.24	114.24	0.91	82.0
			25-250	1.69	0.38	0.998		0.32	0.16	0.996								
Linalool	71	93;121	5-50	1.02	0.27	0.997	3.46	0.33	0.10	0.997	7.12	0.24	0.81	115.24	117.24	114.97	0.88	88.1
			50-500	1.27	0.50	0.999		0.32	0.11	0.998								
Hexanal	72	82;56	1-10	0.85	1.00	0.995	9.50	1.34	1.44	0.998	9.80	0.33	1.10	107.24	108.24	112.04	0.98	114.6
			10-150	0.64	1.52	0.996		0.74	7.23	0.995								
(E)-2-Hexenal	83	98;69;55	1-5	0.82	0.00	0.995	6.20	0.12	0.01	1.000	4.30	0.32	1.07	108.24	110.26	118.24	1.02	115.2
			5-125	0.93	0.16	0.995		1.38	0.20	0.995								
Octanal	84	56;69;81	1-12.5	0.97	0.12	0.997	5.37	0.25	0.09	1.000	9.97	0.18	0.59	98.24	95.24	89.24	0.91	103.8
			12.5-125	1.02	0.65	0.999		0.16	0.22	0.999								
(E)-2-Octenal	83	97;70;55	1-125	1.00	2.03	0.995	9.10	1.26	1.94	0.996	7.80	0.32	1.07	85.27	89.24	95.24	0.93	105.0
Nonanal	57	82;96;98	5-50	1.65	0.37	0.996	5.24	0.20	0.34	1.000	12.48	0.38	1.27	95.24	97.26	87.27	0.88	97.7
			50-500	1.72	1.02	0.997		0.17	0.24	0.995								
Phenol	94	66	1-25	2.02	0.96	0.998	3.24	0.33	0.09	0.999	10.61	0.13	0.42	95.24	96.85	85.24	0.90	119.3
			25-250	2.60	1.24	0.999		0.21	0.44	0.999								
Phenylethyl alcohol	91	92;122	1-25	2.04	0.66	0.996	6.24	0.34	0.06	1.000	11.59	0.14	0.48	97.24	96.24	92.24	0.85	113.6
			25-250	3.09	0.79	0.997		0.21	0.45	0.999								
Ethyl acetate	70	88;61	1-5	0.26	0.81	0.996	9.20	0.10	0.03	1.000	8.90	0.47	1.57	118.26	115.24	116.24	1.57	102.0
			5-100	0.40	0.11	0.995		0.81	2.38	0.996								

11.2.7 Comprehensive chromatographic fingerprinting by image pattern recognition

The comprehensive mapping of the detectable volatilome was realized by the combined untargeted and targeted fingerprinting (*UT fingerprinting*) workflow based on image pattern recognition.[12], [37], [38] The algorithm tracks and aligns untargeted features (*i.e.*, peaks and peak-regions) across all available chromatograms by the principle of template matching.[39] Metadata corresponding to retention times (t_R and ^2t_R), MS fragmentation patterns, and detector responses are extracted for 2D peaks and peak-regions and used to establish correspondences across multiple chromatograms also enabling a temporal realignment in case of random fluctuations of retention times. To make the process highly specific and avoid false positive matches, spectral similarity above a DMF of 750 was used as a constraint together with retention times tolerances, parameters optimized in previous studies.[40] The UT feature reference spectrum used for the *feature template* was the average MS signature (“peak spectra”) from the largest datapoint within the 2D peak detected from all processed chromatograms.

The approach was applied to both the TM GC×GC-TOF MS and to the translated FM GC×GC-qMS/FID methods to cross-map the detectable volatilome and compare fingerprinting information potential. In this study, the process aligned the 102 acquired chromatograms [(30 samples × 3 replicates) + 12 Quality Controls] with the reference TM method and, in a separate workflow, the 76 chromatograms [(30 samples × 2 replicates) + 16 random replicates] acquired with the translated FM method. For each sample/data set, *reliable* peaks for registration were selected within those that positively matched *all-but-one* of the chromatograms (*i.e.*, the most-constrained condition option); the *composite* chromatogram from which the peak-region features were delineated was then built after chromatograms registration in the temporal domain.[41]

The last step enabled features targeting for the most informative compounds; in the TM reference method, they were 266 over 430 UT features while in the translated FM method they were 135 over 420 UT features. Putative identification was based on the 70 eV EI-spectra, according to spectral similarity criteria (DMF above 900 and RMF above 950) and an I^T tolerance of ± 10 units. The target analytes are listed in the **Electronic Supplementary Material, Supplementary Table 2** grouped by chemical class together with t_R , ^2t_R , experimental and library I^T s. **Supplementary Figures 2A and 2C** show the contour plots corresponding to the detectable volatilome of silage inoculated with MPWC (NOT DELAYED – MPWC) as resulting from the TM reference method (2A) or translated FM method (2C). Graphics in **Fig. 2B** and **Fig. 2D** corresponds to UT peak regions delineated in the feature templates for the most comprehensive UT fingerprinting (red areas for non-targeted analytes and light-blue areas for targeted analytes).

11.2.8 Data acquisition, 2D data processing and statistical analysis

Raw chromatographic data were acquired by TOF-DS software (Markes International, Llantrisant, UK) in the TM platform and by MassHunter Workstation (Agilent Technologies) in the FM platform. Raw data were processed by GC Image™ V2020 r1.2 suite (GC Image, LLC Lincoln, NE, USA).

Statistical analysis and chemometrics were performed using GC Investigator™ (GC Image), XLSTAT statistical and data analysis solution (Addinsoft 2020, New York, USA), and Microsoft Office Excel 2016 (Microsoft, Redmond, WA USA).

11.3 Results and discussion

11.3.1 Main fermentative and microbial characteristics of the silages

Silages are representative of the corn silage grown in the Po plain (Italy). The nutritional characteristics of herbage prior to ensiling were in the range of corn harvested at the half milk-line stage, with NDF content of 391 g kg⁻¹ DM ($\pm$ 25), starch content of 285 g kg⁻¹ DM ($\pm$ 29), crude protein content of 81 g kg⁻¹ DM ($\pm$ 2.1), and water soluble carbohydrate content of 92 g kg⁻¹DM ($\pm$ 2.4). The DM of silages was on average 33.3% $\pm$ 0.02. At silo opening all the silages were well fermented with butyric acid content below the detection limit (< 0.1 g kg⁻¹DM). On average silages showed 41.6 $\pm$ 3.8, 10.6 $\pm$ 2.3 and 8.7 $\pm$ 3.2 g kg⁻¹DM of lactic acid, acetic acid, and ethanol, respectively. The pH was on average 3.74, ranging from 3.69 to 3.86. Silages at opening showed a LAB count of 9.07 $\pm$ 0.85 log cfu g⁻¹ FM. Acetic acid bacteria count was on average 5.72 $\pm$ 0.75 log cfu g⁻¹ FM. Yeast count ranged from 1.30 to 6.22 log cfu g⁻¹ FM. Aerobic stability values were highly variable with a range from 39 to 205 h. The delay sealing affected the variability of measured microbial and chemical characteristics, with a higher variability in DELAYED than in NOT DELAYED silages. The high variability in yeast count and aerobic stability was significantly affected by the different LAB treatments (data not shown).

11.3.2 Volatilome compositional diversity and the impact of functional variables

The volatilome of maize silage results from multiple chemical reactions driven by endogenous (plant origin) or exogenous enzymes from both epiphytic and inoculated microorganisms. Silage volatile organic compounds can have a negative, neutral, or sometimes positive influence on silage preference and intake in ruminants.[10] The complex silo environment and the concurrent formation of different fermentation products, may strongly influence intake and consequently animal performances. Several researchers suggested that alcohols (ethanol and propanol) and esters (ethyl acetate and ethyl lactate) can reduce preference and intake of silage.[7], [8], [42] Other VOCs, such as butyric acid, methanol, iso-butyric acid, and iso-valeric acid, seem to have no negative influence on preference or DMI.[10], [42]

These reactions are influenced by environmental changes during fermentation, including anaerobiosis, temperature increases (from 23 to 35°C) in the early days of ensiling (especially in DELAYED silages), and acidification (pH dropping from around 6.0 to between 3.6 and 3.9).

The UT fingerprinting procedure on the detectable volatilome of samples analyzed with the reference methodology (*i.e.*, TM GCxGC-TOF MS) enabled the coverage of 430 peak region features, of them 266 were putatively identified according to their ¹D *I*^T and EI-MS fragmentation pattern. Chemical classes include esters (n=82), alcohols (n=43), aldehydes and ketones (n=33), aromatic derivatives with various functionalities (n=33), hydrocarbons (n=30),

specialized volatile metabolites as terpenes and nor-isoprenoids (n=26), and short chain fatty acids (n=11). In keeping with previous studies, the VOCs composition confirmed the extensive metabolic activity of epiphytic and inoculated microorganisms (homolactic- and heterolactic acid bacteria, and yeasts) accompanied by the plant metabolism reacting to the abiotic and biotic stress factors.[6], [11], [43]

The high relative amounts of short-chain free fatty acids and alcohols, along with temperature increase in the early days of ensiling and acidic conditions increase during fermentation, led to the formation of many esters, which resulted to be the chemical class mostly represented within the detectable volatilome. Additionally, oxidative degradation of lignocellulosic fibers over fermentation increased certain aromatic compounds with antimicrobial properties that could be connected to silage aerobic stability[11] or silage intake by ruminants.[10]

A first unsupervised exploration was conducted by principal component analysis (PCA) on all samples by considering the 430 UT features and their % responses collected from the peak regions included in the feature template. As shown in **Figure 11.1A**, the natural clustering of samples indicates a major impact of the ensiling delay (24 h) on the samples' volatilome. The cluster of DELAYED samples (blue indicators), which includes controls (CONTR) and LAB-treated samples (MP1 and MPWC), has a larger internal variability. The confidence ellipse, set at a 95% confidence level, suggests and confirms that the delay in establishing anaerobiosis at ensiling promotes the production of many volatile metabolites through many concurrent oxidative phenomena (principally plant tissue respiration and microbial oxidative activities). Aerobic respiration, conducted by hexokinase, phosphofructokinase, pyruvate kinase, and other enzymes of the glycolytic pathway and citric acid cycle amplifies the extent of oxidation by producing heat, O₂, and CO₂. [2] At the same time, the activation of lipoxygenases, which catalyze the oxidation of polyunsaturated fatty acids (PUFAs) such as linoleic acid and linolenic acid, generate non-volatile hydroperoxy derivatives that can be converted by hydroperoxy liases or autocatalytic reactions (α - or β -scission) into volatile carbonyls.[44] Moreover, the cross-talk between different metabolic pathways induced by oxidative stress trigger the formation of salicylic acid (SA) and abscisic acid (AC) both signaling molecules that enhance plant resilience.[45]

At the same time, the ensiling delay enables the development of homofermentative and heterofermentative epiphytic lactic acid bacteria (e.g., *Lactiplantibacillus plantarum*, *Pediococcus* spp.) that activate lactic dehydrogenase, aldolase, and enzymes of the pentose phosphate pathway releasing lactic acid and acetic acid as major products. The presence of Enterobacteria (e.g., *Enterobacter* spp.) adds further complexity with the formation of acetic acid and other alcohols.

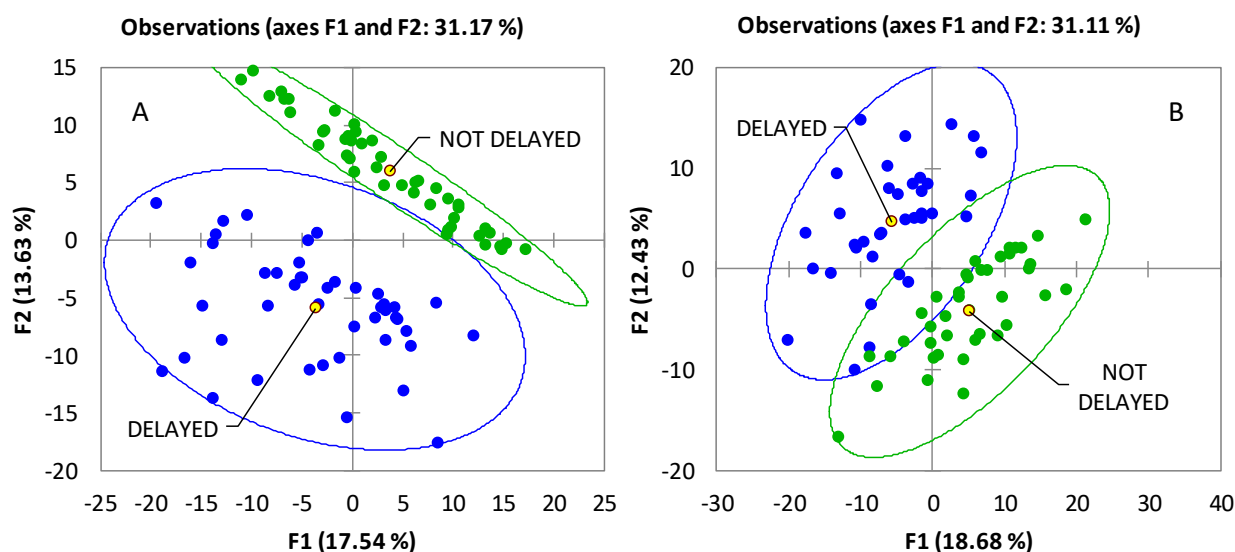


Figure 11.1: Principal component analysis (PCA) scores plot based on % relative responses from UT components as resulting from the TM GC×GC-TOF MS reference methodology (**1A**) and from the translated method by FM GC×GC-qMS/FID (**1B**). UT features computed are 430 for the reference method and 420 for the translated method; the number of samples including replicates are 90 and 76, respectively. Natural samples groups almost independently clustered, as shown by confidence ellipses (95 % of confidence level), correspond to CONTR and LAB treated silage samples subjected or not to ensiling delay of 24 hours (DELAYED blue circles, NOT DELAYED green circles).

Furthermore, extensive oxidation accompanied by the decrease of the pH and increase in temperature have an impact on phenolic specialized metabolites; they are natural substates of polyphenol oxidase(s) and are prone to chemical oxidation, due to their redox potential vs. that of many other substrates (oxylipins, tocopheroxyl radicals, etc.).

All the abovementioned phenomena may had an impact on the DELAYED sample volatile metabolome. By supervised investigation with partial least-squares discriminant analysis (PLS-DA) on the classes DELAYED vs NON-DELAYED the most informative analytes were identified. The model was developed on an estimation set of samples consisting of 70% randomly selected measures/samples (63 samples over 90) and validated on the residual 30% of samples (37 samples over 90). Model performances, after 10 reiterations, were good referring of a 100% of correctness. Discriminant volatile metabolites, based on the variable importance in the projections (VIPs) scores, resulted the series of C6 derivatives (listed in descending order of relevance – VIPs > 1.5: (E)-3-hexen-1-ol, (E)-2-hexenal, (Z)-2-hexen-1-ol, (E)-2-hexen-1-ol, (Z)-3-hexen-1-ol, 1-hexanol, (Z)-3-hexen-1-ol acetate); C5 derivatives [(Z)-2-penten-1-ol, (E)-2-pentenal]; and other characteristic products of LOX pathways [(E,Z)-2,4-heptadienal and (E,E)-2,4-heptadienal]. The signature of heterolactic fermentation from different microorganisms (both epiphytic and inoculated) on the DELAYED samples was confirmed by the pattern of some alcohols, acids, and related esters like 2-butanol, ethyl propanoate, isobutyl acetate, 1-propanol, propyl acetate, 1,3-butanediol, propyl lactate, isoamyl propionate (3-methyl-1-butanol propanoate), ethyl-2-methyl butanoate, and 3-methyl butanoic acid listed in descending order of VIPs values. Specialized metabolites were also impacted, and within this class some oxidized terpene derivatives like linalool oxide II or

nor-isoprenoids (β -damascenone), and phenol derivatives like 2-methoxy phenol, phenylacetaldehyde, 4-ethyl guaiacol, *p*-cresol, and benzaldehyde resulted discriminat.

A further step of the investigation was oriented to highlight the chemical signature of LAB inocula in samples immediately ensiled (NOT DELAYED). The ensiling delay with herbage exposure to ambient air is a non-recommended practice due to reduction of fermentative carbohydrates, increase in temperature, and protein degradation processes.[46] PLS-DA was adopted to discriminate control samples (CONTR) vs. LAB fermented silage (MP1 and MPWC) and unveiled those VOCs characterizing the LAB fermentation on primary and specialized metabolites in the NOT DELAYED group. The model was developed on an estimation set of samples consisting of 70% randomly selected measures/samples (32 samples over 45) and validated on the residual 30% of samples (13 samples over 45). Model performances, after 10 reiterations, were quite good referring of a 70% of correctness accompanied to a 100% specificity and lower sensitivity. The most discriminant variables, *i.e.*, those with VIPs > 1.5 ($\pm$ SD) were 1,2-propandiol-2-acetate, 1,2-propandiol, acetoin, 2-phenylethyl acetate, phenylacetaldehyde, phenylethyl alcohol, 3-methyl-1-butanol acetate, isoamyl propionate, and benzyl alcohol to cite those in common to both commercial LAB species. The presence of many acetates within the most discriminant VOCs confirms the impact of heterolactic fermentation acted by bacteria of *Lentilactobacillus* genus.[47]

11.3.3 Method translation from TM to FM GC $\times$ GC: information capacity and targeted analytes coverage

The extensive chemical complexity of fermented silage VOCs, investigated using GC $\times$ GC-TOF MS, enhances our understanding of the metabolic pathways activated by specific bacterial strains and provides additional insights into the mechanisms of silage stabilization against degradation by yeasts and molds when exposed to air.[1], [6], [11], [47] The approach, by integrating established chemical characterization protocols with extensive/comprehensive VOCs monitoring, has allowed for robust cross-validation of data. However, the VOCs fingerprinting using HS-SPME sampling if not appropriately conducted is not inherently quantitative,[23], [24] therefore although capable of quickly revealing pattern variations across a wide range of concentrations providing evidence of metabolic pathway activation and synergies between bacterial strains, does not support causal yet robust correlation with specific inhibitors or bio-active chemicals responsible of silage stabilization and quality.

This study makes a step ahead in the direction of designing an accurate quantitative approach for extensive quantification of VOCs from silage samples. The transition from a non-quantitative fingerprinting to an accurate quantitative profiling method was realized by integrating into the analytical platform a parallel detection with a FID. Moreover, by choosing a different modulation technology (*i.e.*, differential flow modulation instead of cryogenic/thermal modulation) the analytical platform results cost effective and more robust in case of high sample throughput.[34], [48] Modulator technology, based on capillary microfluidics, requires high-flows in the ^{2} D to produce a suitable band-compression in time. With the careful tuning of chromatographic parameters and the correct translation of the temperature programming, the separation power is almost preserved[19], [49] and ^{2} D post-column splitting into two parallel detectors (*e.g.*, MS and FID) opens to additional

opportunities for accurate quantification and cross-validation. With the FID detection, thanks to its linear response function over many orders of magnitude, analyte quantification in a wide range of concentrations is feasible.[50] Moreover, even without external calibration, FID allows accurate quantification of analytes through response factors (experimentally estimated or predicted based on combustion enthalpies and molecular formulae[51]) in the presence of suitable internal standards added in known amounts to the samples or in the presence of analytes quantified by external calibration.

The translation of the methodology from the reference platform (*i.e.*, TM GC×GC-TOF MS) to the new one realized with FM GC×GC-qMS/FID, followed a workflow described in previous research.[34] In practice, translatable parameters of the reference method were set to preserve ¹D peak elution order, ¹D peak capacity, and chromatographic resolution in the new methodology. By applying the principles of method translation conceptualized for fast GC,[52] temperature programming was modified according to the estimated speed gain and corresponding to the ratio between ¹D column void times (t_M). The operation was supported by the method translation software, available as a free application on the web.[53] The resulting parameters are those reported in the Experimental section. With them, the system peak capacity was reduced to about 80% of the reference method,[19] sensitivity to 60-70%, [19] and information capacity toward confident identification of targeted analytes of 51%. The latter is because of the parallel detection by qMS and FID tuned to a ratio of 3:7 in order to accommodate the high-flow rate in the ²D column necessary for the proper differential-flow modulation dynamics and compensate for the different absolute sensitivity of the two detectors.

The FM GC×GC-qMS/FID 2D patterns corresponding to the 76 samples (including all replicates), when analyzed by UT fingerprinting resulted in 420 UT features included in the feature template and 135 targeted analytes (*i.e.*, targeted features). The targeted analytes are indicated in **Supplementary Table 11.2**. Besides the reduced performances of the translated methodology, the information capacity of the new method was preserved; the PCA loading plot corresponding to the FID % response for all 420 UT features mapped by FM GC×GC is shown in **Figure 11.1B**. The samples at different ensiling times (DELAYED and NOT DELAYED) are almost independently clustered as was observed with the reference method (**Figure 11.1A**). Both approaches result in a fingerprinting capacity with a total explained variance on the first two components (F1 and F2) of about 31%. In the case of the FM GC×GC (**Figure 11.1B**), the cluster of NOT DELAYED samples is connoted by a higher internal variance, if compared to the TM methodology, likely due to the different detector signal exploited (*i.e.*, total ion current - TIC % response for the TM GC×GC-TOF MS and the FID % response for the FM GC×GC) and its peculiar response function.

To further validate the translated method effectiveness in exploring the silage VOCs patterns, PLS-DA was applied to highlight discriminant analytes between the classes of DELAYED vs. NOT DELAYED samples (n=76). The model was developed on an estimation set of samples consisting of 70% randomly selected measures/samples (53 samples over 76) and validated on the residual 30% of samples (23 samples over 76). Model performances, after 10 reiterations, were good referring of a 100% of correctness. VIPs were computed and those with the highest ranking on the Component 1 (VIPs > 1.5) were 2-butanol, 1-propanol, (*E*)-3-hexen-1-ol, ethyl

propanoate, isobutyl acetate, isoamyl propionate, (*E*)-2-hexen-1-ol, (*E*)-2-hexenal, (*Z*)-3-hexen-1-ol, and (*Z*)-2-penten-1-ol. All of them were also targeted by the reference methodology.

11.3.4 Quantitative volatilomics by FM GC×GC-qMS/FID: accuracy validation and predicted FID relative response factors

To accurately quantify multiple volatile targets in a single run over a wide range of analyte concentrations, the multiple headspace extraction (MHE) method is recommended when using high-efficiency, standard multi-component SPME devices in non-equilibrium sampling. Originally introduced by Suzuki *et al.*[54] and McAuliffe[55] for static headspace sampling, MHE measures the area representative of the total amount of the target analyte in the matrix. Kolb and Ettre[56] further developed and modeled MHE, which was later successfully applied to HS-SPME and referred to as MHS-SPME. The principles and fundamental equations of MHE are well-documented in the literature.[57]–[60] In the **Electronic Supplementary Material, Supplementary Figures 3 to 5** visual descriptions of the operative steps are reported.

To strengthen these observations and establish robust correlations, eventually corroborated by quantitative data on absolute concentrations, the quantification was conducted by MHS-SPME on all analytes that, under the sampling conditions (*i.e.*, amount of sample, HS volume, temperature and time), were within the HS linearity interval. To ascertain this, multiple extractions (up to eight) were conducted on a sub-set of selected samples from the different biological replicates of CONTR, MP1 and MPWC to cover the largest possible dynamic range of concentrations. For those analytes, where it was possible, to observe an exponential decrease of the normalized response [*i.e.*, FID trace 2D peak volume normalized over the IS response (A_S)] along with the successive extraction steps, a decay function was estimated (β). In each consecutive sampling step, the partition constant (K_{HS}) between the condensed phase and the HS remains constant, therefore, the sum of the normalized responses A_S from each consecutive extraction step corresponds to the total area or total instrumental response (A_T) for the analyte originally present in the condensed phase (*i.e.*, the sample), which can be calculated from **Equation 11.1**:

Equation 11.1.
$$A_T = \sum_{i=1}^{\infty} A_i = A_1 \left(\frac{1}{1 - e^{-q}} \right) = \frac{A_1}{1 - \beta}$$

where A_T is the total estimated area, A_1 is the area resulting from the first extraction, and q is a constant describing the exponential decay of the area with successive extractions. The term q can be determined by plotting the natural logarithm of chromatographic areas against the number of extraction steps. The β coefficient, which depends on the analyte, remains constant in samples with similar matrix effects. Representing the rate of decay across successive extractions, β intrinsically verifies HS linearity when it is less than or equal to 0.95.

Table 11.1 provides detailed information for fourteen targeted volatiles used to verify method's accuracy and define its relative uncertainty. Calibrated analytes are listed together with calibration ranges (expressed in ng, as the absolute amount of analyte in the HS), calibration functions for both MS and FID channels, determination coefficients (R^2), and intermediate precision expressed as the characteristic % relative standard deviation (RSD%)

from repeated quantitative measurements of a representative sample over two weeks. The choice of external standards, was oriented to major chemical classes represented in silage volatilome (*i.e.*, carbonyls, primary alcohols, aromatics, esters, and terpene alcohols) while covering different volatilities and polarities. About the concentration range spanned by calibration points, it was established based on preliminary experiments conducted to evaluate the HS linearity by applying predicted RRFs for amount estimation. 1-Pentanol, ethyl acetate, (*E*)-2-hexenal, hexanal, linalool, phenol, phenylethyl alcohol, octanal, nonanal, and 2-heptanol required a two-step calibration to achieve linearity on the MS channel. **Table 1** also includes the limit of detection (LOD) and limit of quantitation (LOQ) for the FID channel, estimated according to the EU protocol for food safety analytical methods performance parameters by extrapolation from calibration curves.[36]

Due to the absence of certified reference materials (CRMs), quantitation accuracy was validated through internal cross-matching of MS and FID results from selected calibration points (12.5 and 25 ng). Results are expressed as relative error % (% RE) calculated according to **Equation 11.2** and reported in **Table 11.1**.

Equation 11.2.
$$RE\% = \frac{\text{Estimated amount} - \text{Expected amount}}{\text{Expected amount}} * 100$$

For test samples (silage), accuracy was verified by comparing the FID estimated amount vs. the one obtained from the MS channel due to the higher specificity of this latter which is capable of isolating analytes response to interferences. Accuracy results, reported in **Table 1** confirm a quite good performance when external calibration is performed by MHS-SPME, RE% never exceed the $\pm 30\%$ error tolerance.[61]

With parallel MS and FID detection, additional options for reliable quantitation are available. FID provides stable structure-specific response factors in a wider range of concentrations.[20], [62] In particular, the FM GC $\times$ GC-qMS/FID opens the possibility of adopting FID predicted RRFs based on combustion enthalpies and molecular structure.[20] The alignment of the separation patterns from MS and FID detection enables reliable analyte identification using multiple criteria (retention in two dimensions and MS spectral signature) and across the parallel channels, while quantitation can be extended over uncalibrated compounds by estimating their FID RRFs with **Equation 11.3**.

Equation 11.3.
$$RRF = 10^3 * \left(\frac{MW_i}{MW_{IS}} \right) * (-61.3 + 88.8\eta_C + 18.7\eta_H - 41.3\eta_O + 6.4\eta_N + 64.0\eta_S - 20.2\eta_F - 23.5\eta_{Cl} - 10.2\eta_{Br} - 1.75\eta_I + 127\eta_{benz})^{-1}$$

where η_C , η_H , η_O , η_N , η_S , η_F , η_{Cl} , η_{Br} , η_I , and η_{benz} corresponds to the number of carbon, hydrogen, oxygen, nitrogen, sulfur, fluorine, chlorine, bromine, and iodine atoms, and the number of benzene rings; and MW_i and MW_{IS} are the molecular weights of the analyte *i* and the *IS* adopted for the development of the model by de Saint Laumer *et al.*[20] In this study, quantitation was obtained on the normalized 2D response of each component normalized versus α -thujone as *IS* and corrected with its RRF calculated with **Eq.11.3** to align combustion enthalpies normalized versus methyl octanoate. Validation of the RRFs accuracy was confirmed for calibrated analytes by estimating the %RE on experimental FID calibration vs.

RRFs. Results are reported in **Table 11.1** and confirm a good accuracy, never exceeding the $\pm 20\%$ error tolerance.

Table 11.2 lists the 98 quantified analytes, all within the HS linearity in the sampling conditions, together with their molecular formulae, decay constant β and its variation (%RSD) over 15 different samples tested for multiple extractions, and predicted RRFs estimated over 1-pentanol as IS for quantification. The asterisk (*) highlights those analytes with a non-optimal β value and for which a larger inaccuracy was expected. Precision data on accurate amounts was estimated at intra-week level (repeatability RSD_r %) and at inter-week level (intermediate precision RSD_R %) on QC samples.[63] The adoption of a single IS (*i.e.*, 1-pentanol) for normalizing the predicted RFs was decided based on the accuracy results and because of the excellent linearity of this analyte shown within the calibration range.

11.3.5 Quantitative volatilomics by FM GC×GC-qMS/FID: results

Accurate amounts expressed in ng g⁻¹ of the selected 98 analytes estimated on the subset of NOT DELAYED samples are reported in **Supplementary Table 3** as the mean of replicated analyses (N=2 or 3) accompanied by the relative uncertainty. Of the quantified analytes, ethyl acetate and benzaldehyde were the most abundant, ranging between 17 and 13 µg g⁻¹ respectively. The first reflects the dominant biotransformations resulting in the formation of ethanol and acetic acid. Benzaldehyde on the other hand might result from a complex process involving the interplay between plant-derived precursors (phenylalanine and *trans*-cinnamic acid) and microbial metabolic activities during fermentation conducted by phenolic acid decarboxylase (PAD) and aldehyde dehydrogenase (ADH). In the range between 10-1 µg g⁻¹ there are several oxidized derivatives of plant phenols like 4-ethyl-2-methoxy-phenol, 4-ethyl-phenol, phenylacetaldehyde, phenylethyl alcohol, benzyl alcohol, and benzoic acid accompanied by a series of short-chain fatty acids like butanoic acid, 2-methyl propanoic acid, 3-methyl butanoic acid, 2-methyl pentanoic acid, and hexanoic acid. Alcohols detected in higher amounts (over the quantified analytes) are 1-propanol, 2-methyl-1-propanol, 3-methyl-1-butanol, 2-pentanol, (Z)-3-hexen-1-ol, and 1-octen-3-ol. In the range between 1 µg g⁻¹ – 100 ng g⁻¹ there is the majority of quantified VOCs including also those discriminants for the ensiling delay and LAB treatment. These quantitative results, for the first time determined with a truly accurate approach to an extended set of volatile metabolites, add further knowledge on silage fermentation while indicating new potential markers causally correlated to aerobic stability or other silage quality parameters. By considering the pattern of quantified analytes, the discriminant analytes were identified by the Kruskal–Wallis test combined with Dunn's test (*p*) and Bonferroni correction (*§p*) for nonparametric data distribution (significance level alpha=0.05). Analytes with meaningful variations between the CONTR samples and LAB fermented samples are indicated in **Table 11.2** and **Supplementary Table 11.3** together with *p* values. Differential distribution is shown in box-plots of **Figure 11.2**; analytes and other quality parameters are grouped according to the trend (ascending or decreasing) among sample classes. Although multivariate by nature, the fermentation phenomenon induces down-regulation of some volatiles. Phenylethyl alcohol and the corresponding acetate, propyl acetate, acetoin, 2-methyl propanoic acid, and (Z)-3-hexen-1-ol have all a decreasing trend as the count of acetic acid bacteria, yeasts, and pH. On the contrary, higher log cfu g⁻¹ counts of LABs in MP1 and MPWC classes are correlated with

higher aerobic stability, as expected, and higher amounts of benzaldehyde, phenylmethyl acetate, 2-butanol, 2-pentanol, (*E*)-2-hexen-1-ol, (*E*)-3-hexen-1-ol, (*E*)-2-hexenal, butanoic acid and phenylmethyl acetate.

Table 11.2: Analytes subjected to accurate quantification together with chemical formula, FID predicted RRF normalized over 1-pentanol as IS, decay function (β) estimated on multiple consecutive extractions on 15 different samples randomly selected within the set accompanied by the % relative standard deviation. The asterisk (*) highlights analytes with a decay constant above 0.95 and for which it is expected a larger inaccuracy in quantitative estimations. Precision data refers to intra-week (RSD_r % - repeatability) and intra-week (RSD_R % - intermediate precision) quantification results on Quality Control samples (QCs).

Compound name	Chemical class	Formula	RRF	β	RSD % on β (n=15)	RSD _r % (intra-week)	RSD _R % (inter-week)
(E)-3-Hexenoic acid	Acids	C ₈ H ₁₆ O ₂	1.27	0.86	6.04	14.9	15.6
Benzoic acid	Acids	C ₇ H ₆ O ₂	1.09	0.93	3.61	15.7	18.3
Butanoic acid	Acids	C ₄ H ₈ O ₂	1.57	0.95	1.37	11.6	14.3
3-Methyl butanoic acid	Acids	C ₅ H ₁₀ O ₂	1.35	0.93	2.02	13.7	15.5
Hexanoic acid	Acids	C ₆ H ₁₂ O ₂	1.22	0.89	4.68	12.0	13.4
Octanoic acid	Acids	C ₈ H ₁₆ O ₂	1.07	0.93	2.38	11.9	14.1
Pentanoic acid	Acids	C ₅ H ₁₀ O ₂	1.35	0.90	2.62	12.4	15.0
2-Methyl pentanoic acid	Acids	C ₆ H ₁₂ O ₂	1.22	0.97*	2.67	14.1	14.6
2-Methyl propanoic acid	Acids	C ₄ H ₈ O ₂	1.57	0.98*	0.17	12.0	15.4
1-Pentanol	Alcohols	C ₅ H ₁₂ O	IS	0.87	4.63	11.9	13.2
3-Methyl-1-butanol	Alcohols	C ₅ H ₁₂ O	1.00	0.88	5.16	11.7	11.3
2-Ethyl-1-hexanol	Alcohols	C ₈ H ₁₈ O	0.89	0.90	3.92	17.0	16.0
1-Octanol	Alcohols	C ₈ H ₁₈ O	0.89	0.87	6.94	16.7	16.3
1-Octen-3-ol	Alcohols	C ₈ H ₁₆ O	0.91	0.89	4.87	8.0	9.8
1-Propanol	Alcohols	C ₃ H ₈ O	1.23	0.93	3.01	15.4	18.4
2-Methyl-1-propanol	Alcohols	C ₄ H ₁₀ O	1.08	0.94	1.42	13.9	11.6
(E,E)-2,4-Heptadienal	Alcohols	C ₇ H ₁₀ O	1.00	0.89	4.05	9.8	10.5
2-Butanol	Alcohols	C ₄ H ₁₀ O	1.08	0.92	6.75	6.2	8.9
2-Furanmethanol	Alcohols	C ₅ H ₆ O ₂	1.53	0.94	2.09	7.2	7.9
2-Heptanol	Alcohols	C ₇ H ₁₆ O	0.91	0.77	6.79	18.1	15.4
(E)-2-Hexen-1-ol	Alcohols	C ₆ H ₁₂ O	0.98	0.79	1.84	9.7	7.8
2-Pentanol	Alcohols	C ₅ H ₁₂ O	1.00	0.90	5.89	9.8	9.6
(Z)-2-Penten-1-ol	Alcohols	C ₅ H ₁₀ O	1.05	0.90	4.06	11.1	11.8
1-(2-methoxy-1-methylethoxy)-2-propanol	Alcohols	C ₇ H ₁₆ O ₃	1.29	0.91	3.20	12.0	12.3
(E)-3-Hexen-1-ol	Alcohols	C ₆ H ₁₂ O	0.98	0.81	2.41	7.8	9.4
(Z)-3-Hexen-1-ol	Alcohols	C ₆ H ₁₂ O	0.98	0.69	4.62	9.7	9.5
3-Octanol	Alcohols	C ₈ H ₁₈ O	0.89	0.71	10.98	9.3	12.7
3-Pentanol	Alcohols	C ₅ H ₁₂ O	1.00	0.88	1.49	13.4	13.8
2-Methyl-3-pentanol	Alcohols	C ₆ H ₁₄ O	0.95	0.89	4.26	9.0	12.3
4-Heptanol	Alcohols	C ₇ H ₁₆ O	0.91	0.73	3.73	7.8	12.3
4-Nonanol	Alcohols	C ₉ H ₂₀ O	0.86	0.82	6.40	8.9	9.8
α -Phenylethyl alcohol	Alcohols	C ₈ H ₁₀ O	0.85	0.91	2.12	10.8	12.5
Benzyl alcohol	Alcohols	C ₇ H ₈ O	0.87	0.83	6.35	16.0	16.8
Butoxyethoxyethanol	Alcohols	C ₈ H ₁₈ O ₃	1.21	0.90	6.92	13.2	16.4
6-Methyl-5-hepten-2-ol	Alcohols	C ₈ H ₁₆ O	0.91	0.86	5.44	12.5	13.6
Phenylethyl alcohol	Alcohols	C ₈ H ₁₀ O	0.85	0.91	1.04	7.9	13.2
4-Methyl-decane	Alkanes	C ₁₁ H ₂₄	0.74	0.74	6.87	14.9	17.5

Table 11.2: continue

Compound name	Chemical class	Formula	RRF	β	RSD % on β (n=15)	RSDr % (intra-week)	RSDR % (inter-week)
Decane	Alkanes	C ₁₀ H ₂₂	0.74	0.85	3.37	17.1	15.0
2,4-Di-tert-butylphenol	Aromatics	C ₁₄ H ₂₂ O	0.79	0.93	4.15	17.4	18.8
1,3-Di-tert-butylbenzene	Aromatics	C ₁₄ H ₂₂	0.71	0.88	5.23	15.8	15.4
Cresol	Aromatics	C ₈ H ₁₀ O ₂	1.01	0.88	4.40	16.3	16.4
1-Hydroxy-2-acetyl-4-methylbenzene	Aromatics	C ₉ H ₁₀ O ₂	0.99	0.95	0.95	12.3	17.4
o-Xylene	Aromatics	C ₈ H ₁₀	0.71	0.90	5.07	15.6	15.0
p-Cresol	Aromatics	C ₇ H ₈ O	0.87	0.98	0.57	11.4	16.4
Phenol	Aromatics	C ₆ H ₆ O	0.90	0.93	3.13	20.9	16.8
2-Methoxy-phenol	Aromatics	C ₇ H ₈ O ₂	1.06	0.93	2.67	12.1	15.9
4-Ethyl-phenol	Aromatics	C ₈ H ₁₀ O	0.85	0.93	2.98	14.9	18.6
4-Ethyl-2-methoxy-phenol	Aromatics	C ₉ H ₁₂ O ₂	0.97	0.91	3.13	8.5	7.3
p-Xylene	Aromatics	C ₈ H ₁₀	0.71	0.77	8.39	15.5	20.1
(E)-2-Hexanal	Carbonyls	C ₆ H ₁₀ O	1.02	0.84	6.93	18.0	16.6
Dihydroactinidioidide	Carbonyls	C ₁₁ H ₁₆ O ₂	1.02	0.93	2.67	16.6	17.2
Acetoin	Carbonyls	C ₄ H ₈ O ₂	1.57	0.78	2.41	13.5	15.4
(E)-2-Heptanal	Carbonyls	C ₇ H ₁₂ O	0.97	0.77	8.56	11.7	12.4
(E)-2-Octenal	Carbonyls	C ₈ H ₁₄ O	0.93	0.93	1.77	12.0	11.3
6-Methyl-5-hepten-2-one	Carbonyls	C ₈ H ₁₄ O	0.93	0.82	3.83	11.8	14.8
Benzaldehyde	Carbonyls	C ₇ H ₆ O	0.90	0.81	6.54	14.2	18.2
Phenylacetaldehyde	Carbonyls	C ₈ H ₈ O	0.87	0.94	1.44	12.6	13.3
Furfural	Carbonyls	C ₅ H ₄ O ₂	1.65	0.93	4.62	17.9	16.0
Hexanal	Carbonyls	C ₆ H ₁₂ O	0.98	0.86	2.72	9.1	12.1
Nonanal	Carbonyls	C ₉ H ₁₈ O	0.88	0.86	4.85	9.7	13.8
2-Methyl-nonane	Carbonyls	C ₁₀ H ₂₂	0.74	0.77	4.81	15.8	17.2
Octanal	Carbonyls	C ₈ H ₁₆ O	0.91	0.83	5.82	15.1	16.9
3-Methyl-1-butanol acetate	Esters	C ₇ H ₁₄ O ₂	1.13	0.74	8.11	15.4	16.1
Phenylmethyl acetate	Esters	C ₉ H ₁₀ O ₂	0.99	0.81	7.12	14.9	12.0
Ethyl benzoate	Esters	C ₉ H ₁₀ O ₂	0.99	0.86	5.35	10.6	11.6
Methyl benzoate	Esters	C ₈ H ₈ O ₂	1.04	0.88	1.06	15.2	16.6
2-Pentanol acetate	Esters	C ₇ H ₁₄ O ₂	1.13	0.79	8.13	8.0	10.1
(Z)-Hexen-1-ol acetate	Esters	C ₈ H ₁₄ O ₂	1.10	0.60	8.33	7.3	9.6
3-Ethyl hexanoate	Esters	C ₈ H ₁₄ O ₂	1.10	0.75	3.78	9.0	12.0
(Z)-3-Methyl hexenoate	Esters	C ₇ H ₁₂ O ₂	1.17	0.83	3.94	12.9	17.4
(E)-4-Ethyl heptenoate	Esters	C ₉ H ₁₆ O ₂	1.05	0.70	7.09	12.3	12.7
2-Phenylethyl acetate	Esters	C ₁₀ H ₁₂ O ₂	0.96	0.84	7.78	9.6	11.6
Hexyl acetate	Esters	C ₈ H ₁₆ O ₂	1.07	0.78	2.36	20.2	15.5
Methyl acetate	Esters	C ₃ H ₆ O ₂	2.03	0.94	2.15	19.1	19.0
Ethyl phenylacetate	Esters	C ₁₀ H ₁₂ O ₂	0.96	0.83	9.01	18.5	15.7
Ethyl phenylpropanoate	Esters	C ₁₁ H ₁₄ O ₂	0.94	0.93	2.71	14.7	15.8
Diethyl butanedioate	Esters	C ₈ H ₁₄ O ₄	1.50	0.90	1.77	11.2	12.2
Ethyl butanoate	Esters	C ₆ H ₁₂ O ₂	1.22	0.83	5.71	9.9	15.0
Ethyl decanoate	Esters	C ₁₂ H ₂₄ O ₂	0.94	0.87	2.56	13.5	15.8
Ethyl dodecanoate	Esters	C ₁₄ H ₂₈ O ₂	0.90	0.93	1.48	14.8	18.8
Ethyl Acetate	Esters	C ₄ H ₈ O ₂	1.57	0.94	5.31	12.2	16.5
Ethyl heptanoate	Esters	C ₉ H ₁₈ O ₂	1.02	0.80	5.88	15.5	17.0
Ethyl hexanoate	Esters	C ₈ H ₁₆ O ₂	1.07	0.68	13.57	13.6	15.4
Methyl hexanoate	Esters	C ₇ H ₁₄ O ₂	1.13	0.89	3.20	18.0	19.0
Isobutyl acetate	Esters	C ₆ H ₁₂ O ₂	1.22	0.88	8.64	18.9	14.7
Isopentyl hexanoate	Esters	C ₁₁ H ₂₂ O ₂	0.96	0.81	8.89	14.8	19.2
Methyl salicylate	Esters	C ₈ H ₈ O ₃	1.22	0.85	3.42	11.0	10.4
n-Propyl acetate	Esters	C ₅ H ₁₀ O ₂	1.35	0.88	4.87	13.0	11.6
Ethyl octanoate	Esters	C ₁₀ H ₂₀ O ₂	0.99	0.92	3.32	17.1	14.4
Ethyl pentanoate	Esters	C ₇ H ₁₄ O ₂	1.13	0.85	9.58	12.4	16.9
Ethyl propanoate	Esters	C ₅ H ₁₀ O ₂	1.35	0.88	3.57	12.4	15.5
2-Butyl acetate	Esters	C ₆ H ₁₂ O ₂	1.22	0.87	3.79	19.1	16.7
γ -Caprolactone	Lactones	C ₆ H ₁₀ O ₂	1.27	0.92	2.37	14.0	17.4
γ -Nonalactone	Lactones	C ₉ H ₁₆ O ₂	1.05	0.93	1.62	14.5	17.7
β -Cyclocitral	Terpenes	C ₁₀ H ₁₆ O	0.90	0.85	7.27	19.8	19.6
β -Damascenone	Terpenes	C ₁₃ H ₁₈ O	0.88	0.86	6.35	11.5	14.0
Linalool	Terpenes	C ₁₀ H ₁₈ O	0.88	0.72	2.11	12.1	15.5
Dihydroedulan I	Terpenes	C ₁₃ H ₂₂ O	0.85	0.65	2.94	7.1	10.3

This study, although focused on a limited sample set, provides evidence of meaningful quantitative changes in the volatile metabolome of silage after ensiling and when heterofermentative strains are added to improve aerobic stability and quality. These variations, since monitored with accurate amounts, enable robust marker discovery and validation over many experiments and among intervention studies conducted in large time frames and conditions.

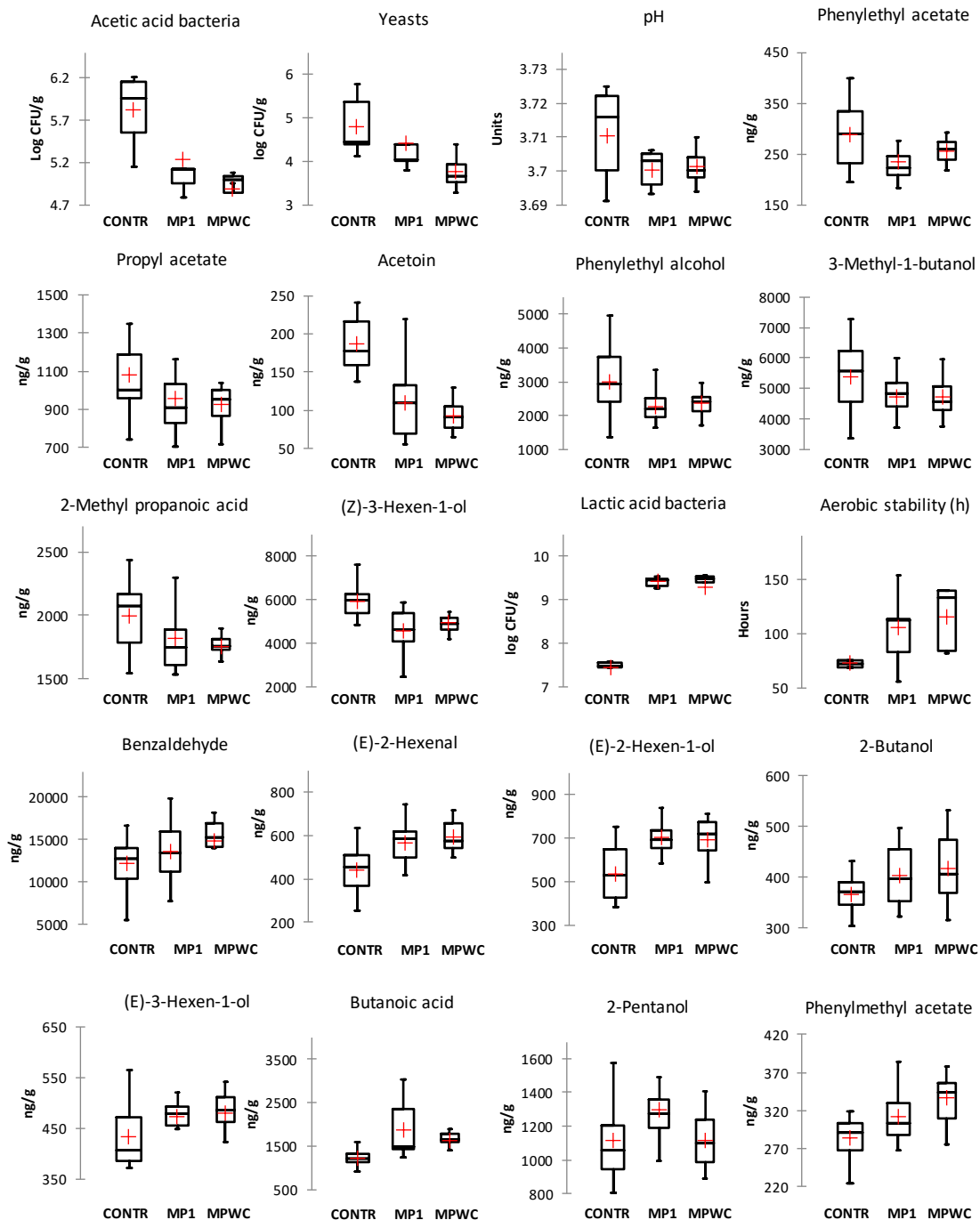
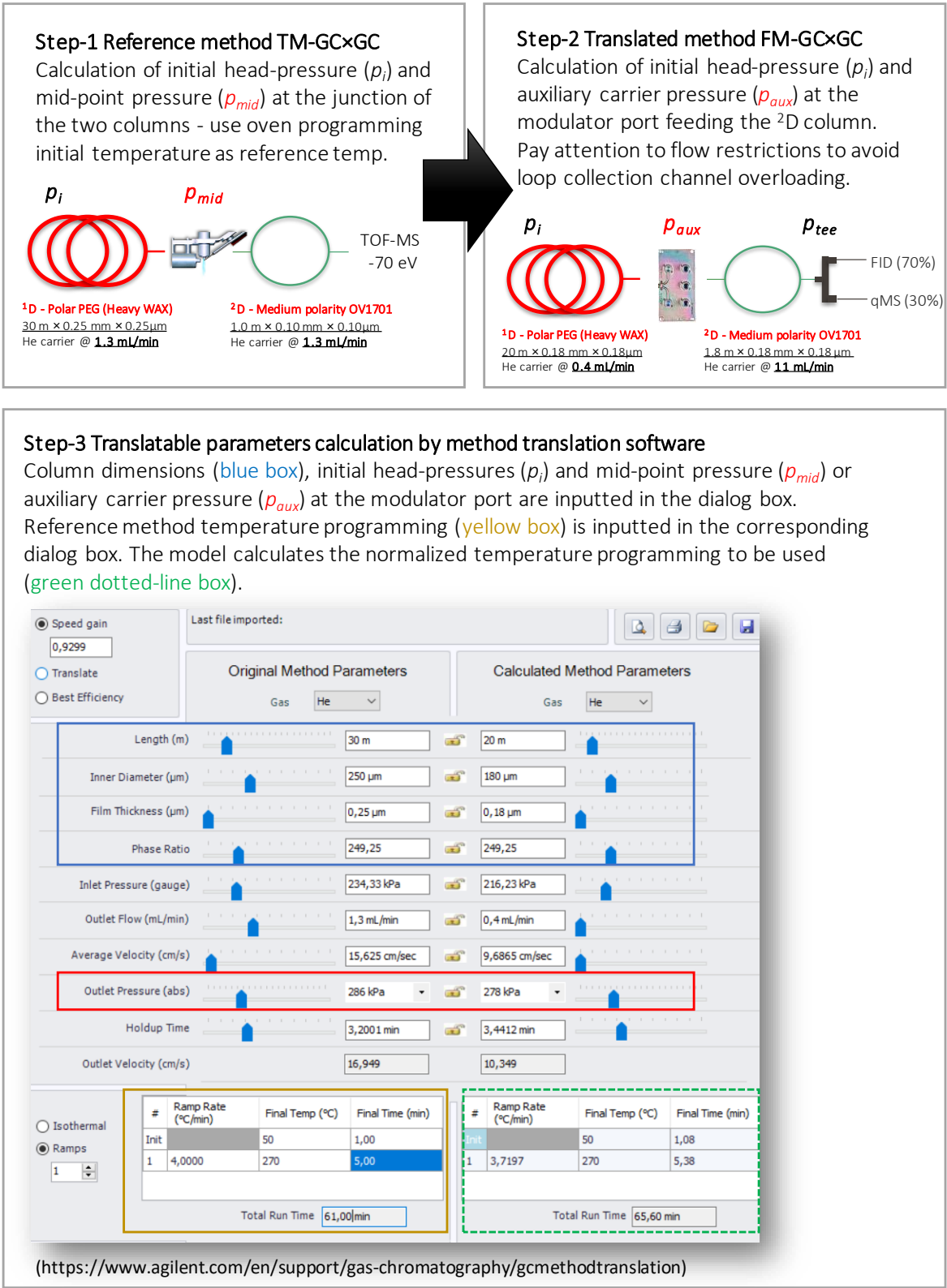


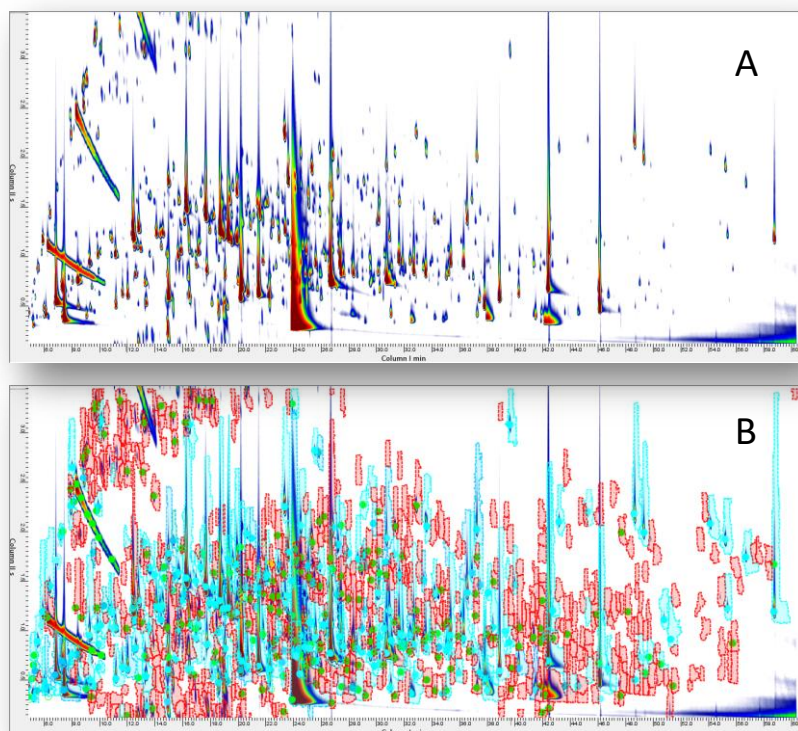
Figure 11.2: Box-plots illustrating the concentration (mean, median, min/max) of the most discriminant volatiles in NOT DELAYED samples belonging to the CONTR or LAB treated groups. Analyte trends are accompanied by additional indicators (i.e., microbial counts and pH) of relevance for the silage quality and stability.

Supplementary Information

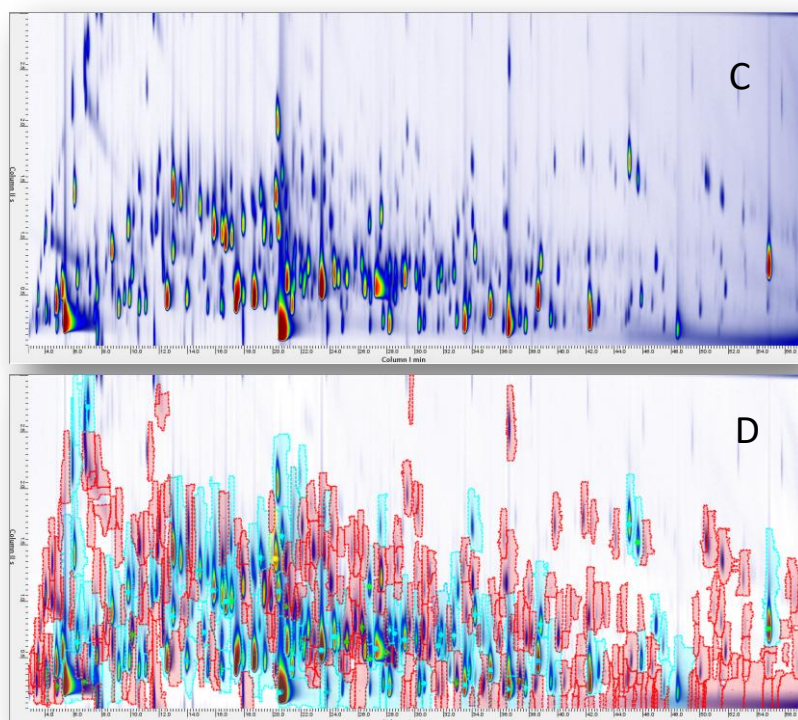


Supplementary Figure 1: step-by-step procedure for chromatographic parameters translation from TM-GC×GC to FM-GC×GC.

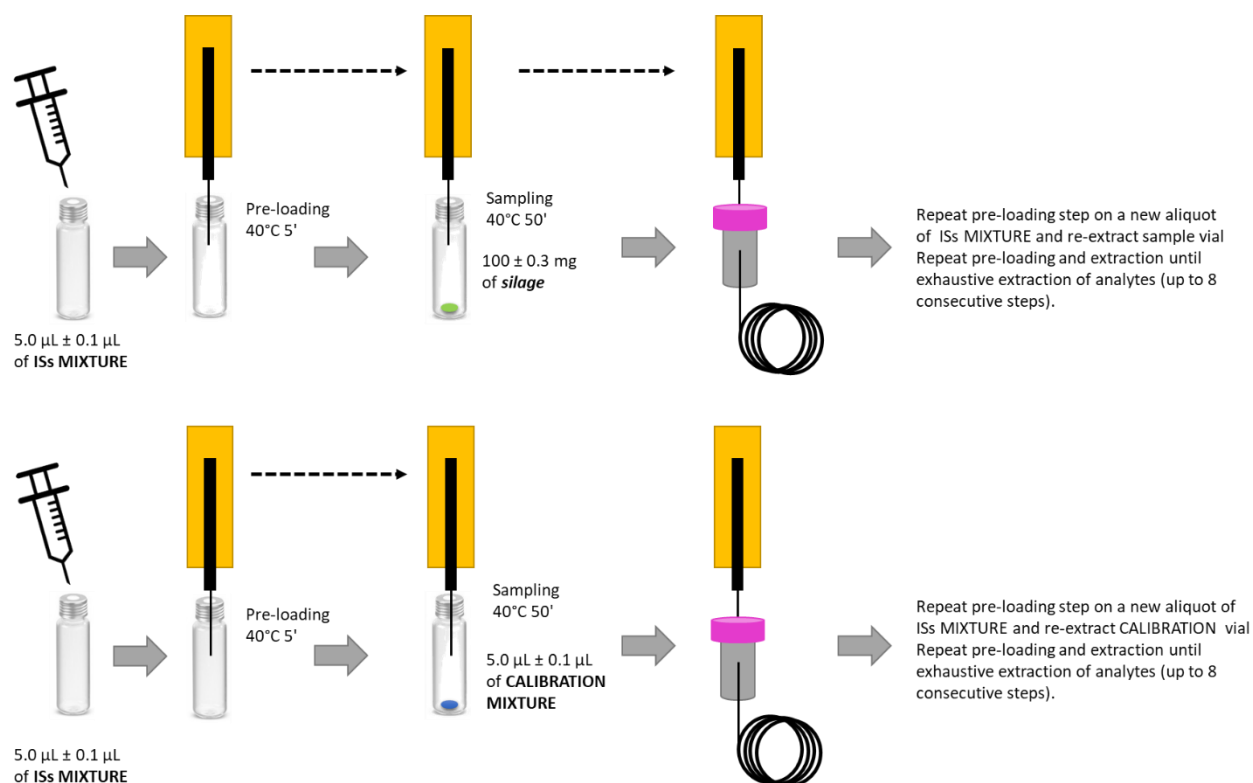
Reference method TM GC×GC-TOF MS



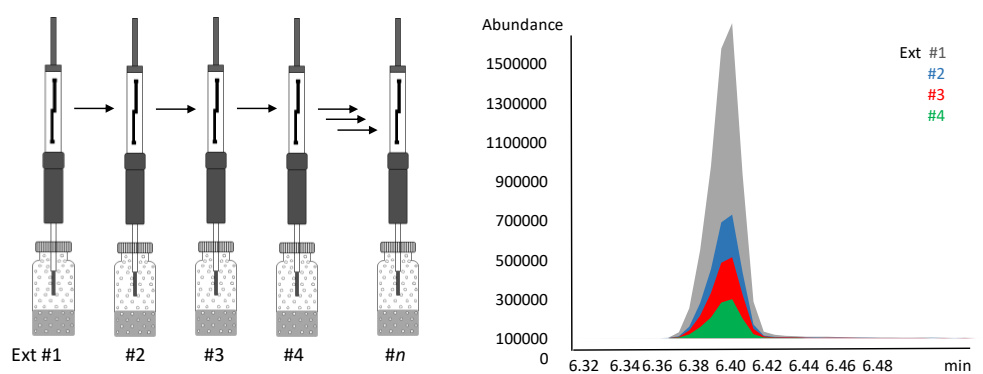
Translated method FM GC×GC-MS/FID - FID trace



Supplementary Figure 2: contour plots corresponding to the detectable volatilome of silage inoculated with the MPWC strain (NOT DELAYED – MPWC) as resulting from the TM reference method (2A) or translated FM method (2C). Graphics in Fig. 2B and Fig. 2D corresponds to UT peak regions delineated in the feature templates for the most comprehensive UT fingerprinting (red areas for non-targeted analytes and light-blue areas for targeted analytes).

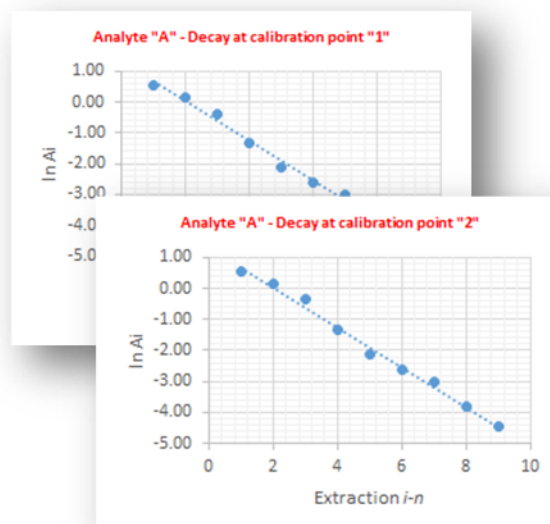


Exponential decay of an analyte subjected to consecutive extraction steps.



Supplementary Figure 3: Sequential steps corresponding to the exhaustive extraction of analytes from a silage sample or standard reference compounds from calibration solutions in HS linearity conditions by MHS-SPME approach.

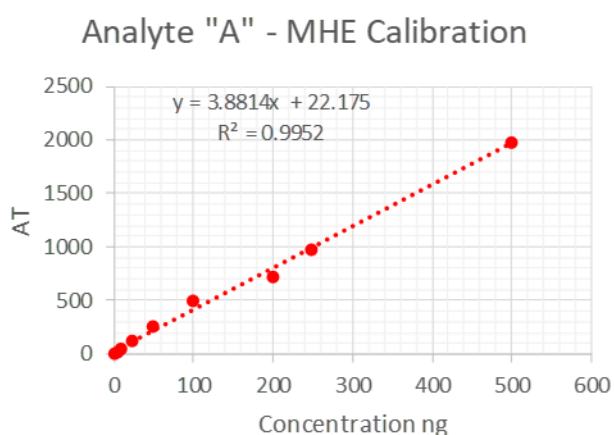
Analyte "A" - Calibration point "1"			
Extraction #n	i-1	A _i	ln A _i
1	0	1.76	0.57
2	1	1.17	0.15
3	2	0.69	-0.37
4	3	0.26	-1.35
5	4	0.12	-2.12
6	5	0.07	-2.62
7	6	0.05	-3.00
8	7	0.02	-3.79
9	8	0.01	-4.45
Regression equation $y = -0.6404x + 1.3147 - R^2 0.995$			
β 0.550			



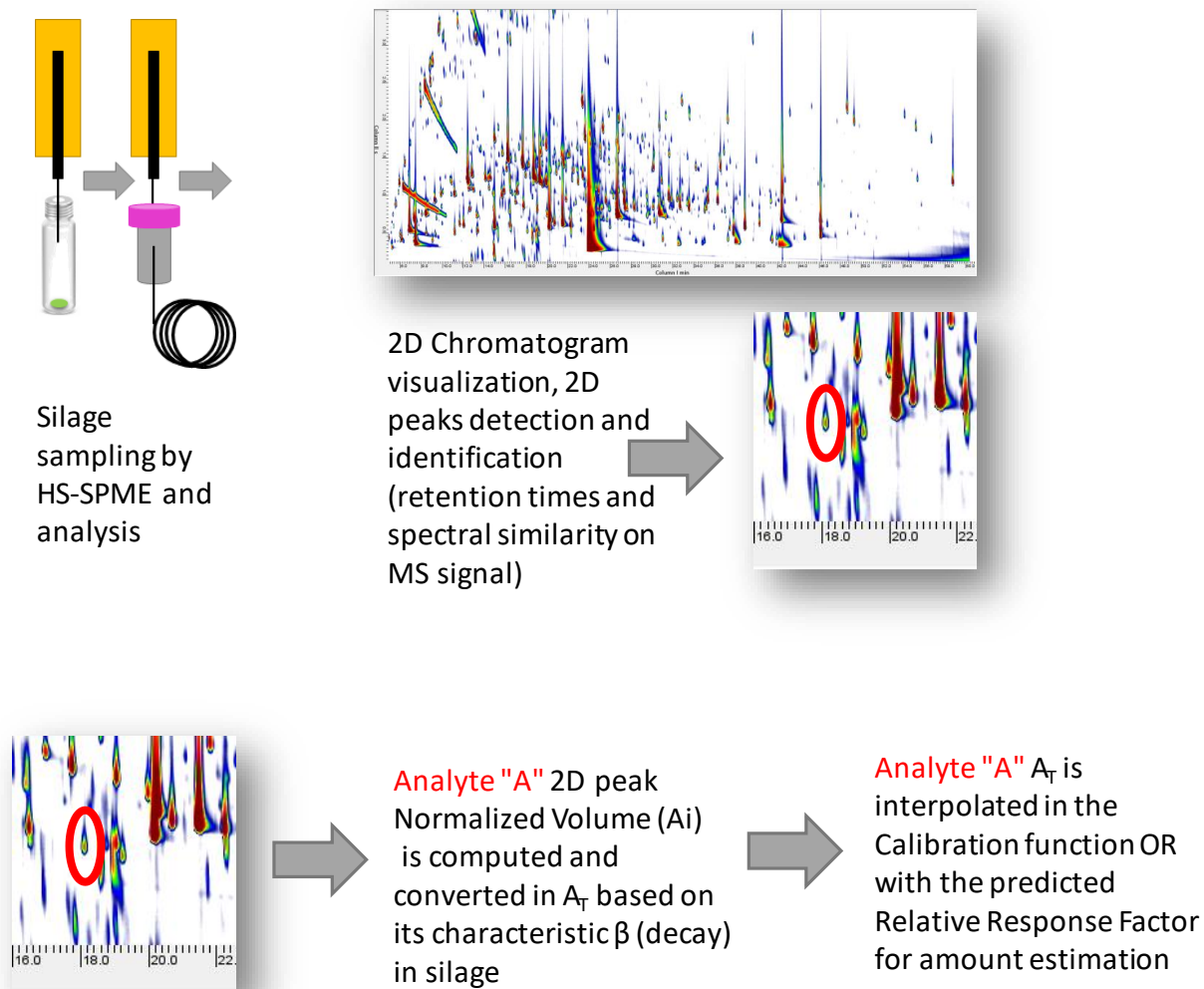
[....]

[....]

Analyte "A" - MHE Calibration			
Calibration Point #	Conc ng	A _i	A _T
1	1	2	5
2	5	10	24
3	10	23	54
4	25	50	118
5	50	109	258
6	100	208	492
7	200	305	721
8	250	412	974
9	500	839	1983
Regression equation $y = 3.8814x + 22.175 - R^2 0.995$			



Supplementary Figure 4: Sequential steps for Calibration curve with external standards and MHE-SPME extraction. Example for Analyte "A" subjected to up to 9 successive extractions at "n" different concentration levels. The Calibration Curve for Analyte A is then built by computing the A_T value obtained by applying Equation 1 at each calibration level/point.



Supplementary Figure 5: Sequential steps for Analyte “A” quantification. The procedure starts with the sample analysis after HS-SPME sampling (first extraction). The 2D chromatogram is visualized, pre-processed and once Analyte “A” is located, its response (2D Peak Volume) corresponding to the first extraction (A_i) is computed and used to estimate A_T (Equation 1). A_T can be interpolated in the Analyte “A” calibration curve (in the case of externally calibrated compounds) or with predicted Relative Response Factors to estimate the amount in the sample.

Supplementary Table 1: samples information about the in-field batch #ID, treatment (control samples CONTR or LAB inoculated samples), LAB commercial strain code, and ensiling delay.

Batch#	Treatment	LAB commercial strain code	Ensiling delay
7002	CONTR	-	NO
7003	CONTR	-	NO
7004	CONTR	-	NO
7005	CONTR	-	NO
7006	CONTR	-	NO
7012	LAB	MP1	NO
7013	LAB	MP1	NO
7014	LAB	MP1	NO
7015	LAB	MP1	NO
7016	LAB	MP1	NO
7022	LAB	MPWC	NO
7023	LAB	MPWC	NO
7024	LAB	MPWC	NO
7025	LAB	MPWC	NO
7026	LAB	MPWC	NO
6997	CONTR	-	24h
6998	CONTR	-	24h
6999	CONTR	-	24h
7000	CONTR	-	24h
7001	CONTR	-	24h
7007	LAB	MP1	24h
7008	LAB	MP1	24h
7009	LAB	MP1	24h
7010	LAB	MP1	24h
7011	LAB	MP1	24h
7017	LAB	MPWC	24h
7018	LAB	MPWC	24h
7019	LAB	MPWC	24h
7020	LAB	MPWC	24h
7021	LAB	MPWC	24h

Supplementary Table 2: Targeted analytes listed according to chemical class and in ascending retention order on the ¹D polar column. For the reference TM GC×GC-TOF MS method are reported ¹D and ²D retention times (¹t_R, ²t_R), linear retention index (I^T) experimental and tabulated on the NIST 2023 database. For the FM GC×GC-TOF MS information include analytes detected and/or quantified.

Compound name	Chemical class	TM GC×GC-TOF MS				FM GC×GC-TOF MS	
		¹ t _R min	² t _R sec	I ^T exp	I ^T tab	Detection	Quantification
Acetic Acid	Acids	23.9	0.4	1447	1452	X	
2-Methyl propanoic acid	Acids	27.8	0.4	1568	1570	X	X
Butanoic acid	Acids	28.5	0.4	1583	1581	X	X
3-Methyl butanoic acid	Acids	32.5	0.4	1672	1680	X	X
Pentanoic acid	Acids	33.7	0.5	1723	1720	X	
2-Methyl hexanoic acid	Acids	35.0	0.4	1758	1757		
2-Methyl pentanoic acid	Acids	35.1	0.5	1768	1770	X	X
Hexanoic acid	Acids	38.0	0.4	1868	1866	X	X
(E)-3-Hexenoic acid	Acids	40.2	0.6	1933	1930	X	X
Octanoic acid	Acids	44.2	0.5	2082	2080	X	X
Nonanoic acid	Acids	47.0	0.5	2175	2173	X	X
Benzoic acid	Acids	53.3	0.8	2410	2407	X	X
Ethanol	Alcohols	7.4	0.4	937	944	X	
2-Butanol	Alcohols	9.3	0.6	1027	1036	X	X
1-Propanol	Alcohols	10.0	0.5	1045	1051	X	X
2-Methyl-1-propanol	Alcohols	11.3	0.6	1097	1101	X	X
3-Pentanol	Alcohols	11.6	0.7	1108	1111	X	X
2-Pentanol	Alcohols	12.0	0.7	1112	1112	X	X
2-Methyl-3-pentanol	Alcohols	13.4	0.8	1158	1167	X	X
1-Penten-3-ol	Alcohols	13.4	0.6	1161	1158	X	
3-Methyl-1-butanol	Alcohols	14.9	0.7	1202	1211	X	X
2-Hexanol	Alcohols	15.4	0.8	1213	1222		
1-Pentanol	Alcohols	16.6	0.7	1243	1252	X	X
4-Heptanol	Alcohols	17.7	0.9	1276	1285	X	X
2-3-Dimethyl-1-butanol	Alcohols	18.3	0.7	1290	1291		
(Z)-2-Penten-1-ol	Alcohols	19.0	0.6	1311	1317	X	X
2-Heptanol	Alcohols	19.0	0.9	1311	1319	X	X
(Z)-4-Hepten-2-ol	Alcohols	20.0	0.8	1315	1310		
1-Hexanol	Alcohols	20.1	0.7	1344	1344		
(E)-3-Hexen-1-ol	Alcohols	20.7	0.7	1356	1352	X	X
6-Methyl-2-heptanol	Alcohols	21.1	1.0	1372	1379		
(Z)-3-Hexen-1-ol	Alcohols	21.4	0.7	1375	1373	X	X
3-Octanol	Alcohols	21.8	1.0	1389	1398	X	X
(Z)-2-Hexen-1-ol	Alcohols	22.2	0.7	1401	1401	X	
2-Octanol	Alcohols	22.7	1.0	1411	1405		
(E)-2-Hexen-1-ol	Alcohols	22.7	1.1	1415	1411	X	X
1-Heptanol	Alcohols	24.0	0.8	1447	1447		
1-Octen-3-ol	Alcohols	24.7	0.9	1458	1460	X	X
6-Methyl-5-hepten-2-ol	Alcohols	24.8	0.9	1459	1465	X	X
4-Nonanol	Alcohols	25.0	1.1	1481	1479	X	X
1-(2-methoxy-1-methylethoxy)-2-propanol	Alcohols	25.1	0.9	1481	1478	X	X
2-Ethyl-1-hexanol	Alcohols	25.4	1.1	1489	1488	X	X
(Z)-4-Hepten-1-ol	Alcohols	25.6	0.7	1496	1502		
(E)-4-Hepten-1-ol	Alcohols	25.7	0.7	1500	1504		
2,3-Butanediol	Alcohols	27.5	0.5	1539	1545		
1-Octanol	Alcohols	27.7	0.9	1556	1555	X	X
4-Methyl-cyclohexanemethanol	Alcohols	28.5	0.9	1559	1563		
1,3-Butanediol	Alcohols	28.5	0.5	1576	1578		
1,2-Propandiol	Alcohols	28.8	0.4	1585	1585	X	
2-Octen-1-ol	Alcohols	29.6	0.8	1611	1617	X	
2-(2-Ethoxyethoxy) ethanol	Alcohols	29.8	0.7	1613	1615		
2-Furanmethanol (Furfuryl alcohol)	Alcohols	31.0	0.5	1654	1651	X	X
1-Nonanol	Alcohols	31.1	1.0	1657	1663	X	

Supplementary Table 2: continue

Compound name	Chemical class	TM GC×GC-TOF MS				FM GC×GC-TOF MS	
		¹ t _R min	² t _R sec	I ⁺ exp	I ⁺ tab	Detection	Quantification
(Z)-3-Nonen-1-ol	Alcohols	31.9	0.9	1678	1682		
2-(2-Butoxyethoxy) ethanol	Alcohols	35.4	0.4	1780	1786	x	x
Toluene	Aromatics	10.0	0.9	1045	1054	x	
Ethylbenzene	Aromatics	12.5	1.2	1129	1136		
1,3-Dimethyl benzene (m-Xylene)	Aromatics	13.2	1.2	1147	1150	x	
1-Ethyl-3-methyl benzene	Aromatics	15.9	1.4	1225	1216		
Styrene	Aromatics	16.9	1.0	1254	1264		
1,2,4-Trimethyl benzene	Aromatics	17.9	1.3	1281	1280	x	
2-Ethyl-1,4-dimethylbenzene (p-Xylene)	Aromatics	20.5	1.5	1350	1364	x	x
1-Ethyl-2,3-dimethylbenzene (o-Xylene)	Aromatics	21.1	1.5	1368	1369	x	x
Indane	Aromatics	21.2	1.3	1370	1365		
1,3-Bis(1,1-dimethylethyl) benzene	Aromatics	23.3	2.3	1430	1427	x	x
1,2,4,5-Tetramethylbenzene	Aromatics	23.6	1.4	1435	1438		
Benzaldehyde	Aromatics	26.7	0.8	1531	1529	x	x
1,2,3,4-Tetrahydro-1,6,8-trimethyl-naphthalene	Aromatics	30.6	1.6	1617	1610		
Phenylacetaldehyde	Aromatics	29.9	0.8	1619	1625	x	x
1,2,3,4-Tetrahydro-1,5,8-trimethyl-naphthalene	Aromatics	30.9	1.6	1649	1655		
1-Ethenyl-4-methoxy benzene	Aromatics	31.8	0.9	1668	1670		
Naphthalene	Aromatics	33.6	1.1	1721	1719		
1,1,6-Trimethyl-1,2-dihydronaphthalene	Aromatics	34.6	1.4	1750	1747		
1-Phenylethanol	Aromatics	36.3	0.6	1820	1818	x	x
Benzyl alcohol	Aromatics	37.6	0.6	1860	1864	x	x
2-Methoxy phenol	Aromatics	37.7	0.6	1865	1862	x	x
Phenylethyl alcohol	Aromatics	39.6	0.6	1917	1915	x	x
o-Cresol	Aromatics	40.3	0.7	1950	1956	x	x
Phenol	Aromatics	42.0	0.5	1996	1994	x	x
4-Ethyl-2-methoxy-phenol (4-Ethyl guaiacol)	Aromatics	42.4	0.7	2025	2022	x	x
p-Cresol	Aromatics	43.6	0.5	2051	2057	x	x
2-Phenoxy ethanol	Aromatics	45.3	0.6	2122	2126		
2,3-Dimethyl phenol	Aromatics	46.1	0.5	2160	2155		
Eugenol	Aromatics	46.0	0.7	2166	2164		
4-Ethyl-phenol	Aromatics	47.0	0.9	2172	2180	x	x
1-Hydroxy-2-acetyl-4-methylbenzene (o-Acetyl-p-cresol)	Aromatics	47.4	0.7	2176	2178	x	x
(E)-2,6-Dimethyl-2,7-octadien-1,6-diol	Aromatics	49.4	0.6	2287	2285		
2,4-Di-tert-butylphenol	Aromatics	49.6	0.8	2299	2294	x	x
Acetaldehyde	Carbonyls	5.2	0.4	720	715		
2-Methyl-propanal	Carbonyls	6.0	0.5	816	813		
Acetone	Carbonyls	6.0	0.5	816	821	x	
2-Propenal	Carbonyls	6.2	0.4	842	851		
3-Methyl-butanal	Carbonyls	7.3	0.7	921	916	x	
2,3-Butanedione	Carbonyls	8.4	0.6	979	978		
(E)-2-Butenal	Carbonyls	9.9	0.7	1040	1046		
Hexanal	Carbonyls	11.1	1.1	1081	1084	x	x
(E)-2-Pentenal	Carbonyls	12.5	0.9	1128	1121		
2-Methyl-2-pentenal	Carbonyls	13.4	1.1	1160	1171		
2-Heptanone	Carbonyls	14.4	1.3	1182	1184	x	
4-Methyl-2-heptanone	Carbonyls	15.2	1.5	1208	1206		
(E)-2-Hexenal	Carbonyls	15.6	1.1	1219	1220	x	x
(Z)-4-Heptenal	Carbonyls	16.5	1.1	1241	1244		
3-Octanone	Carbonyls	16.9	1.5	1254	1250	x	
3-Hydroxy-2-butanone (Acetoin)	Carbonyls	18.1	0.6	1286	1287	x	x

Supplementary Table 2: continue

Compound name	Chemical class	TM GC×GC-TOF MS				FM GC×GC-TOF MS	
		¹ t _R min	² t _R sec	I ⁺ exp	I ⁺ tab	Detection	Quantification
Octanal	Carbonyls	18.2	1.5	1289	1291	X	X
2,2,6-Trimethylcyclohexanone	Carbonyls	19.4	1.6	1321	1312		
(E)-2-Heptenal	Carbonyls	20.1	1.5	1340	1336	X	X
6-Methyl-5-hepten-2-one	Carbonyls	20.8	1.7	1362	1365	X	X
Nonanal	Carbonyls	22.1	1.6	1394	1392	X	X
3-Furancarboxaldehyde	Carbonyls	23.2	0.6	1426	1426		
(E)-2-Octenal	Carbonyls	23.3	1.3	1429	1434	X	X
2-Furancarboxaldehyde (Furfural)	Carbonyls	24.4	0.6	1453	1455	X	X
(E,Z)-2,4-Heptadienal	Carbonyls	24.8	1.0	1466	1464		
(E,E)-2,4-Heptadienal	Carbonyls	25.6	1.0	1491	1493	X	X
Decanal	Carbonyls	25.8	1.7	1507	1505		
(E)-2-Nonenal	Carbonyls	26.2	1.4	1524	1530		
3,5,5-Trimethyl-2-cyclohexen-1-one (α-Isophorone)	Carbonyls	28.5	1.3	1583	1591		
6-Methyl-3,5-heptadiene-2-one	Carbonyls	28.8	1.1	1584	1582		
Safranal	Carbonyls	30.7	1.3	1640	1648		
(E)-2-Phenyl-2-butenal	Carbonyls	39.6	0.9	1913	1910		
6,10,14-Trimethyl-2-pentadecanone	Carbonyls	45.2	2.3	2108	2110		
Methyl acetate	Esters	6.1	0.5	826	832	X	X
Ethyl acetate	Esters	6.8	0.6	880	870	X	X
Methyl propionate	Esters	7.2	0.5	907	910	X	X
Ethyl propanoate	Esters	8.0	0.8	958	957	X	X
Propyl acetate	Esters	8.3	0.8	976	977	X	X
Isobutyl acetate	Esters	9.2	1.0	1015	1018	X	X
Ethyl butanoate	Esters	9.8	1.1	1035	1035	X	X
Propyl propanoate	Esters	10.0	1.1	1044	1047		
Ethyl-2-methyl butanoate	Esters	10.3	1.3	1052	1052		
Ethyl-3-methyl butanoate	Esters	10.7	1.3	1069	1071		
Butyl acetate	Esters	10.9	1.1	1073	1075	X	X
2-Pentanol acetate	Esters	10.9	1.2	1073	1075	X	X
3-Methyl-1-butanol acetate	Esters	12.4	1.3	1123	1121	X	X
Ethyl pentanoate	Esters	12.8	1.4	1135	1137	X	X
Pentyl acetate	Esters	13.9	1.3	1174	1175		
Methyl hexanoate	Esters	14.5	1.3	1188	1184	X	X
3-Methyl-1-butanol propanoate (Isoamyl propionate)	Esters	14.6	1.6	1191	1183		
Butyl butanoate	Esters	15.7	1.6	1221	1220	X	
Ethyl hexanoate	Esters	16.2	1.6	1235	1232	X	X
3-Methyl-2-buten-1-ol acetate	Esters	16.6	1.1	1246	1248		
Methyl-(3Z)-3-hexenoate	Esters	17.1	1.1	1259	1252	X	X
2-Heptyl acetate	Esters	17.3	1.7	1263	1266		
3-Methylbutyl butanoate	Esters	17.4	1.8	1267	1259		
Hexyl acetate	Esters	17.6	1.5	1273	1277	X	X
Ethyl (3Z)-3-hexenoate	Esters	18.3	1.3	1292	1295		
Ethyl (3E)-3-hexenoate	Esters	18.7	1.3	1302	1301	X	X
Methyl 2-hydroxy propanoate (Methyl lactate)	Esters	19.2	0.5	1318	1328		
(Z)-3-Hexen-1-ol acetate	Esters	19.3	1.3	1319	1313	X	X
(E)-2-Hexen-1-ol acetate	Esters	19.8	1.2	1333	1338		
Ethyl heptanoate	Esters	19.9	1.7	1335	1337	X	X
2-Hydroxyethyl propanoate	Esters	20.1	0.7	1341	1339	X	
2-Methylpropyl hexanoate	Esters	20.6	2.0	1354	1347		
3-Methylbutyl pentanoate	Esters	20.9	2.0	1363	1354		
3-Hexenyl hexanoate	Esters	31.3	1.8	1372	1371		
Heptyl acetate	Esters	21.4	1.6	1376	1377		
Ethyl (4E)-4-heptenoate	Esters	21.5	1.5	1379	1382	X	X
Methyl octanoate	Esters	21.9	1.6	1390	1399		
(Z)-3-Butyl hexanoate	Esters	22.9	1.7	1416	1421		
2-Hydroxypropyl propanoate	Esters	23.2	0.7	1426	1424	X	

Supplementary Table 2: continue

Compound name	Chemical class	TM GC×GC-TOF MS				FM GC×GC-TOF MS	
		^t _{tr} min	^t _{tr} sec	I ^t exp	I ^t tab	Detection	Quantification
2-Hydroxy-3-methyl butanoate	Esters	23.3	0.9	1427	1421		
Ethyl octanoate	Esters	23.6	1.8	1435	1440	x	x
2-Hydroxy-2-methylpropyl propanoate	Esters	24.4	0.8	1451	1455		
Isopentyl hexanoate	Esters	24.5	2.1	1456	1451	x	x
Octyl acetate	Esters	25.0	1.7	1472	1480		
Ethyl-7-octenoate	Esters	25.6	1.5	1493	1486		
(E,E)-Ethyl-2,4-hexadienoate	Esters	25.6	1.1	1495	1501		
Ethyl nonanoate	Esters	27.2	1.9	1532	1530		
α-Methyl-γ-butyrolactone	Esters	27.7	0.8	1555	1557		
Ethyl-2-hydroxy-4-methylpentanoate	Esters	28.5	0.9	1569	1562		
Nonyl acetate	Esters	28.9	1.8	1587	1582		
Hexyl hexanoate	Esters	29.9	2.1	1595	1599		
1,2-Propandiol-2-acetate	Esters	30.1	0.6	1624	1621	x	
Ethyl decanoate	Esters	30.3	2.0	1626	1624	x	x
Methyl benzoate	Esters	30.5	0.9	1633	1631	x	x
γ-Butyrolactone	Esters	30.6	0.7	1633	1635		
γ-Vinyl-γ-valerolactone	Esters	31.7	0.9	1671	1669		
Diethyl succinate	Esters	31.9	1.0	1672	1670	x	x
Dihydro-3,5-dimethyl-2(3H)-furanone	Esters	32.0	0.9	1677	1672		
Ethyl benzoate	Esters	31.6	1.0	1680	1673	x	x
Benzyl acetate	Esters	32.8	0.9	1682	1690	x	x
γ-Hexalactone	Esters	32.7	0.9	1692	1689	x	x
5-Ethylidihydro-2(3H)-furanone	Esters	32.8	0.6	1696	1694		
(Z)-3-Hexenyl-(Z)-3-hexenoate	Esters	33.4	1.6	1709	1715	x	
2-Hydroxy-(E)-3-hexenyl propionate ((E)-3-Hexenyl lactate)	Esters	34.0	0.9	1734	1727		
Methyl-2-phenylacetate	Esters	34.5	0.9	1743	1749	x	
Ethyl phenylacetate	Esters	35.3	1.0	1773	1775	x	x
Methyl salicylate	Esters	35.4	0.9	1780	1778	x	x
2-Phenylethyl acetate	Esters	36.2	1.0	1816	1820	x	x
Ethyl dodecanoate	Esters	37.2	2.1	1840	1848	x	x
Ethyl phenylpropanoate	Esters	37.6	1.2	1878	1880	x	x
γ-Octalactone	Esters	38.4	1.0	1902	1910		
Ethyl-3-phenyl propionate	Esters	39.3	1.1	1912	1914	x	
γ-Nonalactone	Esters	42.5	1.1	2017	2023	x	x
Ethyl tetradecanoate	Esters	43.2	2.2	2053	2046		
Ethyl hexadecanoate	Esters	48.6	2.2	2257	2254		
Ethyl-2-hydroxy-3-phenylpropanoate	Esters	49.0	0.8	2271	2273		
Ethyl-9-hexadecenoate	Esters	49.3	2.1	2273	2277		
Dihydroactinolide	Esters	50.8	1.2	2330	2332	x	x
Ethyl Oleate	Esters	54.1	2.1	2472	2470		
Ethyl linoleate	Esters	55.2	2.0	2511	2519		
Methyl linolenate	Esters	56.6	1.8	2577	2583		
Ethyl sorbate	Esters	26.1	1.1	1507	-		
2-Ethyl furane	Heterocycle	7.9	0.7	952	955	x	
2-Pentyl furane	Heterocycle	16.1	1.4	1232	1230		
2-Ethyl-3-methyl maleimide	Heterocycle	48.7	0.6	2254	2260		
Pentane	Hydrocarbons	4.7	0.4	500	500	x	
n-Hexane	Hydrocarbons	4.8	0.5	600	600		
Heptane	Hydrocarbons	5.1	0.4	700	700	x	
2-Methyl pentane	Hydrocarbons	5.1	0.7	711	-		
Methyl cyclopentane	Hydrocarbons	5.2	0.6	737	-		
Cyclohexane	Hydrocarbons	5.3	0.6	750	741		
2,5-Dimethyl hexane	Hydrocarbons	5.3	0.8	753	742		
2-Methyl heptane	Hydrocarbons	5.5	0.9	774	764		
Octane	Hydrocarbons	5.8	1.1	800	800	x	
2,4-Dimethyl heptane	Hydrocarbons	5.8	1.2	805	797		
4-Methyl octane	Hydrocarbons	6.4	1.5	850	840		

Supplementary Table 2: continue

Compound name	Chemical class	TM GC×GC-TOF MS				FM GC×GC-TOF MS	
		^a t _R min	^a t _R sec	I ^T exp	I ^T tab	Detection	Quantification
n-Nonane	Hydrocarbons	6.9	1.8	900	900	x	
2,7-Dimethyl octane	Hydrocarbons	7.2	2.0	915	-		
2,6-Dimethyl octane	Hydrocarbons	7.3	2.1	921	921		
2-Methyl nonane	Hydrocarbons	7.9	2.3	952	-	x	x
2,2,4,6,6-Pentamethyl-heptane	Hydrocarbons	8.0	2.7	958	957	x	
3-Methyl nonane	Hydrocarbons	8.1	2.4	964	967		
n-Decane	Hydrocarbons	8.8	2.6	1000	1000	x	x
(E)-4-Decene	Hydrocarbons	9.6	2.3	1033	1039		
1-Decene	Hydrocarbons	9.9	2.3	1040	1048		
5-Methyl decane	Hydrocarbons	10.0	3.2	1046	1049		
4-Methyl decane	Hydrocarbons	10.1	3.2	1047	1054	x	x
n-Undecane	Hydrocarbons	11.4	3.6	1100	1100	x	
2-Methyl undecane	Hydrocarbons	13.3	3.8	1151	1155		
3-Ethylidene cycloheptane	Hydrocarbons	14.1	1.6	1178	1180		
n-Dodecane	Hydrocarbons	14.8	3.8	1200	1200	x	
1-Dodecene	Hydrocarbons	16.5	3.1	1241	1233	x	
n-Tridecane	Hydrocarbons	18.6	0.5	1300	1300		
n-Tetradecane	Hydrocarbons	22.2	0.6	1400	1400		
1-Tetradecene	Hydrocarbons	23.7	3.4	1436	1428		
Dimethyl sulfide	Others	5.4	0.4	764	757		
Hexanenitrile	Others	18.6	1.0	1301	1303		
3-Methylthio-propanal (Methional)	Others	24.4	0.8	1452	1451		
3-(Methylthio)-propyl acetate	Others	30.1	1.0	1622	1620		
Diphenyl ether	Others	41.5	1.0	1983	1991		
α-Pinene	Terpenes	9.5	1.8	1028	1026	x	
δ-3-Carene	Terpenes	13.2	1.9	1147	1142		
β-Myrcene	Terpenes	13.4	1.7	1161	1155		
Limonene	Terpenes	14.8	1.8	1200	1194	x	
trans-Anhydrolinalool oxide	Terpenes	15.3	1.6	1210	1209		
1,8-Cineole	Terpenes	15.3	2.0	1211	1212	x	
cis-Linalool oxide	Terpenes	23.8	1.2	1444	1441	x	
Cyclosativene	Terpenes	25.5	2.9	1489	1487		
α-Copaene	Terpenes	25.8	2.9	1498	1493		
Bornylene	Terpenes	26.6	1.7	1528	1536		
Dihydroedulan I	Terpenes	26.8	2.1	1532	1529	x	x
Linalool	Terpenes	27.3	1.0	1536	1544	x	x
Theaspirane	Terpenes	27.5	2.1	1543	1540		
α-Ionone	Terpenes	27.9	1.8	1559	1565		
β-Cyclocitral	Terpenes	29.9	1.4	1616	1611	x	x
Menthol	Terpenes	30.6	1.1	1634	1631	x	
α-Terpineol	Terpenes	32.4	1.0	1689	1687	x	
β-Bisabolene	Terpenes	33.6	2.1	1715	1723		
trans-β-Damascenone	Terpenes	36.5	1.4	1820	1815	x	x
Dihydro-β-ionone	Terpenes	36.9	1.6	1827	1825		
Trans-geraniol	Terpenes	36.9	0.9	1839	1841		
Geranylacetone	Terpenes	37.5	1.5	1846	1852		
Dihydrodehydro-β-ionone	Terpenes	38.0	1.4	1870	1865		
Neophytadiene	Terpenes	39.6	3.1	1917	1915		
β-Ionone	Terpenes	40.0	1.5	1918	1926		
Nerolidol	Terpenes	42.7	1.3	2024	2023		

Supplementary Table 3: Accurate amounts expressed in ng g⁻¹ for the 98 volatiles on the sub-set of NOT DELAYED samples. Amounts, mean of replicated analyses, accompanied by their relative uncertainty. The p value as resulting from the Kruskal-Wallis non-parametric significance test for the classes CONTR vs. LAB treated is reported; in bold values where the difference between classes is meaningful (significance level alpha=0.05).

The table is available at the link: <https://pubs.acs.org/doi/10.1021/acs.jafc.4c05877>

References

- [1] R. E. Muck, E. M. G. Nadeau, T. A. McAllister, F. E. Contreras-Govea, M. C. Santos, and L. Kung, "Silage review: Recent advances and future uses of silage additives," *J. Dairy Sci.*, vol. 101, no. 5, pp. 3980–4000, May 2018.
- [2] G. Borreani, E. Tabacco, R. J. Schmidt, B. J. Holmes, and R. E. Muck, "Silage review: Factors affecting dry matter and quality losses in silages," *J. Dairy Sci.*, vol. 101, no. 5, pp. 3952–3979, May 2018.
- [3] D. H. Kleinschmit and L. Kung, "A Meta-Analysis of the Effects of *Lactobacillus buchneri* on the Fermentation and Aerobic Stability of Corn and Grass and Small-Grain Silages," *J. Dairy Sci.*, vol. 89, no. 10, pp. 4005–4013, Oct. 2006.
- [4] F. Driehuis, S. J. W. H. Oude Elferink, and P. G. Van Wikselaar, "Fermentation characteristics and aerobic stability of grass silage inoculated with *Lactobacillus buchneri*, with or without homofermentative lactic acid bacteria," *Grass Forage Sci.*, vol. 56, no. 4, pp. 330–343, Dec. 2001.
- [5] H. Danner, M. Holzer, E. Mayrhuber, and R. Braun, "Acetic Acid Increases Stability of Silage under Aerobic Conditions," *Appl. Environ. Microbiol.*, vol. 69, no. 1, pp. 562–567, Jan. 2003.
- [6] F. Ferrero, E. Tabacco, S. Piano, M. Casale, and G. Borreani, "Temperature during conservation in laboratory silos affects fermentation profile and aerobic stability of corn silage treated with *Lactobacillus buchneri*, *Lactobacillus hilgardii*, and their combination," *J. Dairy Sci.*, vol. 104, no. 2, pp. 1696–1713, Feb. 2021.
- [7] S. J. Krizsan, F. Westad, T. Ådnøy, E. Odden, S. E. Aakre, and Å. T. Randby, "Effect of volatile compounds in grass silage on voluntary intake by growing cattle," *Animal*, vol. 1, no. 2, pp. 283–292, 2007.
- [8] K. Weiss, B. Kroschewski, and H. Auerbach, "Effects of air exposure, temperature and additives on fermentation characteristics, yeast count, aerobic stability and volatile organic compounds in corn silage," *J. Dairy Sci.*, vol. 99, no. 10, pp. 8053–8069, Oct. 2016.
- [9] S. D. Hafner *et al.*, "Emission of volatile organic compounds from silage: Compounds, sources, and implications," *Atmos. Environ.*, vol. 77, pp. 827–839, Oct. 2013.
- [10] N. Bandla, K. H. Südekum, and K. Gerlach, "Review: Role of silage volatile organic compounds in influencing forage choice behavior and intake in ruminants," *Animal Feed Science and Technology*, vol. 307, no. December 2023. Elsevier B.V., p. 115853, 2024.
- [11] S. Squara, F. Ferrero, E. Tabacco, C. Cordero, and G. Borreani, "Effect of Inoculation with *Lentilactobacillus buchneri* and *Lactocaseibacillus paracasei* on the Maize Silage Volatilome: The Advantages of Advanced 2D-Chromatographic Fingerprinting Approaches," *J. Agric. Food Chem.*, vol. 70, no. 38, pp. 12232–12248, Sep. 2022.
- [12] A. Caratti *et al.*, "Boosting comprehensive two-dimensional chromatography with artificial intelligence: Application to food-omics," *TrAC Trends Anal. Chem.*, vol. 174, p. 117669, May 2024.

- [13] F. Stilo, C. Bicchi, A. Robbat, S. E. Reichenbach, and C. Cordero, "Untargeted approaches in food-omics: The potential of comprehensive two-dimensional gas chromatography/mass spectrometry," *TrAC Trends Anal. Chem.*, vol. 135, p. 116162, Feb. 2021.
- [14] N. I. of S. and Technology, *NIST/EPA/NIH Mass Spectral Library with Search Program: (Data Version: NIST 08, Software Version 2.0f)*, vol. 2010, no. 1/22. Gaithersburg MD: National Institute of Standards and Technology (NIST), 2005.
- [15] I. L. Malkina, A. Kumar, P. G. Green, and F. M. Mitloehner, "Identification and Quantitation of Volatile Organic Compounds Emitted from Dairy Silages and Other Feedstuffs," *J. Environ. Qual.*, vol. 40, no. 1, pp. 28–36, Jan. 2011.
- [16] S. K. Roberts, M. McAinsh, H. Cantopher, and S. Sandison, "Calcium Dependence of Eugenol Tolerance and Toxicity in *Saccharomyces cerevisiae*," *PLoS One*, vol. 9, no. 7, p. e102712, Jul. 2014.
- [17] C. Lv, P. Wang, L. Ma, M. Zheng, Y. Liu, and F. Xing, "Large-Scale Comparative Analysis of Eugenol-Induced/Repressed Genes Expression in *Aspergillus flavus* Using RNA-seq," *Front. Microbiol.*, vol. 9, May 2018.
- [18] N. Jiang *et al.*, "Transcriptomic analysis of inhibition by eugenol of ochratoxin A biosynthesis and growth of *Aspergillus carbonarius*," *Food Control*, vol. 135, no. 1800, p. 108788, May 2022.
- [19] F. Stilo *et al.*, "A step forward in the equivalence between thermal and differential-flow modulated comprehensive two-dimensional gas chromatography methods," *J. Chromatogr. A*, vol. 1627, p. 461396, Sep. 2020.
- [20] J. Y. De Saint Laumer *et al.*, "Prediction of response factors for gas chromatography with flame ionization detection: Algorithm improvement, extension to silylated compounds, and application to the quantification of metabolites," *J. Sep. Sci.*, vol. 38, no. 18, pp. 3209–3217, Jul. 2015.
- [21] F. Magagna *et al.*, "Advanced fingerprinting of high-quality cocoa: Challenges in transferring methods from thermal to differential-flow modulated comprehensive two dimensional gas chromatography," *J. Chromatogr. A*, vol. 1536, pp. 122–136, 2018.
- [22] L. Nicolotti *et al.*, "Quantitative fingerprinting by headspace-Two-dimensional comprehensive gas chromatography-mass spectrometry of solid matrices: Some challenging aspects of the exhaustive assessment of food volatiles," *Anal. Chim. Acta*, vol. 798, pp. 115–125, 2013.
- [23] B. Sgorbini, C. Cagliero, E. Liberto, P. Rubiolo, C. Bicchi, and C. Cordero, "Strategies for Accurate Quantitation of Volatiles from Foods and Plant-Origin Materials: A Challenging Task," *J. Agric. Food Chem.*, vol. 67, no. 6, pp. 1619–1630, Feb. 2019.
- [24] H. H. Jeleń and M. N. Wieczorek, "Commentary: 'Quantitative' vs quantitative Headspace Solid-Phase Microextraction (HS-SPME) in food volatile and flavor compounds analysis," *J. Food Compos. Anal.*, vol. 115, no. September 2022, p. 104955, Jan. 2023.
- [25] P. Huhtanen *et al.*, "Prediction of the relative intake potential of grass silage by dairy

cows," *Livest. Prod. Sci.*, vol. 73, no. 2, pp. 111–130, 2002.

- [26] Å. T. Randby, I. Selmer-Olsen, and L. Baevre, "Effect of Ethanol in Feed on Milk Flavor and Chemical Composition," *J. Dairy Sci.*, vol. 82, no. 2, pp. 420–428, 1999.
- [27] Y. Wang, J. O'Reilly, Y. Chen, and J. Pawliszyn, "Equilibrium in-fibre standardisation technique for solid-phase microextraction," *J. Chromatogr. A*, vol. 1072, no. 1, pp. 13–17, Apr. 2005.
- [28] M. E. V. A. Canale, A. Ciotti, "Metodi analitici per la valutazione della qualità di conservazione degli insilati," 1983.
- [29] S. F. Spoelstra, M. G. Courtin, and J. A. C. Van Beers, "Acetic acid bacteria can initiate aerobic deterioration of whole crop maize silage," *J. Agric. Sci.*, vol. 111, no. 1, pp. 127–132, Aug. 1988.
- [30] L. Setkova, S. Risticvic, C. M. Linton, G. Ouyang, L. M. Bragg, and J. Pawliszyn, "Solid-phase microextraction-gas chromatography-time-of-flight mass spectrometry utilized for the evaluation of the new-generation super elastic fiber assemblies," *Anal. Chim. Acta*, vol. 581, no. 2, pp. 221–231, 2007.
- [31] B. Kolb and L. S. Ettre, *Static Headspace–Gas Chromatography*. New York: Wiley-VCH, 2006.
- [32] B. Kolb and L. S. Ettre, *Static headspace-gas chromatography : theory and practice*. New York: Wiley-VCH, 2006.
- [33] M. Giardina, J. D. McCurry, P. Cardinael, G. Semard-Jousset, C. Cordero, and C. Bicchi, "Development and validation of a pneumatic model for the reversed-flow differential flow modulator for comprehensive two-dimensional gas chromatography," *J. Chromatogr. A*, vol. 1577, pp. 72–81, 2018.
- [34] C. Cordero *et al.*, "Method translation and full metadata transfer from thermal to differential flow modulated comprehensive two dimensional gas chromatography: Profiling of suspected fragrance allergens," *J. Chromatogr. A*, vol. 1480, pp. 70–82, 2017.
- [35] W. Khummueng, J. Harynuk, and P. J. Marriott, "Modulation ratio in comprehensive two-dimensional gas chromatography," *Anal. Chem.*, vol. 78, no. 13, pp. 4578–4587, 2006.
- [36] T. Wenzl, J. Haedrich, A. Schaechtele, P. Robouch, and J. Stroka, *Guidance Document on the Estimation of LOD and LOQ for Measurements in the Field of Contaminants in Feed and Food. EUR 28099 EN*. 2016.
- [37] F. Magagna *et al.*, "Combined untargeted and targeted fingerprinting with comprehensive two-dimensional chromatography for volatiles and ripening indicators in olive oil," *Anal. Chim. Acta*, vol. 936, pp. 245–258, Sep. 2016.
- [38] M. Cialiè Rosso *et al.*, "Exploring extra dimensions to capture saliva metabolite fingerprints from metabolically healthy and unhealthy obese patients by comprehensive two-dimensional gas chromatography featuring Tandem Ionization mass spectrometry," *Anal. Bioanal. Chem.*, vol. 413, no. 2, pp. 403–418, Jan. 2021.
- [39] S. E. Reichenbach, P. W. Carr, D. R. Stoll, and Q. Tao, "Smart Templates for peak pattern

matching with comprehensive two-dimensional liquid chromatography,” *J. Chromatogr. A*, vol. 1216, no. 16, pp. 3458–3466, 2009.

- [40] F. Stilo, E. Liberto, S. E. Reichenbach, Q. Tao, C. Bicchi, and C. Cordero, “Untargeted and Targeted Fingerprinting of Extra Virgin Olive Oil Volatiles by Comprehensive Two-Dimensional Gas Chromatography with Mass Spectrometry: Challenges in Long-Term Studies,” *J. Agric. Food Chem.*, vol. 67, no. 18, pp. 5289–5302, May 2019.
- [41] A. Caratti *et al.*, “Augmented visualization by computer vision and chromatographic fingerprinting on comprehensive two-dimensional gas chromatographic patterns: Unraveling diagnostic signatures in food volatilome,” *J. Chromatogr. A*, vol. 1699, p. 464010, Jun. 2023.
- [42] K. Gerlach, K. Weiß, and K.-H. Südekum, “Effects of ethyl ester supplementation to forage on short-term dry matter intake and preference by goats,” *Arch. Anim. Nutr.*, vol. 73, no. 2, pp. 127–139, Mar. 2019.
- [43] L. Comino, E. Tabacco, F. Righi, A. Revello-Chion, A. Quarantelli, and G. Borreani, “Effects of an inoculant containing a *Lactobacillus buchneri* that produces ferulate-esterase on fermentation products, aerobic stability, and fibre digestibility of maize silage harvested at different stages of maturity,” *Anim. Feed Sci. Technol.*, vol. 198, pp. 94–106, Dec. 2014.
- [44] E. Aprea, F. Gasperi, E. Betta, G. Sani, and C. Cantini, “Variability in volatile compounds from lipoxygenase pathway in extra virgin olive oils from tuscan olive germoplasm by quantitative SPME/GC-MS,” *J. Mass Spectrom.*, vol. 53, no. 9, pp. 824–832, Sep. 2018.
- [45] P. M. Dewick, *Medicinal Natural Products*. 2009.
- [46] D. J. C. Jones, “The Biochemistry of Silage (2nd edn), by p. McDonald, A. R. Henderson & S. J. E. Heron. 340 pp. Kingston, Kent: Chalcombe Publications (1991). £49.50 (UK) £55.00 (elsewhere) (hardback). ISBN 0 948617 22 5,” *J. Agric. Sci.*, vol. 117, no. 3, pp. 386–386, Dec. 1991.
- [47] E. Tabacco, S. Piano, A. Revello-Chion, and G. Borreani, “Effect of *Lactobacillus buchneri* LN4637 and *Lactobacillus buchneri* LN40177 on the aerobic stability, fermentation products, and microbial populations of corn silage under farm conditions,” *J. Dairy Sci.*, vol. 94, no. 11, pp. 5589–5598, 2011.
- [48] C. Cordero *et al.*, “Potential of the reversed-inject differential flow modulator for comprehensive two-dimensional gas chromatography in the quantitative profiling and fingerprinting of essential oils of different complexity,” *J. Chromatogr. A*, vol. 1417, pp. 79–95, 2015.
- [49] I. Aloisi *et al.*, “Towards the determination of an equivalent standard column set between cryogenic and flow-modulated comprehensive two-dimensional gas chromatography,” *Anal. Chim. Acta*, vol. 1105, pp. 231–236, 2020.
- [50] E. Belhassen *et al.*, “Routine quantification of 54 allergens in fragrances using comprehensive two-dimensional gas chromatography-quadrupole mass spectrometry with dual parallel secondary columns. Part I: Method development,” *Flavour Fragr. J.*, vol. 33, no. 1, pp. 63–74, 2018.

- [51] J. Y. De Saint Laumer, E. Cicchetti, P. Merle, J. Egger, and A. Chaintreau, "Quantification in gas chromatography: Prediction of flame ionization detector response factors from combustion enthalpies and molecular structures," *Anal. Chem.*, vol. 82, no. 15, pp. 6457–6462, 2010.
- [52] M. S. Klee and L. M. Blumberg, "Theoretical and Practical Aspects of Fast Gas Chromatography and Method Translation," *J. Chromatogr. Sci.*, vol. 40, no. 5, pp. 234–247, 2002.
- [53] "Agilent Method Translation software." [Online]. Available: <https://www.agilent.com/en/support/gas-chromatography/gcmethodtranslation>.
- [54] M. Suzuki, S. Tsuge, and T. Takeuchi, "Gas Chromatographic Estimation of Occluded Solvents in Adhesive Tape by Periodic Introduction Method," *Anal. Chem.*, vol. 42, no. 14, pp. 1705–1708, 1970.
- [55] MCAULLIFE C, "GC determination of solutes by multiple phase equilibration," *Chem. Tech.*, no. 1, pp. 46–51, 1971.
- [56] B. Kolb and L. S. Ettre, *Static Headspace – Gas Chromatography : Theory and Practice Second Edition*, Second. New York: Wiley-VCH, 2006.
- [57] R. Costa, L. Tedone, S. De Grazia, P. Dugo, and L. Mondello, "Multiple headspace-solid-phase microextraction: An application to quantification of mushroom volatiles," *Anal. Chim. Acta*, vol. 770, pp. 1–6, 2013.
- [58] Ó. Ezquerro, G. Ortiz, B. Pons, and M. T. Tena, "Determination of benzene, toluene, ethylbenzene and xylenes in soils by multiple headspace solid-phase microextraction," *J. Chromatogr. A*, vol. 1035, no. 1, pp. 17–22, Apr. 2004.
- [59] W.-Q. Xie, Y.-X. Gong, and K.-X. Yu, "Enhancing the Sensitivity of Full Evaporation Technique Using Multiple Headspace Extraction Analysis," *Chromatographia*, vol. 80, no. 8, pp. 1263–1268, 2017.
- [60] C. Cordero *et al.*, "Comprehensive two-dimensional gas chromatography coupled with time of flight mass spectrometry featuring tandem ionization: Challenges and opportunities for accurate fingerprinting studies," *J. Chromatogr. A*, vol. 1597, pp. 132–141, Jul. 2019.
- [61] Commission Decision 657/2002, "Commission Decision 2002/657/EC implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results," *Off. J. Eur. Union*, vol. L221, no. 23 May 1996, pp. 8–36, 2002.
- [62] C. Cordero *et al.*, "Odorants quantitation in high-quality cocoa by multiple headspace solid phase micro-extraction: Adoption of FID-predicted response factors to extend method capabilities and information potential," *Anal. Chim. Acta*, vol. 1052, pp. 190–201, 2019.
- [63] E. Commission, "Commission Implementing Regulation (EU) 2021/808 of 22 March 2021 on the performance of analytical methods for residues of pharmacologically active substances used in food-producing animals and on the interpretation of results as well as on the methods to," *Off. J. Eur. Union*, vol. 180, no. 401, pp. 84–109, 2021.

Chapter 12

Beyond Current Quality Indices: Quantitative Volatilomics Unrevealed Cultivar Traits, Harvesting Practices Impact, and Aroma Blueprint of Extra-Virgin Olive Oils

Virgin olive oil, derived from *Olea europaea* L., is a staple in the Mediterranean diet and is classified into quality categories by EU regulations and the International Olive Council. Extra Virgin Olive Oil (EVOO), the highest quality, is valued for its nutritional and sensory attributes, driving consumer willingness to pay a premium. However, this makes EVOO susceptible to fraud, necessitating robust quality control methods. This study combines advanced untargeted fingerprinting and quantitative volatilomics using headspace solid-phase microextraction (HS-SPME) coupled to comprehensive two-dimensional gas chromatography with parallel mass spectrometry and flame ionization detection (GC×GC-MS/FID) on EVOOs from *Hojiblanca* and *Picual* cultivars to define robust markers of quality, named cultivar markers and cultivation practice indicators. The comprehensive analysis identified over 190 peak features and 84 compounds, revealing distinct volatile profiles influenced by cultivar, cultivation methods, and ripening stages. Of the 125 diagnostic volatile features (Fisher value > 4) for cultivar discrimination, alkenes, carbonyls, and alcohols predominate. Key compounds, such as (*E*)-2-octenal, (*E*)-2-hexenal, and (*Z*)-3-hexen-1-ol, reflect lipoxygenase pathway activity and maturation stages across cultivars. Quantification confirmed distinct aroma blueprints between cultivars, driven by differences in OAVs of key-odorants. However, minimal differences between odor active markers for organic and conventional cultivation practices, suggest consumers are unlikely to perceive variations in aroma between the two.

This chapter was published as Original Research Paper in the Journal of Food Composition and Analysis: [A. Caratti](#), A. Fina, F. Trapani, E. Liberto, B. Jiménez-Herrera, L. Arce, R.M. Callejón, C. Cordero, Beyond current quality indices: Quantitative volatilomics unveiled cultivar traits, harvesting practices impact, and aroma blueprint of extra-virgin olive oils, J. Food Compos. Anal. 137 (2025) 106975. DOI: 10.1016/j.jfca.2024.106975.

12.1 Introduction

Virgin olive oil, derived from *Olea europaea* L., is central to the Mediterranean diet and primarily produced in the Mediterranean region [1], [2]. Classified by organizations like Codex Alimentarius, International Olive Council, and subjected to EU Regulations [3], its quality categories depend on sensory and physicochemical parameters. Extra Virgin Olive Oil (EVOO), the highest quality category, is prized for its nutritional traits and sensory attributes [4], [5], driving consumer demand and willingness to pay premiums for specific qualities, such as health benefits, organic certification, or geographic origin [6]. However, this high demand also increases the risk of fraud, necessitating stringent quality control measures to maintain consumer confidence [7]–[9].

Characterizing EVOO involves analyzing a wide range of sensory and physicochemical parameters, with a growing need for advanced analytical methods due to the limited effectiveness of current techniques and evolving adulteration methods [7]. Technological advancements have been embraced by the olive sector to meet quality demands, with producers and researchers focusing on optimizing production under various agronomic and environmental conditions. Efforts aim to ensure the highest quality of olive oils despite these variables [10]–[12].

Aroma is a crucial quality criterion for EVOO, significantly influenced by factors such as olive variety/cultivar, geographic origin, fruit maturation, irrigation, crop season, agronomic practices, and oil extraction conditions [13]–[19]. The volatile profile and overall aroma are especially affected by olive variety/cultivar and ripening stage, driving interest in monovarietal oils. Despite promising analytical methods, no official approach exists to authenticate monovarietal EVOOs, highlighting the need for robust authentication methods [15], [20]–[22].

The type of farming, particularly organic versus conventional methods, is another significant factor due to consumer demand for organic products and strict agricultural regulations [23]. While organic olive farming has grown, studies comparing organic and conventional EVOOs are limited and often inconclusive due to uncontrolled variables. Future research needs controlled samples to advance knowledge on organic EVOOs, prevent fraud, and provide valuable information to the industry, producers, and consumers.

Gas chromatography-mass spectrometry (GC-MS) is widely used to profile olive oil volatile compounds. Combined with headspace (HS) approaches like solid-phase microextraction (SPME) or dynamic headspace (D-HS), it enables high-throughput profiling and monitoring of many aroma-active compounds [11], [12], [24]. The introduction of comprehensive two-dimensional GC (GC×GC) has significantly improved the understanding of EVOO's chemical complexity, providing higher peak capacity, resolution, lower detection thresholds, and more confident analyte identification [25]–[28]. Moreover, adopting non-targeted data processing [e.g., combined untargeted and targeted fingerprinting *UT fingerprinting* [29]], the investigation of complex fraction of volatile organic compounds (VOCs) by GC×GC has opened to volatilomics, the discipline that connects the chemical information encrypted in the volatilome (or volatome) with biological properties [30]–[32]. The volatilome “contains all of the volatile metabolites as well as other volatile organic and inorganic compounds that

originate from an organism" [33] super-organism, or ecosystem; as part of the metabolome it brings information worthy to be investigated for product valorization and traceability.

Analytical platforms that implement parallel detection, combining mass spectrometry (MS) and flame ionization detector (FID), offer great opportunities allowing for reliable identification and accurate quantification of analytes without external calibration. This enables true quantitative results for robust marker selection and correlation with quality traits [34], [35]. The use of GC×GC-MS/FID in olive oil research in particular, allows for a comprehensive mapping of detectable VOCs, provides accurate amounts on selected analytes, and enables causal correlation with biological variables such as olive cultivar, ripening stages, geographical origin, agricultural practices, and technological processing [29], [34]–[36]. The quantitative mapping of key-aroma compounds delineates the aroma blueprint of EVOOs, i.e., the unique pattern of odor active compounds responsible of the aroma identity of the product [37]

This study, for the first time, combines advanced untargeted fingerprinting and quantitative volatilomics using HS-SPME-GC×GC-MS/FID on a selection of monocultivar EVOOs to define both robust quality tracers and distinctive chemical blueprints predicting aroma features. Samples were obtained from *Hojiblanca* and *Picual* olive cultivars harvested in the Cordoba region (Spain) under different farming procedures (organic and conventional) and ripening stages from November 2021 to January 2022.

12.2 Materials and methods

12.2.1 Extra virgin olive oil samples

Extra virgin olive oils (EVOOs) were obtained from *Picual* and *Hojiblanca* olive cultivars harvested in 2021. Olive trees were subjected to two different cultivation methods, *organic* and *conventional*. Trees were located at the Agricultural Research Training Centre in “Cabra” in the province of Cordoba (Spain) [plot coordinates 37°29'36.4"N 4°25'45.8"W], which has a continental Mediterranean climate with dry summers and mild winters. Rainfall occurs from autumn to spring, with a mean annual rainfall of 400 mm. The average annual temperature is 17 °C, reaching -1.3 °C in winter and 43 °C in summer. Soils have low depths, being mainly over limestone and siliceous stones. They have a loamy texture and an alkaline pH (8-8.5 range). The electrical conductivity is low, as is the organic matter content and the carbon/nitrogen ratio. Twenty olive trees were randomly selected (always among the most loaded to guarantee sampling) of the *Picual* and *Hojiblanca* cultivars from conventional cultivation and twenty trees of the two cultivars from organic cultivation. Organic cultivation was conducted according to EU Regulations and certified by Instituto Andaluz de Investigación y Formación Agraria, Pesquera, Alimentaria y de la Producción Ecológica (IFAPA) (Junta de Andalucía) [38].

Olive harvesting was done by randomly hand-picking healthy fruits (without any type of infection or physical damage), beginning in November 2021 and ending in January 2022, and comprising different harvest periods. It conforms to a total of 3 ripening stages: 15 of November (stage I), 13 of December (stage II), and 1 of January (stage III). The ripening index

(RI) of each harvest was determined according to the methodology proposed by Uceda and Frias [39]. The fruit RI values for each modality are shown in **Table 12.1**.

In each harvesting stage, 200 kg of olives were collected without any discriminant selection before immediate processing in the experimental mill of IFAPA Venta del Llano (Mengíbar, Jaén, Spain) to obtain the oils. Thus, the extraction of the oils was carried out in a two-phase semi-industrial continuous system (Il Molinetto, Gruppo Pieralisi, Italy) equipped with a hammer mill, a thermo-blender, and a horizontal centrifuge. The 200 kg of olives were ground and then, the resulting paste was beaten for 45 min at 28 °C, after which the oil was separated by centrifuging the beaten paste at $\approx 3,250$ G. The oils obtained were decanted and filtered. Samples for volatilome screening were immediately sent to the University of Turin laboratory in dark-glass bottles. Before analysis they were kept at -18°C away from UV exposure. Sample descriptions and details are reported in **Table 12.1**.

Olive oils were also subjected to the evaluation of the sensory panel of the Priego de Córdoba Protected Denomination of Origin (PDO), following the standards of the International Olive Council (IOC) [40]. According to sensory results, all samples considered in this work resulted “Extra-Virgin” Olive Oils.

Table 12.1: EVOO samples characteristics and coding

Sample coding	Cultivar	Cultivation method	Olives harvesting	Maturity stage	Ripening Index
EVOO_1	<i>Picual</i>	Conventional	November 15	I	3.5
EVOO_2	<i>Picual</i>	Conventional	December 13	II	4.8
EVOO_3	<i>Picual</i>	Conventional	January 1	III	5.7
EVOO_4*	<i>Picual</i>	Conventional	December 13	II	4.8
EVOO_5	<i>Picual</i>	Organic	November 15	I	4.1
EVOO_6	<i>Picual</i>	Organic	December 13	II	4.8
EVOO_7	<i>Picual</i>	Organic	January 1	III	5.9
EVOO_8	<i>Hojiblanca</i>	Conventional	November 15	I	1.4
EVOO_9	<i>Hojiblanca</i>	Conventional	December 13	II	3.8
EVOO_10	<i>Hojiblanca</i>	Conventional	January 1	III	4.0
EVOO_11*	<i>Hojiblanca</i>	Conventional	December 13	II	3.8
EVOO_12	<i>Hojiblanca</i>	Organic	November 15	I	1.2
EVOO_13	<i>Hojiblanca</i>	Organic	December 13	II	2.6
EVOO_14	<i>Hojiblanca</i>	Organic	January 1	III	4.7

* Technological replicates from a different batch

12.2.3 Reference standards and solvents

Pure reference compounds for identity confirmation of marker volatiles and potent odorants listed in **Table 12.2**, *n*-alkanes (*n*-C₇ to *n*-C₂₅) for Linear Retention Indices (*I*^T) calibration, and α/β -thujone used as internal standards for response stability check were obtained from Merck (Milan, Italy).

The external standard for predicted relative response factors (RRF) based on combustion enthalpies calibration was hexanal (99% purity, Merck). The reference solution was prepared in cyclohexane (99% purity, Merck) at a final concentration of 10.00 g/L and then diluted in dibutyl phthalate (99% purity, Merck) to suitable concentrations for linearity evaluation.

Table 12.2: Identified compounds in the volatile fraction of EVOOs together with analytical information. See text for identification criteria.

Compound Name	¹ D t _R min	SD	² D t _R sec	SD	Exp I [†]	Lit. I [†]	MW	Formulae	PRRF	β (±SD)	Odor quality	OT (ng/g)*
<i>Analytes subjected to accurate quantification</i>												
Propanal	4.83	0.073	0.42	0.019	792	786	58.1	C ₃ H ₆ O	1.38	0.90 (±0.07)	Fresh, fruity, malty	9.4 [1]
Ethyl acetate	6.44	0.028	0.37	0.070	897	898	88.1	C ₄ H ₈ O ₂	1.60	0.92 (±0.08)	Fruity, sweet, winery	940 [3]
1-Penten-3-one	10.15	0.016	0.68	0.049	1017	1021	84.1	C ₅ H ₈ O	1.12	0.65 (±0.05)	Pungent, spicy	1.6 [1]
α-Pinene	10.24	0.087	1.68	0.055	1017	1017	136.2	C ₁₀ H ₁₆	0.79	0.93 (±0.07)	Herbal, woody,terpenic	274 [4]
Hexanal (Internal Standard for FID)	12.57	0.027	0.91	0.048	1079	1080	100.2	C ₆ H ₁₂ O	-	0.73 (±0.07)	Green, grass	300 [1]
β-Pinene	13.34	0.062	1.85	0.067	1097	1100	136.2	C ₁₀ H ₁₆	0.79	0.92 (±0.06)	Herbal, pine	-
(E)-2-Pentalenal	14.45	0.011	0.73	0.049	1123	1131	84.1	C ₅ H ₈ O	1.12	0.75 (±0.02)	Pungent, apple-like	300 [3]
2-Pentanol	14.99	0.016	0.44	0.052	1136	1138	88.1	C ₅ H ₁₂ O	1.02	0.88 (±0.11)	Musty, fermented	380 [2]
(Z)-3-Hexenal	15.14	0.020	0.77	0.050	1139	1133	98.1	C ₆ H ₁₀ O	1.04	0.56 (±0.03)	Green, grassy	1.7 [1]
δ-3-Carene	15.16	0.040	1.77	0.085	1139	1144	136.2	C ₁₀ H ₁₆	0.79	0.9 (±0.04)	Citrus, pine, herbal	770 [4]
1-Penten-3-ol	15.75	0.006	0.42	0.052	1153	1157	86.1	C ₅ H ₁₀ O	1.07	0.57 (±0.02)	Pungent, butter	400 [3]
Heptanal	16.99	0.024	1.02	0.062	1181	1181	114.2	C ₇ H ₁₄ O	0.96	0.93 (±0.05)	Citrus-like, fatty	500 [1]
Limonene	17.48	0.027	1.67	0.050	1192	1195	136.2	C ₁₀ H ₁₆	0.79	0.92 (±0.06)	Citrus, terpenic	250 [2]
3-Methyl-1-butanol	17.80	0.006	0.45	0.050	1199	1211	88.1	C ₅ H ₁₂ O	1.02	0.63 (±0.03)	Malty, ethereal	100 [3]
Eucalyptol	17.88	0.036	1.86	0.089	1201	1205	154.2	C ₁₀ H ₁₈ O	0.90	0.94 (±0.03)	Herbal, minty	15 [2]
(E)-2-Hexenal	18.40	0.006	0.79	0.051	1212	1216	98.1	C ₆ H ₁₀ O	1.04	0.6 (±0.02)	Bitter almond, green, fruity	320 [1]
Hexyl acetate	21.05	0.014	1.15	0.050	1271	1275	144.2	C ₈ H ₁₆ O ₂	1.09	0.7 (±0.08)	Fruity, pear-like	1040 [3]
Octanal	21.70	0.006	1.10	0.053	1285	1289	128.2	C ₈ H ₁₆ O	0.92	0.94 (±0.04)	Citrus-like, fatty	140 [1]
(Z)-2-Penten-1-ol	22.50	0.006	0.42	0.052	1303	1306	86.1	C ₅ H ₁₀ O	1.07	0.57 (±0.02)	Green, almond	250 [2]
(E)-2-Penten-1-ol	22.84	0.022	0.40	0.047	1310	1313	86.1	C ₅ H ₁₀ O	1.07	0.68 (±0.03)	Mushroom, earthy	250 [2]
(Z)-3-Hexenyl acetate	23.01	0.025	1.01	0.047	1315	1317	142.2	C ₈ H ₁₄ O ₂	1.12	0.55 (±0.04)	Green	200 [1]
(E)-2-Heptenal	23.17	0.025	0.86	0.031	1318	1320	112.2	C ₇ H ₁₂ O	0.99	0.81 (±0.05)	Green, fatty	1200 [1]
6-Methyl-5-hepten-2-one	23.80	0.006	0.95	0.054	1332	1339	126.2	C ₈ H ₁₄ O	0.95	0.69 (±0.03)	Pungent, green, fruity-like	1000 [3]
1-Hexanol	24.30	0.044	0.48	0.054	1343	1348	102.2	C ₆ H ₁₄ O	0.97	0.66 (±0.02)	Fruity, banana	400 [3]
(Z)-3-Hexen-1-ol	25.69	0.022	0.46	0.048	1374	1380	100.2	C ₆ H ₁₂ O	1.00	0.6 (±0.01)	Banana, fresh, grass	1100 [5]
(E,E)-2,4-Hexadienal	26.30	0.006	0.63	0.051	1388	1397	96.1	C ₆ H ₈ O	1.08	0.68 (±0.12)	Green, soapy	270 [2]
Nonanal	26.35	0.006	1.13	0.054	1389	1390	142.2	C ₉ H ₁₈ O	0.90	0.95 (±0.02)	Fatty, waxy, pungent	610 [1]
(E)-2-Hexen-1-ol	26.55	0.011	0.62	0.050	1394	1398	100.2	C ₆ H ₁₂ O	1.00	0.7 (±0.07)	Green, grass	500 [1]
(E)-2-Octenal	27.70	0.009	0.40	0.054	1420	1427	126.2	C ₈ H ₁₄ O	0.95	0.92 (±0.04)	Fatty, nutty	120 [1]
Acetic acid	28.51	0.023	0.24	0.051	1440	1440	60.1	C ₂ H ₄ O ₂	3.62	0.73 (±0.09)	Sour, vinegary	350 [1]
(E,E)-2,4-Heptadienal	29.20	0.006	0.70	0.055	1456	1461	110.2	C ₇ H ₁₀ O	1.02	0.86 (±0.05)	Fatty, green, oily	710 [1]
α-Copaene	30.55	0.016	2.37	0.052	1489	1491	204.4	C ₁₅ H ₂₄	0.78	0.94 (±0.06)	Woody	-
Benzaldehyde	31.59	0.022	0.66	0.051	1513	1524	106.1	C ₇ H ₆ O	0.92	0.91 (±0.04)	Almond, burnt sugar, erthy	60 [4]
Popanoic acid	32.15	0.013	0.26	0.055	1526	1534	74.1	C ₃ H ₆ O ₂	2.06	0.81 (±0.05)	Acidic, pungent	720 [4]
(E)-2-Nonenal	32.26	0.026	0.91	0.058	1530	1534	140.2	C ₉ H ₁₆ O	0.92	0.95 (±0.02)	Fatty, green, soapy	140 [1]
1-Octanol	33.01	0.022	0.55	0.055	1548	1553	130.2	C ₈ H ₁₈ O	0.90	0.86 (±0.03)	Nut, mushroom	27 [2]
Butanoic acid	35.75	0.006	0.25	0.053	1616	1613	88.1	C ₄ H ₈ O ₂	1.60	0.85 (±0.05)	Cheesy, sour	34 [1]
(E,E)-α-Farnesene	40.60	0.015	1.56	0.051	1742	1744	204.4	C ₁₅ H ₂₄	0.78	0.78 (±0.03)	Sweet, floral	-

Table 12.2: continue

Compound Name	¹ D t _R min	SD	² D t _R sec	SD	Exp I ^T	Lit. I ^T	MW	Formulae	PRRF	β (±SD)	Odor quality	OT (ng/g)*
<i>Analytes not quantified</i>												
Hexane	3.58	0.135	0.45	0.049	600	600	86.2	C ₆ H ₁₄	-	-	Alkane	-
Octane	4.87	0.066	0.99	0.063	800	800	114.2	C ₈ H ₁₈	-	-	Solvent	940 [5]
Acetone	5.05	0.017	0.45	0.049	850	845	58.1	C ₃ H ₆ O	-	-	Pungent	-
2-Methyl butanal	7.08	0.140	0.64	0.073	924	926	86.1	C ₅ H ₁₀ O	-	-	Malty	10 [5]
Ethanol	7.35	0.015	0.37	0.051	928	933	46.1	C ₂ H ₆ O	-	-	Ethanol-like	-
3,4-Diethyl-1,5-hexadiene (RS/SR)	7.90	0.018	1.21	0.049	946	952	138.2	C ₁₀ H ₁₈	-	-	-	-
3-Methylbutanal	8.00	0.133	0.72	0.066	939	930	86.1	C ₅ H ₁₀ O	-	-	Malty	13 [5]
3,4-Diethyl-1,5-hexadiene (meso)	8.45	0.145	1.40	0.083	956	956	138.2	C ₁₀ H ₁₈	-	-	-	-
Pentanal	8.76	0.027	0.75	0.050	975	974	86.1	C ₅ H ₁₀ O	-	-	Almond-like, pungent, malt	150 [4]
(Z)-1-Methoxy-3-hexene	9.74	0.025	1.17	0.050	1007	997	114.2	C ₇ H ₁₄ O	-	-	-	-
(5Z)-3-Ethyl-1,5-octadiene	9.93	0.026	1.69	0.050	1012	1006	138.2	C ₁₀ H ₁₈	-	-	-	-
(5E)-3-Ethyl-1,5-octadiene	10.47	0.027	1.69	0.049	1026	1032	138.2	C ₁₀ H ₁₈	-	-	-	-
Toluene	10.73	0.050	0.91	0.057	1029	1036	92.1	C ₇ H ₈	-	-	Chemical-like	-
(E,Z)-3,7-Decadiene	12.52	0.028	1.92	0.051	1078	1068	138.2	C ₁₀ H ₁₈	-	-	-	-
(E,E)-3,7-Decadiene	12.85	0.016	1.89	0.049	1085	1082	138.2	C ₁₀ H ₁₈	-	-	-	-
(Z)-2-Pentenal	13.49	0.017	0.74	0.049	1101	1105	84.1	C ₅ H ₈ O	-	-	Pungent, apple-like	300 [3]
3-Methylbutyl acetate	14.42	0.053	0.97	0.060	1123	1126	130.2	C ₇ H ₁₄ O ₂	-	-	Banana-like, fruity	-
2-Heptanone	16.85	0.027	0.98	0.086	1178	1171	130.2	C ₇ H ₁₄ O	-	-	Sweet, fruity	470 [4]
(Z)-2-Hexenal	17.66	0.022	0.82	0.048	1197	1193	98.1	C ₆ H ₁₀ O	-	-	Fruity	-
γ-Terpinene	19.59	0.024	1.71	0.058	1239	1247	98.1	C ₆ H ₁₀ O	-	-	Petrol-like	-
trans-β-Ocimene	19.95	0.006	1.44	0.050	1246	1249	136.2	C ₁₀ H ₁₆	-	-	Citrus-like, soapy, geranium-like	-
o-Cymene	20.72	0.027	1.42	0.050	1264	1268	134.2	C ₁₀ H ₁₄	-	-	-	-
3-Hydroxy-2-butanone	21.45	0.021	0.46	0.059	1279	1283	88.1	C ₄ H ₈ O ₂	-	-	Buttery	-
2-Octanone	21.50	0.014	1.07	0.052	1280	1284	128.2	C ₈ H ₁₆ O	-	-	Mould, green	510 [4]
(E)-3-Hexen-1-ol acetate	22.49	0.020	1.57	0.050	1303	1300	142.2	C ₈ H ₁₄ O ₂	-	-	Fruity	-
Dioxa-1,6-spiro[4.5]decane	23.54	0.032	1.44	0.077	1326	-	142.2	C ₈ H ₁₄ O ₂	-	-	-	-
3-Methyl-cyclopentanol	23.68	0.034	0.65	0.064	1329	1342	100.1	C ₆ H ₁₂ O	-	-	-	-
(E)-3-Hexen-1-ol	24.78	0.031	0.47	0.044	1354	1352	100.2	C ₆ H ₁₂ O	-	-	green, grassy	-
Methoxymethyl-benzene	26.05	0.006	0.93	0.053	1382	1382	122.2	C ₈ H ₁₀ O	-	-	-	-
(Z)-2-Hexen-1-ol	26.62	0.035	0.43	0.029	1396	1405	101.2	C ₆ H ₁₂ O	-	-	Green, grass	-
Dimethyl Sulfoxide	32.60	0.015	0.49	0.051	1537	1549	78.1	C ₂ H ₆ OS	-	-	-	-
3-Methyl-2-hexen-4-one	34.60	0.000	0.47	0.054	1587	1586	112.2	C ₇ H ₁₂ O	-	-	-	-
Benzoic acid, methyl ester	35.66	0.028	0.75	0.060	1613	1614	136.1	C ₈ H ₈ O	-	-	Starfruit-like, sweet	-
1,4-Cyclohex-2-enedion	39.93	0.025	0.53	0.047	1725	-	110.1	C ₆ H ₆ O ₂	-	-	-	-
Pentanoic acid	39.93	0.024	0.25	0.050	1725	1730	102.1	C ₅ H ₁₀ O ₂	-	-	Sweaty, fruity	400 [1]
2,3-Dimethylbenzaldehyde	39.99	0.052	0.66	0.049	1725	1736	134.2	C ₉ H ₁₀ O	-	-	-	-
5-ethyl-2(5H)-Furanone	40.73	0.024	0.51	0.045	1746	1757	112.1	C ₆ H ₈ O ₂	-	-	Sweet, spicy	-
Methyl salicylate	41.53	0.029	0.72	0.046	1768	1765	152.1	C ₈ H ₈ O ₃	-	-	Fatty, tallowy, terpene-like	-
2-(2-butoxyethoxy)Ethanol	42.09	0.024	0.55	0.054	1783	1786	162.2	C ₈ H ₁₈ O ₃	-	-	-	-

Table 12.2: continue

Compound Name	¹ D t _R min	SD	² D t _R sec	SD	Exp I ^r	Lit. I ^r	MW	Formulae	PRRF	β (±SD)	Odor quality	OT (ng/g)*
Hexanoic acid	43.84	0.022	0.26	0.049	1831	1839	116.2	C ₆ H ₁₂ O ₂	-	-	Goat-like, sweaty	460 [1]
Benzyl alcohol	44.78	0.025	0.38	0.046	1858	1857	108.1	C ₇ H ₈ O	-	-	Sweet, fruity	-
Phenylethyl alcohol	46.02	0.029	0.44	0.053	1893	1904	122.2	C ₈ H ₁₀ O	-	-	Floral, honey-like	-
(E)-2-Hexenoic acid	47.89	0.022	0.30	0.047	1947	1941	114.1	C ₆ H ₁₀ O ₂	-	-	-	-
Phenol	49.10	0.006	0.26	0.054	1983	1987	94.1	C ₆ H ₆ O	-	-	Ink-like, phenolic	-
Benzoic acid, 2-methoxy-, methyl ester	51.31	0.045	0.66	0.060	2049	2032	166.2	C ₉ H ₁₀ O ₃	-	-	-	-
*Odor threshold references: [1] [41]; [2] [42]; [3] [43]; [4] [44]; [5] [27] Retention times (¹t_R min, ²t_R sec), standard deviation (SD), experimental and tabulated linear retention indexes (I^r), molecular weight (MW), predicted FID relative retention factor (RRF), decay function (β), odor threshold (OT) in ng/g												

12.2.4 Multiple headspace solid phase microextraction: devices and conditions

Volatiles from EVO oils were extracted by HS-SPME with a divinylbenzene/carboxen/polydimethyl siloxane (DVB/CAR/PDMS) fiber (d_f 50/30 μm ; 2 cm length) from Supelco (Bellefonte, PA, USA) according to a previously optimized procedure [45]. The SPME fiber was conditioned before use as recommended by the manufacturer. Sampling was carried out on 0.100 ± 0.003 g of oil, precisely weighed in a 20 mL headspace vial, and kept at 40 °C for 60 min under constant agitation. The very low amount of sample was chosen to match for HS linearity conditions for most of the characteristic analytes of the EVO oil volatilome [34]. After extraction, the SPME device was automatically transferred to the split/splitless injection port of the GC $\times$ GC system, kept at 250 °C, and thermal desorption was for 5 min. Samples were analyzed in four replicates randomly distributed over two weeks.

Multiple headspace extraction by SPME (MHS-SPME) of samples and calibration solutions were conducted by applying the above-indicated conditions, and the number of consecutive extraction steps was set to four, achieving an almost exhaustive extraction of the analytes under study [34]

12.2.5 Comprehensive two-dimensional gas chromatography: instrument set-up and conditions

Automated MHS-SPME was performed by a multipurpose sampler, model MPS-2 (Gerstel, Mülheim a/d Ruhr, Germany), installed on a GC $\times$ GC system equipped with a reverse-inject differential-flow modulator based on capillary flow technologyTM (Agilent Technologies, Little Falls, DE, USA). The Agilent 7890B GC unit was coupled to an Agilent 5977B equipped with a high-efficiency source (HES) and fast quadrupole MS analyzer (Agilent Technologies). The MS was operating in electron ionization mode at 70 eV. Ion source and transfer-line temperatures were set at 280 °C, and the quadrupole temperature was set at 240 °C. The scan range was set between 40 and 250 m/z , achieving a data acquisition frequency of 30 Hz. Parallel detection was by a fast FID with base temperature held at 280 °C; H_2 flow 40 mL/min, air flow 350 mL/min, and sampling frequency 200 Hz.

The column set was configured as follows: first dimension (^1D) HeavyWaxTM column (100% polyethylene glycol - PEG; 20 m $\times$ 0.18 mm d_c $\times$ 0.18 μm d_f) coupled with second dimension (^2D) DB17 column (50% phenyl-methylpolysiloxane; 1.8 m $\times$ 0.18 mm d_c $\times$ 0.18 μm d_f), both from Agilent Technologies. The connection between the ^2D column and deactivated silica capillaries (Agilent Technologies) toward MS (0.5 m $\times$ 0.1 mm d_c) and FID (1.1 m $\times$ 0.18 mm d_c) for parallel detection was by a three-way unpurged capillary microfluidic splitter (G3181B, Agilent Technologies). The resulting split ratio was 70:30 FID/MS.

The GC split/splitless injector port was set at 250 °C and operated in pulsed-split mode (250 kPa overpressure applied to the injection port until 2 min) with a split ratio 1:20. The carrier gas was helium at a nominal flow of 0.4 mL/min along the ^1D column and 10 mL/min along the ^2D column. The oven temperature program was set as: from 40 °C (2.29 min) to 240 °C (11') at 3.06 °C/min. The modulation period (P_M) was set at 3s and pulse time at 250 ms. The n -alkanes liquid sample solution for I^T determination was analyzed under the following

conditions: split/splitless injector in split mode, split ratio 1:50, injector temperature 250 °C, and injection volume 1 µL.

12.2.6 Analytes identification criteria

Analyte identification was by combining retention data (carried out using experimental I^T with ± 10 units tolerance vs. tabulated ones) and comparing electron ionization (EI)-MS spectral signature with reference spectra in commercial and in-house databases by using the NIST identity search algorithm with direct match factor (DMF) and reverse match factor (RMF) scores threshold ≥ 900 .

12.2.7 Analytical data acquisition, processing and mining

Combined Untargeted and Targeted (UT) fingerprinting by smart templates

A template is a pattern of 2D peaks and/or graphic objects (features) created over a reference chromatogram or image (single or cumulative). This template is used to identify similar patterns of 2D peaks in a set of analyzed chromatograms or images [36], [46]. The process involves specific matching functions that establish correspondences between features across multiple chromatograms. Specificity is achieved by defining confidence thresholds for retention times and MS spectral similarity to account for variable peak detection, and by using suitable transform functions to correct for retention time inconsistencies between runs [45], [47]. Once the correspondences between features are established, the template's metadata (chemical name, retention times, mass spectra, informative ions, and their relative ratios) are transferred to candidate peaks and/or graphic objects (peak regions) in the analyzed chromatogram. This feature template, which includes untargeted reliable peaks and peak regions, is obtained through a fully automated workflow in GC Image Investigator™ (GC Image™, GC Image LLC).

The workflow consists of the following steps:

- (a) Match peak patterns between all chromatogram pairs from a set of representative samples, using optimized parameters: S/N threshold of 50, MS constraint of 700 for both the direct match factor (DMF) and reverse match factor (RMF) based on the NIST similarity algorithm [47], [48].
- (b) Select reliable peaks across the analyzed chromatograms, using a relaxed reliability criterion that includes peaks matching at least 50% +1 of the chromatograms [49]. This template of reliable peaks is used for re-aligning chromatograms in the temporal domain.
- (c) Align chromatograms of representative samples to the average retention times of the reliable peaks, then sum/fuse them into a composite chromatogram.
- (d) Generate a comprehensive untargeted feature template from the composite chromatogram, including reliable peaks and peak regions defined by the footprint of all detected peaks [49].

This process was used to generate a composite image from a selection of sample chromatograms representative of the different expressions of functional variables. From the composite image, a feature template consisting of 190 UT features (i.e., peak regions

corresponding to untargeted and targeted components) was defined and applied to single chromatogram images (*i.e.*, samples and replicates) for reliable tracking and alignment of features. The list of targeted features is reported in **Table 2** together with chemical names, ¹D and ²D retention times (¹*t_R*; ²*t_R*), experimental *I^T*, tabulated *I^T*, odor quality, and odor thresholds (OTs) in oil as from reference literature.

12.2.8 Data acquisition and statistical analysis

Raw chromatographic data were acquired by MassHunter Workstation (Agilent Technologies). Raw data were processed by GC Image™ V2020 r1.2 suite (GC Image, LLC Lincoln, NE, USA).

Statistical analysis and chemometrics were performed using GC Investigator™ (GC Image), XLSTAT statistical and data analysis solution (Addinsoft 2020, New York, USA), and Microsoft Office Excel 2016 (Microsoft, Redmond, WA USA).

12.3 Results and discussion

EVO oil *volatilome* encrypts information on many key-quality variables (*e.g.*, olives cultivar and geographical origin, cultivation methodologies, olives harvesting stage, oil auto-oxidation level, and sensory profile [43], [50], [51], by GC×GC it was possible to consistently monitor more than 190 *reliable* peak features (including targeted and untargeted compounds [29]), of which 84 were identified according to ¹D retention index (*I^T* ± 10 units of tolerance) combined to EI-MS spectral similarity (Direct Match Factor threshold >900) with commercial and in-house databases. **Table 12.2** lists putatively identified compounds (*i.e.*, targeted features) and potent odorants together with retention data, odor qualities, odor thresholds (OTs) in oil, and FID predicted RRFs calculated according to hexanal as reference internal standard. **Table S1**, provided as supplementary material, lists targeted and untargeted features comprehensively mapping the detectable volatilome of analyzed samples.

As a first approach to understanding the existence of natural samples' clusters according to volatile patterns distribution, unsupervised exploration by Principal Component Analysis (PCA) was conducted. **Fig. 12.1A** illustrates PCA results obtained by examining the data matrix including untargeted and targeted (UT) peak features (analytes) % responses for all *Picual* oils (blue indicators) and *Hojiblanca* oils (green indicators) (190 × 32 – features × samples replicates). It represents a total explained variance of 41.62% (PC₁ and PC₂) and, along the PC₁ with 31.32% of the total explained variance, the two *cultivars* form natural clusters as shown by ellipses, set at 95% confidence. Besides some outliers, the cultivars' chemical signature is almost clear confirming that the volatile fraction *in toto* encrypts information on olives' chemotype useful for their discrimination. This result agrees with other authors who observed that botanical traits have the most influence on the volatile profile of EVOOs [40].

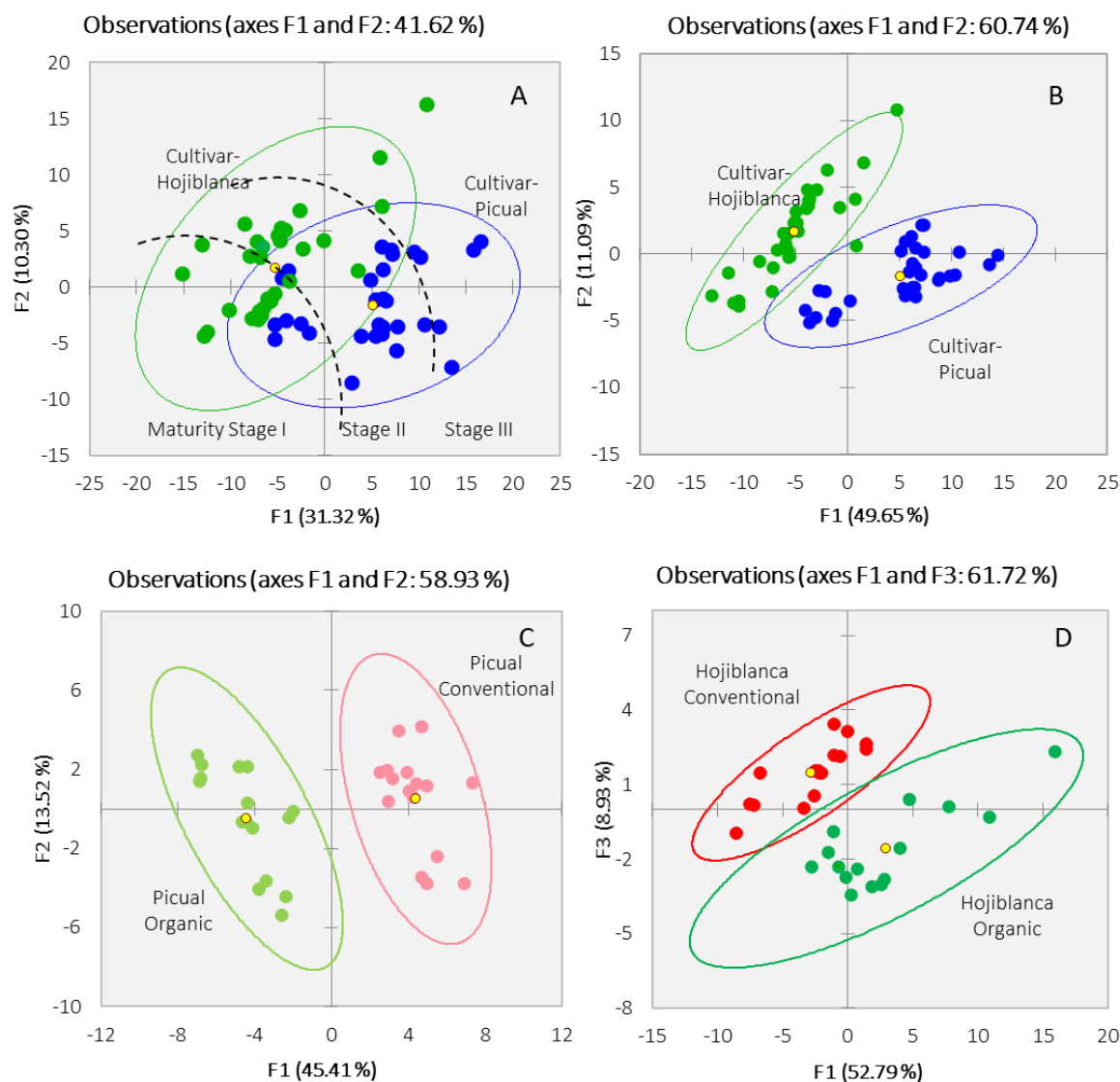


Figure 12.1. PCA on untargeted and targeted (UT) peak features (analytes) % responses; ellipses set at 95% of confidence level, show cultivar natural clusters. In **Fig. 1A** all 190 UT features were computed for all samples; dotted lines indicate Maturity Stage evolution according to sample loadings. **Fig. 1B** results from Fisher ratio features reduction considering cultivar discrimination (*Picual* vs. *Hojiblanca*) the PCA refers to 92 UT peaks with a $F > 10$. **Fig. 1C** Fisher ratio features reduction considering cultivation type on *Picual* oils (organic vs. conventional) the PCA refers to 49 UT peaks with a $F > 10$. **Fig. 1D** Fisher ratio features reduction considering cultivation type on *Hojiblanca* oils (organic vs. conventional) the PCA refers to 49 UT peaks with a $F > 10$.

Nevertheless, a certain degree of overlap between the two clusters suggests that for some samples the volatilome has similar composition, possibly due to the influence of the harvest region with its common pedoclimatic and soil characteristics. However, the simultaneous presence of many variables influencing the volatiles' qualitative and quantitative profiles, such as the ripening stage of the fruits, might have a confounding effect. According to previous investigations, the ripening stage of the fruit was, after the variety/cultivar, the most influential factor on the volatile fraction [29], [40], [52]. A deeper observation of the samples' distribution on the Cartesian plane (Fig. 1A), indicates a clear trend along PC₁ with oils obtained by olives harvested at Stage I characterized by lower loadings and those obtained by

olives at Stage III with higher values. This is in accordance to the increasing ripening index as reported in **Table 12.1**. Analytes responsible for this distribution are, among the others: (*E,E*)- α -farnesene, (*5Z*)-3-ethyl-1,5-octadiene, octane, (*E,Z*)-3,7-decadiene, 3,4-diethyl-1,5-hexadiene (meso), 3,4-diethyl-1,5-hexadiene (RS+SR), (*Z*)-3-hexenal, (*Z*)-2-hexenal, phenylethyl alcohol, 6-methyl-5-hepten-2-one, (*Z*)-3-hexen-1-ol, (*E,E*)-3,7-decadiene, 1-penten-3-one, 1-penten-3-ol, and nonanal. Results are in keeping with previous studies on ripening indices [29], [52], [53].

To identify cultivar discriminating variables, the UT features % response data were filtered by the Fisher ratio (F) value criterion [54]. Since $F_{crit}(1,15)$ with $\alpha=0.05$ is 4.54, UT variables with a $F_{calc} > 10$ (arbitrarily set above the critical value) were retained ($n=92$ UT features), and by applying this features selection criterion, the resulting PCA shown in **Fig. 12.1B** clearly clusters samples according to the cultivar type. The total explained variance raises 60.74% (PC1 and PC2) and samples are now independently clustered and discriminated along the PC1 (49.65% of the total explained variance). F values referred to UT features discriminating cultivars are reported in **Table S1**.

The impact of cultivation methodology (i.e., conventional vs. organic) was then explored. The PCA was conducted on the data matrix including UT peak features (analytes) % responses with an F value > 10 for *organic vs. conventional* classes. F values discriminating cultivation methods are listed in **Table S1**. PCA loadings plot based on the 35 features with $F > 10$ is shown in **Fig. 12.2**. Results confirm the major role played by cultivar traits on the volatilome expression, with a total explained variance of 61.58% the first two PCs do not show any independent clustering for organic and conventional cultivated olive tree samples. However, for *Picual* oils, a distinction of the two sub-groups along the PC1 appears (pink and light green indicators).

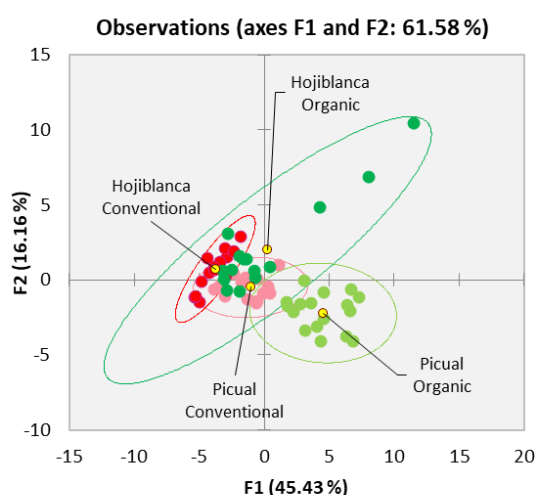


Figure 12.2. PCA on untargeted and targeted (UT) peak features (analytes) % responses after Fisher ratio reduction according to classes organic vs. conventional and $F > 10$ (35 UT features). Red indicators are for *Hojiblanca* conventional, pink is for *Hojiblanca* organic, dark green is for *Picual* conventional, and light green is for *Picual* organic.

Due to the concurrent effect of external variables, as cultivar and cultivation practices, the identification of marker volatiles for organic cultivation was conducted on samples belonging to the same cultivar with the aid of supervised strategies.

12.3.1 *Picual* olive oils: Organic vs. conventional cultivation diagnostic volatiles signature

Considering the *Picual* cultivar, the unsupervised statistics on the two cultivation methodologies ($F_{\text{calc}} > 5$ for organic vs. conventional classes) show two independent groups along PC₁, with a total explained variance of 58.93% (e.g., **Fig. 12.1C**). At the same time, heatmap visualization (e.g., **Fig. S1A**) captures diagnostic patterns of UT volatiles and by hierarchical clustering (HC) based on Pearson correlation of variables form two independent clusters corresponding to cultivation methodologies. Sub-clusters are coherent with biological and technical replicates.

To select statistically relevant yet informative compounds related to cultivation methodologies on *Picual* oils, supervised analysis by partial least squares-discriminant analysis (PLS-DA) was adopted [55]. The model was developed on an estimation set consisting of 70% randomly selected measures/samples (20 samples over 28) and validated on the residual 30% (8 samples over 28). Model performances, after 10 reiterations, were good referring of a 98% of correctness. The variable importance in the projection scores (VIPs) were used to identify meaningful variables, resulting from the classification model on *Picual* EVO oils conventional vs. organic, they are visualized as histogram in **Fig. 12.3** (orange bars). **Table S2** lists VIPs values together with relative standard deviation. Discriminant analytes between the cultivation methodologies belongs to the class of potent odorants with 2-pentanol, 1-hexanol, (Z)-3-hexen-1-ol likely discriminating the *Hojiblanca* cultivar from the *Picual* which has a characteristic pattern of odorants with (E)-2-penten-1-ol, heptanal, and (E)-2-pentenal. Variables showing a role in the discrimination are also validated by HC results. Of interest are carbonyls [2-heptanone, 2-octanone, heptanal, (E)-2-pentenal] and alkenes [(5E)-3-ethyl-1,5-octadiene, (E,E)-3,7-decadiene, 3,4-diethyl-1,5-hexadiene (RS+SR), (5Z)-3-ethyl-1,5-octadiene, (E,Z)-3,7-decadiene]; this last class of chemicals known for their correlation with olives ripening [53].

The relative distribution of the most informative variables is illustrated by box-plots in **Fig. 12.4**. Analytes with higher % response in conventional cultivation are methoxymethylbenzene [α -methylbenzyl ether already documented[56]], (5E)-3-ethyl-1,5-octadiene, (5Z)-3-ethyl-1,5-octadiene, 3,4-diethyl-1,5-hexadiene (RS+SR), (E,E)-3,7-decadiene, (E,Z)-3,7-decadiene, (Z)-3-hexenyl acetate, (E)-3-hexen-1-ol acetate, and α -copaene. Compounds with an opposite trend are 2-heptanone, 2-octanone, and butanoic acid.

According to the literature, (5E)-3-ethyl-1,5-octadiene, (5Z)-3-ethyl-1,5-octadiene, 3,4-diethyl-1,5-hexadiene (RS or SR), (E, E)-3,7-decadiene and (E, Z)-3,7-decadiene have been found in *Picual* EVOOs and were considered as markers of early ripening stages, independently of the variety/cultivar and geographical origin [52]. In addition, 3-ethyl-1,5-octadiene, a compound deriving from the lipoxygenase pathway, recorded significant variations according to the cultivar and/or environmental conditions on olive tree growing [57].

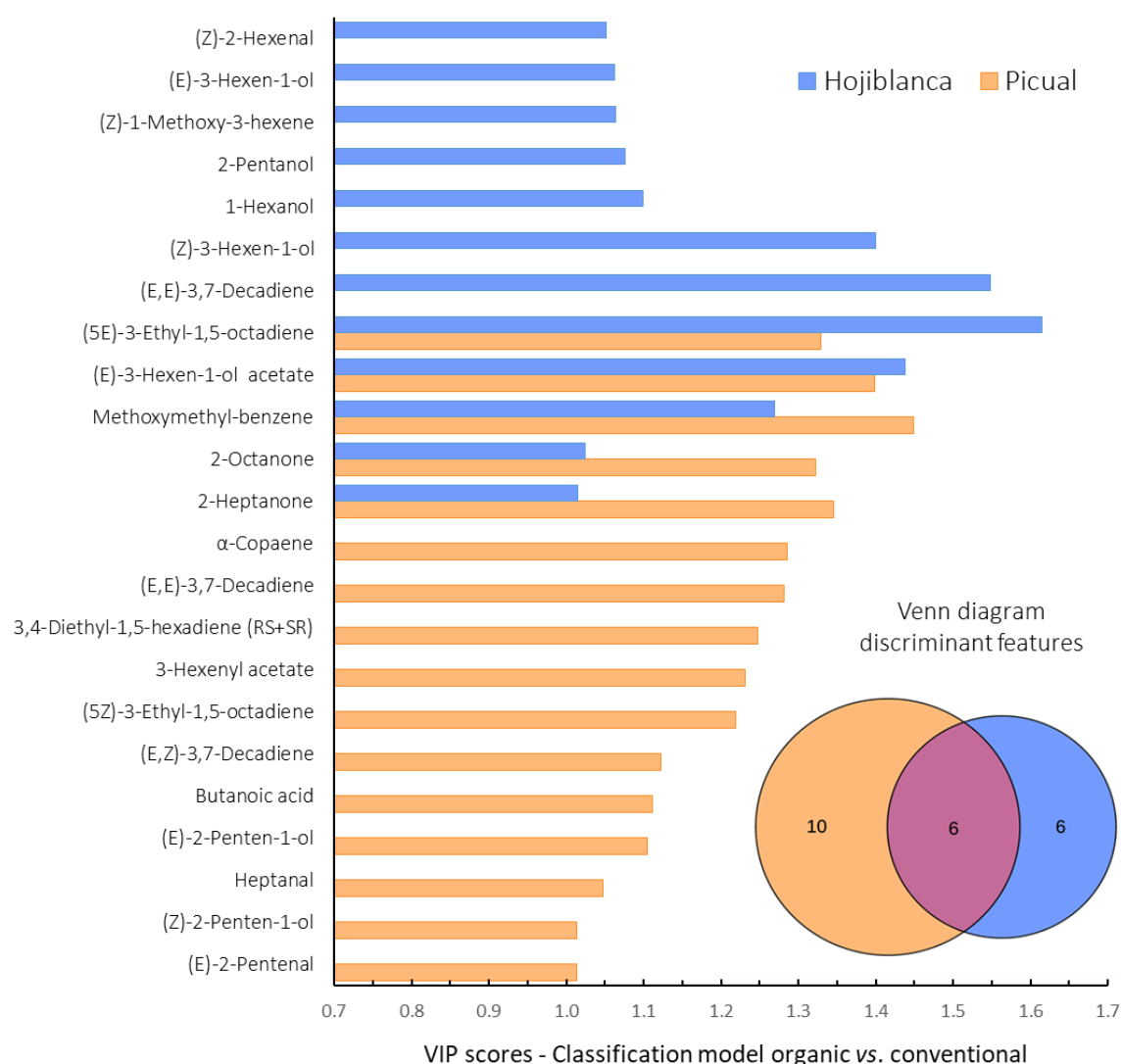


Figure 12.3: Histogram showing the VIPs values derived by the PLS-DA classification models conventional vs. organic classes. Blue bars are for *Hojiblanca* and orange bars for *Picual*. The Venn diagram visualizes the number of targeted analytes (features) in common ($n=6$) or unique ($n=10$ for *Picual* and $n=6$ for *Hojiblanca*) for the two cultivars.

(Z)-3-Hexenyl acetate is a compound responsible for sensory attributes as *green leaves*, and it was shown to be relevant in the differentiation between non-defective (EVO) and defective (non-EVO) olive oil samples, being present in higher concentration in EVOO samples [40]. Non-defective (EVO) olive oil is of the highest quality and free of sensory defects, while defective (non-EVO) olive oil does not meet these strict standards and is of lower quality. In addition, aligned with the current evidence, (Z)-3-hexenyl acetate was differentially distributed in organic and conventional samples considered in a previous study [58].

α -Copaene is a mono-unsaturated sesquiterpene that has already been detected in Spanish oils obtained from olives of the *Hojiblanca* cultivar and *Picual* [59], [60], and together with α -muurolene and α -farnesene were the terpenes that aided the discrimination between extra virgin olive oil according to the cultivar and geographical origin [61]–[63] and they have been

suggested as markers of olive oil differentiation [57]. According to the literature, butanoic acid has been associated with *vinegary*, *musty*, and *rancid* defects [64], [65].

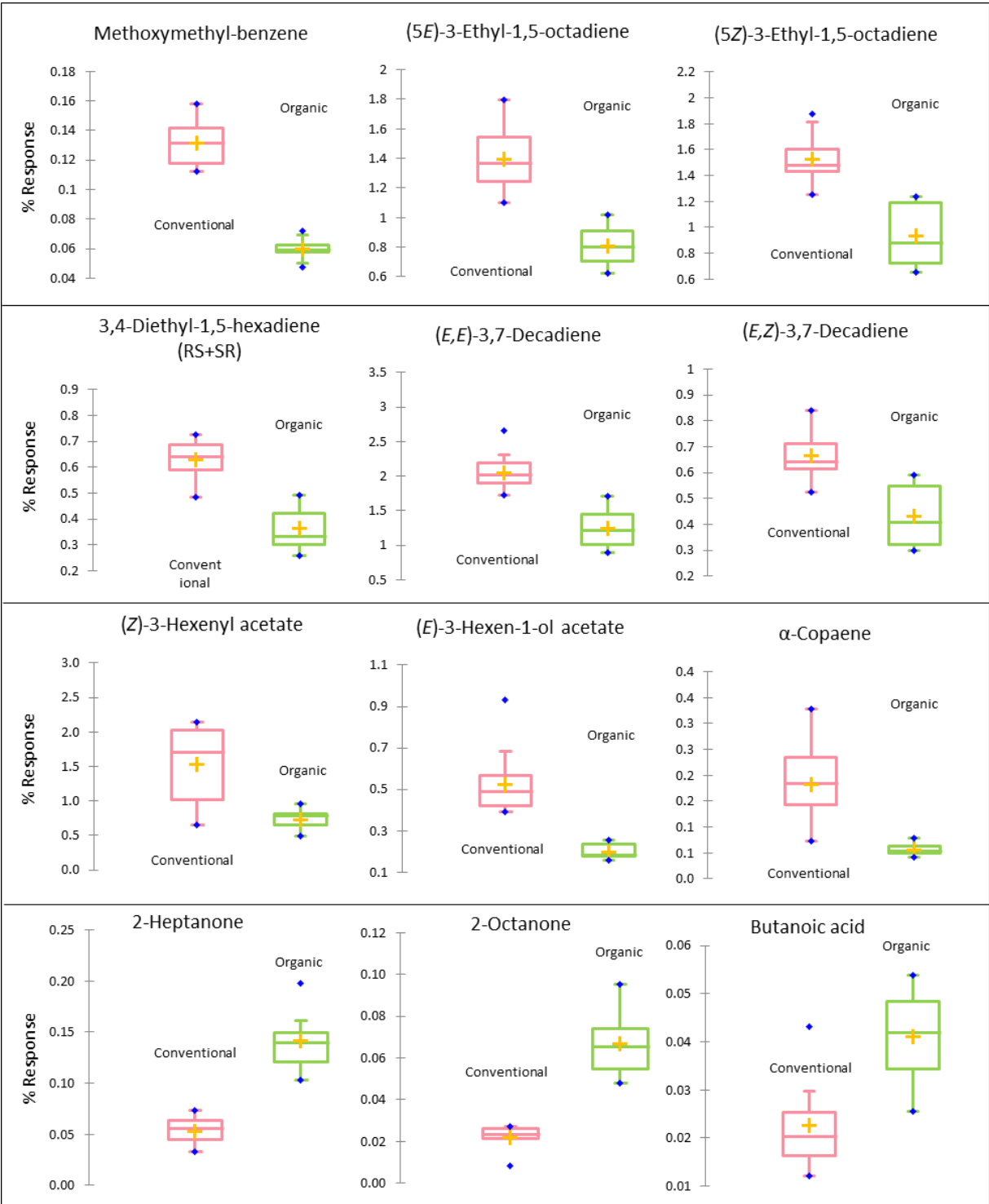


Figure 12.4. Box plots of discriminant compounds (from PLS-DA modeling) between organic and conventional cultivation on *Picual* cultivar samples. Green boxes represent data from organic farming (mean/median) while pink boxes are for conventional farming (mean/median).

12.3.2 *Hojiblanca* olive oils: Organic vs. conventional cultivation diagnostic volatiles signature

A first exploratory approach was conducted on *Hojiblanca* samples considering the cultivation practice. **Fig. 12.1D** shows the PCA resulting from UT features filtered by Fisher ratio value ($F_{\text{calc}} > 10$). Samples, as expected, are clearly clustered according to cultivation practice and the total explained variance for PC₁-PC₃ achieves 61.72%. At the same time, heatmap visualization (provided as supplementary material **Fig. S1**) captures diagnostic patterns of UT volatiles and by hierarchical clustering (HC) based on Pearson correlation of variables form two independent clusters corresponding to cultivation methodologies. Sub-clusters are coherent with biological and technical replicates. Chemical variables highlighted by PLS-DA as distinctive for the cultivation methods in *Hojiblanca* EVO oil (VIPs > 1) are listed in **Table S2** together with VIPs value and relative standard deviation. The model was developed on an estimation set of samples consisting of 70% randomly selected measures/samples (20 samples over 28) and validated on the residual 30% of samples (8 samples over 28). Model performances, after 10 reiterations, were good referring of a 90% of correctness.

Some analytes capable of discriminating between two different cultivation practices in *Hojiblanca* EVO oils were already found in *Picual* samples, they are visualized with a Venn diagram in **Fig. 12.3** together with a the histogram showing the VIPs values derived by the classification model (blue bars for *Hojiblanca* and orange bars for *Picual*). **Fig. S2** illustrates with box plots the relative distribution in all sample classes for (5*E*)-3-ethyl-1,5-octadiene, (5*E*)-3,7-decadiene, and (5*E*)-3-hexen-1-ol, acetate. The trend for (5*E*)-3-ethyl-1,5-octadiene and (5*E*)-3,7-decadiene is similar in both cultivars; they are more abundant in samples from conventional practices than in organic ones. However, a different behavior was observed for (5*E*)-3-hexen-1-ol acetate that was more abundant in the organic oils from *Hojiblanca* and lower in the *Picual*..

Hojiblanca EVO oil samples showed additional markers for organic cultivation, relative distribution is shown in box-plots of **Fig. S3**; they are (Z)-3-hexen-1-ol and 2-pentanol. These analytes were more abundant in organic samples. According to previous studies those cultivars with the highest percentage of C6 alcohols, mainly represented by (Z)-3-hexen-1-ol were the same cultivars with the highest content of (Z)-3-hexenyl acetate. In our case, (Z)-3-hexen-1-ol and (5*E*)-3-hexen-1-ol acetate resulted to be more abundant in organic *Hojiblanca* samples [11]. This could be due to a different expression/activity of the isomerase, an enzyme within the Lipoxygenase (LOX) pathway.

12.3.3 Accurate quantification of aroma active compounds: Artificial Intelligence smelling

Many of the informative compounds identified in the first part of the study have minimal or no sensorial impact, therefore they would not be perceived by a consumer during oil consumption. In this perspective, the possibility to highlight possible sensory-relevant compositional differences between oils obtained from olives harvested with organic practices assumes a great interest from a marketing perspective. If the farming does not have an impact on the typical yet univocal sensory trait of the cultivar, consumer's experience is likely aligned

to expectations. However, to proceed in this direction, accurate quantification of sensory active analytes is required [66], [67].

This step of the study performed accurate quantification of an extended list of target analytes to provide a solid foundation for samples' aroma blueprinting [27], [34], [35]. Moreover, by targeting key-aroma compounds elucidated by sensomics [41], the implementation of the Artificial Intelligence smelling concept was possible as the prediction of oils aroma features was reliable and objective [68].

The selection of analytes for accurate quantification was driven by their impacting sensory qualities (potent odorants and key aromas) in the definition of high-quality EVO oils aroma and/or information role already described in reference literature [43]. Alcohols and carbonyls contribute to *fresh-green* and *fruity* notes that are considered within the positive attributes mandatory for EVO oils qualification [69]. They are (*E*)-2-pentenal, (*Z*)-2-penten-1-ol, 1-penten-3-ol, (*E,E*)-2,4-hexadienal, (*E*)-2-hexenal, (*Z*)-3-hexen-1-ol, hexanal, and 1-hexanol. Compounds responsible for sensorial defects were also monitored: heptanal, (*E*)-2-octenal, and (*E*)-2-nonenal are linear saturated and unsaturated aldehydes deriving from fatty acids hydroperoxides cleavage. They have *rancid* and *fatty* notes and provide information on the autoxidation process and shelf-life evolution [41], [65]. The targeted compounds list was completed by analytes correlated to peculiar sensory defects described in the International Olive Council standards [27], [34], [70]: (*E*)-2-penten-1-ol, with *mushroom-like* and *earthy* sensations; 2-pentanol, related to *musty* and *fermented* perception; and ethyl acetate, responsible for the *winey* note.

Other compounds, deriving from the LOX pathway, are 1-penten-3-one, hexyl acetate, (*Z*)-3-hexenal, (*Z*)-3-hexenyl acetate, and (*E*)-2-hexen-1-ol, all correlated to *green* and *fruity* notes [71]. Terpenes and terpenoids, as limonene, α -pinene, β -pinene, δ -3-carene, eucalyptol, α -copaene, and (*E,E*)- α -farnesene are mainly known to be cultivar and/or geographical markers [72] bringing *herbal*, *pine*, and *citrus-like* notes to the overall perception [12], [34], [35], [73]. 6-Methyl-5-hepten-2-one is a ripening indicator with a *fruity-like* odor and propanal is present in high-quality EVOOs and described as *fresh*, *fruity*, and *malty* notes [43]. Acetic acid, propanoic acid, and butanoic acid are responsible for *vinegary*, *musty*, and *rancid* defects [27], [41]; 1-octanol is usually found in oxidized samples responsible for a *mushroom-like* odor; and benzaldehyde imparts *moldy* and *earthy* notes. Finally, octanal, nonanal, (*E*)-2-heptenal, and (*E,E*)-2,4-heptadienal contribute to *rancid* and *fatty* sensations.

The analytical strategy involving GC $\times$ GC with parallel MS and FID detection offers several advantages for accurate quantification and analyte identity confirmation. MS detection provides suitable specificity through the selection of specific ion traces in case of co-elutions, while FID offers stable and structure-specific response factors over a wide range of concentrations. This allows the use of FID-predicted RRFs, based on combustion enthalpies and molecular structure, for accurate quantification without external calibration. In contrast, 1D-GC, with its lower peak capacity, fails to achieve suitable chromatographic resolution for applying FID response factors. **Fig. S4** shows the contour plot of a *Picual* sample (EVOO_3 conventional cultivation, ripening stage III), while enlarged areas highlight elution regions

with critical pairs where 1D-GC would fail in providing full/suitable resolution for accurate quantification.

Quantified analytes showed an exponential decay in multiple consecutive extractions from the HS (MHS-SPME approach) indicating that they were not saturating the HS [*i.e.*, HS linearity range [74]]. The use of a multicomponent fiber, *i.e.*, CAR/PDMS/DVB, that mixes partition and adsorption phenomena prevents a correct application of non-exhaustive extraction approaches and consequently the possibility to adopt external calibration. Once the HS linearity for selected analytes was verified, their quantification was by predicted RRFs based on combustion enthalpies. **Equation 12.1** illustrates how to calculate RRFs [75].

Equation 12.1:

$$RRF = 10^3 * \left(\frac{MW_i}{MW_{IS}} \right) * (-61.3 + 88.8\eta_C + 18.7\eta_H - 41.3\eta_O + 6.4\eta_N + 64.0\eta_S - 20.2\eta_F - 23.5\eta_{Cl} - 10.2\eta_{Br} - 1.75\eta_I + 127\eta_{benz})^{-1}$$

where η_C , η_H , η_O , η_N , η_S , η_F , η_{Cl} , η_{Br} , η_I , and η_{benz} correspond to the number of carbon, hydrogen, oxygen, nitrogen, sulfur, fluorine, chlorine, bromine, and iodine atoms, and the number of benzene rings; and MW_i and MW_{IS} are the molecular weights of the analyte *i* and the *IS* adopted for the development of the reference model [75]. In this study, quantification was achieved by considering the analytes' response (*i.e.*, 2D peak volume), as resulting by FID signal, and its predicted RRF calculated with **Eq. 12.1**. To align combustion enthalpies *versus* hexanal, here taken as internal standard and accurately quantified by external calibration, the analyte-specific RRF was corrected to the hexanal/methyl octanoate ratio ($RRF_{i, \text{hexanal}} = RRF_{i, \text{methyl octanoate}}/0.967$) [76]. RRFs are listed in **Table 12.2** together with analytes molecular weights (g/mol), formulae, decay constant β , odor qualities, and odor thresholds (OTs) in oil.

The quantification results are reported in **Table S3**; the analytes' amount in the oil samples is expressed in ng/g (mean value over four replicates) accompanied by absolute uncertainty.

12.3.4 Aroma Blueprinting of organic extra-virgin olive oils

Based on quantification results, odor activity values (OAVs), as the ratio between the odorant amount and its odor threshold (OT) in oil, were calculated therefore obtaining a more realistic picture of the actual contribution to the sensory profile of the most potent odorants on the overall EVO oils aroma, *i.e.* the aroma blueprint [37], [41].

In this study, we referred to both known key odorants qualified through the sensomic protocol [41] and potent odorants highlighted in the reference literature. They are: (*E*)-2-pentenal, (*E*)-2-penten-1-ol, (*Z*)-2-penten-1-ol, 1-penten-3-ol, 3-methyl-1-butanol, 2-pentanol, ethyl acetate, (*E,E*)-2,4-hexadienal, (*E*)-2-hexenal, (*Z*)-3-hexen-1-ol, hexanal, 1-hexanol, heptanal, (*E*)-2-octenal, and (*E*)-2-nonenal [50], [73], [77], [78].

Spider-diagrams in **Fig. 12.5A-C**, evoking descriptive sensory analysis results, show the EVO oils aroma blueprint by plotting OAVs in $\log_{10}$ scale for selected potent odorants. In **Fig. 12.5A-B** spider diagrams show blueprint differences between cultivation practices on *Hojiblanca* and *Picual* cultivars observed separately while **Fig. 12.5C-D** show the cultivar aroma blueprint as a function of olives harvest methodology. To note, aroma blueprinting supports effective comparative analysis of samples based on their aroma code but cannot replace a human panel in expressing preferences or likeness.

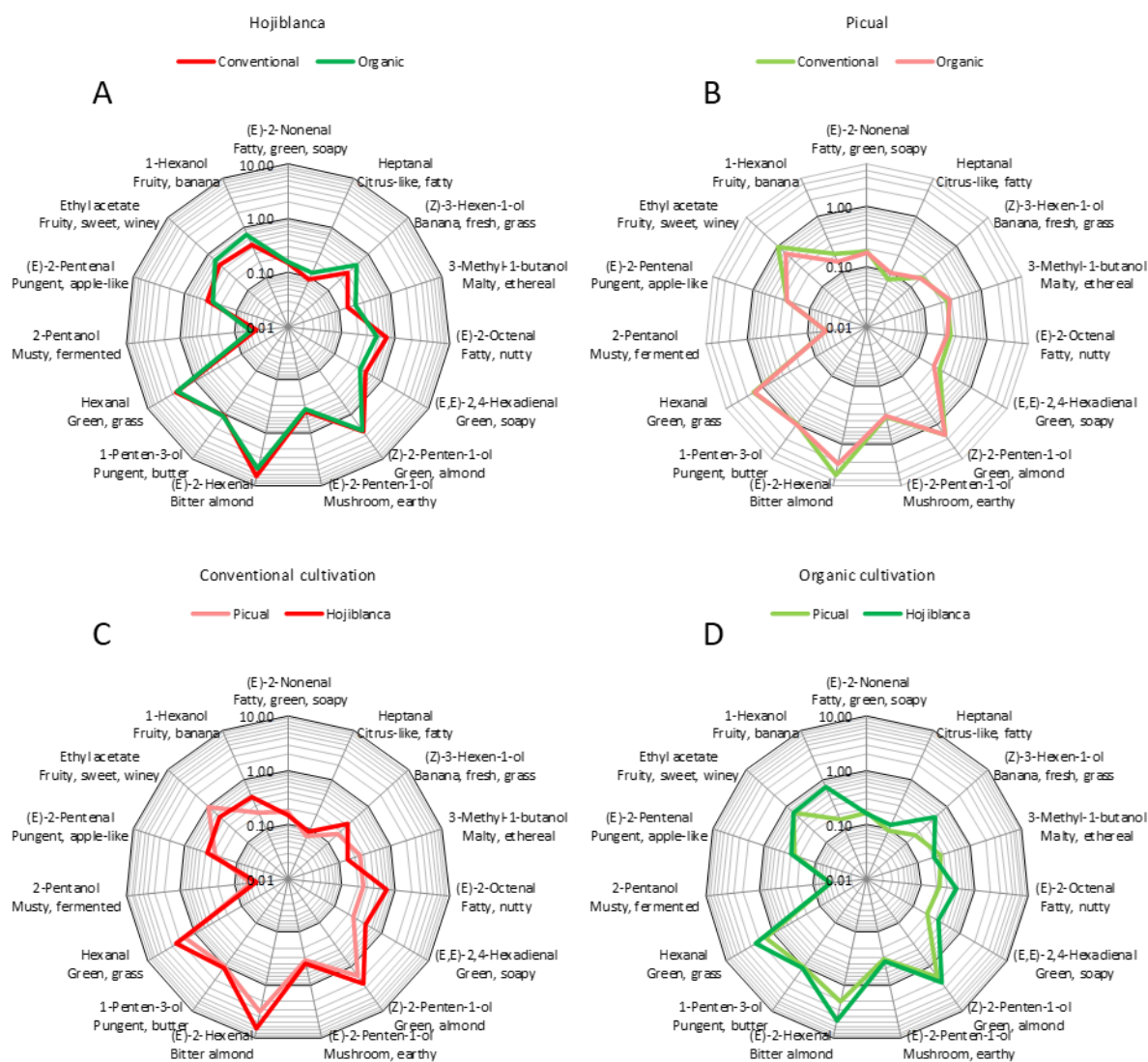


Figure 12.5: Spider diagrams illustrating OAVs in log₁₀ scale for the most potent odorants. Cultivation methods are compared for *Hojiblanca* (Fig. 5A) and *Picual* (Fig. 5B) while cultivar differences on aroma blueprint are highlighted in Fig. 5C for the conventional cultivation method and in Fig. 5D for the organic one.

Considering the impact of cultivation methods on the *Picual* cultivar, the profiles between organic and conventional harvesting are nearly identical. However, discriminating analytes identified through PLS-DA modeling (*Picual* organic vs. conventional, VIPs >1) showed lower absolute amounts of 1-hexanol (-29%), (*E*)-2-hexenal (-35%), and (*E,E*)-2,4-hexadienal (-25%) in oils from organically cultivated olives. For the *Hojiblanca* samples, the most significant differences (PLS-DA modeling, *Hojiblanca* organic vs. conventional, VIPs >1) were a 64% higher concentration of 1-hexanol in organic samples. Other notable odorants with significant quantitative differences were (*Z*)-3-hexen-1-ol (+57%) and (*E*)-2-octenal (-34%). Despite these statistically relevant variations, all were below OAV unit values, suggesting no perceivable differences in the overall aroma profile.

On the other hand, by comparing the impact of olives' harvesting practices on the cultivar aroma identity, it appeared that the organic cultivation methodology had an effect in emphasizing concentration differences of the cultivar's specific blueprint odorants. In conventional cultivation, *Hojiblanca* samples showed higher amounts of (*E*)-2-octenal, (*E*)-2-hexenal, and (*Z*)-3-hexen-1-ol (+170.54%, +107.74%, +85.58% respectively) compared to the *Picual* oils (Fig. 12.5C) likely impacting on the *fatty, fresh and green, and bitter almond* notes. In case of organic cultivation, potent odorants differences were even higher (Fig. 12.5D). The (*E*)-2-octenal was +99 %, (*E*)-2-hexenal +122 %, and (*Z*)-3-hexen-1-ol +207 % in *Hojiblanca* vs. *Picual*.

12.4 Conclusions

The volatilome of extra-virgin olive oil (EVOO) effectively encodes critical quality parameters such as cultivar, geographical origin, cultivation methodologies, harvesting stage, oxidation level, and sensory profile. Through comprehensive analysis using GC×GC, over 190 reliable peak features were monitored, with 84 compounds identified. Principal Component Analysis (PCA) revealed distinct clusters for *Picual* and *Hojiblanca* cultivars, demonstrating the significance of botanical traits in the volatile profiles. However, some overlap indicated shared influences, likely from regional pedoclimatic conditions and fruit ripening stages.

Filtering variables by Fisher ratio and further PCA provided clearer discrimination, especially for cultivar type, explaining up to 60.74% of the variance. Cultivation methodologies (organic vs. conventional) were less distinguishable, but for *Picual* oils, significant differences were observed using supervised strategies as PLS-DA, which identified key volatiles with high accuracy.

Diagnostic volatiles for cultivar discrimination, 125 UT features above a $F > 4$, included alkenes [(*E,E*)-3,7-decadiene, (*5E*)-3-ethyl-1,5-octadiene, 3,4-diethyl-1,5-hexadiene (RS+SR), 3,4-diethyl-1,5-hexadiene (meso), (*5Z*)-3-ethyl-1,5-octadiene, and (*E,Z*)-3,7-decadiene), carbonyls [1-penten-3-one, (*E*)-2-octenal, (*E*)-2-nonenal, (*Z*)-2-hexenal, 3-hydroxy-2-butanone, (*Z*)-3-hexenal, 3-methylbutanal, (*E*)-2-hexenal, 2-methyl butanal, and 6-methyl-5-hepten-2-one], and alcohols [2-pentanol, (*Z*)-2-penten-1-ol, (*Z*)-3-hexen-1-ol, 1-hexanol, 3-methyl-1-butanol, 1-penten-3-ol, and (*E*)-2-hexen-1-ol] reflecting LOX pathway activity while also intercepting maturation stages likely aligned between cultivars. Accurate quantification of aroma-active compounds confirmed that the two cultivars have distinctive blueprint traits due to the different OAVs for (*E*)-2-octenal, (*E*)-2-hexenal, and (*Z*)-3-hexen-1-ol. Differences that were minimal between the cultivation practices suggesting that consumer would probably not perceived any difference between organic and conventional products.

In conclusion, the study underscores the complex interplay of factors shaping EVOO's volatile profile. While cultivar remains the dominant influence, cultivation practices and ripening stages also play crucial roles. The findings highlight the potential for using volatilome analysis in EVOO quality control and marketing, ensuring alignment of consumer expectations with sensory experiences.

Acknowledgements

Financial support from “Ayudas a la I + D + i, en el ámbito del Plan Andaluz de Investigación, Desarrollo e Innovación (PAIDI 2020)” co-funded by SOVENA Group SA (Project number: p18-tp-2850) and for “Fondo Europeo de Desarrollo Regional (FEDER) y Consejería de Transformación Económica, Industria, Conocimiento y Universidades de la Junta de Andalucía, dentro del Programa Operativo FEDER 2014-2020” (US-1380836) is acknowledged.

Supplementary Information

Table S1: UT features mapped across all samples by feature template matching. Fisher ratio values (F) are those estimated between classes. F value cultivar (*Hojiblanca* vs. *Picual*) and F value cultivation (organic vs. conventional). In green F values above the F_{crit} .

UT feature ID	RT I	RT II	Exp. I^t	F Value Cultivar	F Value Cultivation
Hexane	3.58	0.45	797	9.42	0.37
(143)	3.97	0.54	812	0.22	3.21
Propanal	4.83	0.42	833	1.04	2.34
Octane	4.87	0.99	844	20.47	0.78
Acetone	5.05	0.45	850	16.81	6.64
(147)	6.15	0.61	887	4.18	1.71
Ethyl acetate	6.44	0.37	897	14.07	0.12
2-Methyl butanal	7.08	0.64	924	20.13	3.03
Ethanol	7.35	0.37	928	59.74	3.12
3,4-Diethyl-1,5-hexadiene (meso)	7.90	1.21	946	16.29	0.58
3-Methylbutanal	8.00	0.72	939	26.35	0.03
3,4-Diethyl-1,5-hexadiene (RS+SR)	8.45	1.40	956	18.54	21.14
Pentanal	8.76	0.75	975	0.03	1.30
(150)	9.58	0.42	1003	0.00	2.02
(154)	9.73	0.86	1007	1.45	3.90
(Z)-1-Methoxy-3-hexene	9.74	1.17	1007	9.29	2.04
(5Z)-3-Ethyl-1,5-octadiene	9.93	1.69	1012	9.72	41.62
1-Penten-3-one	10.15	0.68	1017	55.44	0.54
α -Pinene	10.24	1.68	1017	13.31	3.39
(5E)-3-Ethyl-1,5-octadiene	10.47	1.69	1026	53.42	32.46
(163)	10.62	0.37	1032	35.15	0.93
Toluene	10.73	0.91	1029	2.39	11.31
(126)	11.34	0.67	1047	22.95	1.73
(138)	11.96	0.66	1064	14.05	3.49
(E,Z)-3,7-Decadiene	12.52	1.92	1078	5.85	30.38
Hexanal	12.57	0.91	1079	6.81	1.35
(157)	12.78	0.44	1084	41.34	2.77
(E,E)-3,7-Decadiene	12.85	1.89	1085	60.48	27.47
(161)	13.20	1.36	1094	10.71	1.09
β -Pinene	13.34	1.85	1097	1.51	0.13
(Z)-2-Pentenal	13.49	0.74	1101	0.46	0.01
3-Methylbutyl acetate	14.42	0.97	1123	50.65	0.14
(E)-2-Pentenal	14.45	0.73	1123	1.81	0.27
(13)	14.87	0.59	1133	6.45	1.89
(167)	14.92	0.78	1134	31.52	1.98
2-Pentanol	14.99	0.44	1136	45.70	5.62
(165)	15.00	1.26	1136	2.33	2.29
(Z)-3-Hexenal	15.14	0.77	1139	28.45	1.19
δ -3-Carene	15.16	1.77	1139	3.72	0.20
1-Penten-3-ol	15.75	0.42	1153	17.75	5.87
(169)	15.91	1.16	1154	0.39	8.15
(168)	16.05	1.48	1159	9.68	2.79
(89)	16.38	0.71	1168	17.44	1.53
2-Heptanone	16.85	0.98	1178	1.67	48.44
Heptanal	16.99	1.02	1181	0.04	15.58
(97)	17.07	0.68	1186	15.69	1.61
(256)	17.20	1.67	1186	45.13	1.60
Limonene	17.48	1.67	1192	0.68	0.02
(53)	17.49	0.65	1191	0.22	0.33
(Z)-2-Hexenal	17.66	0.82	1197	29.65	1.14
3-Methyl-1-butanol	17.80	0.45	1199	17.93	0.72
(39)	17.80	1.53	1199	28.30	4.95
(58)	17.86	0.53	1199	7.96	0.10
Eucalyptol	17.88	1.86	1201	0.48	0.04
(87)	18.38	0.59	1212	5.54	0.00
(E)-2-Hexenal	18.40	0.79	1212	24.35	9.80

Table S1: continue

UT feature ID	RT I	RT II	Exp. I ^T	F Value Cultivar	F Value Cultivation
(111)	18.96	0.58	1224	9.49	0.05
γ-Terpinene	19.59	1.71	1239	0.74	4.11
(50)	19.65	0.53	1241	27.59	0.26
(252)	19.77	2.29	1243	145.59	8.36
trans-β-Ocimene	19.95	1.44	1246	15.29	0.26
(139)	20.01	0.91	1249	7.70	3.62
(110)	20.58	0.57	1260	40.12	0.42
o-Cymene	20.72	1.42	1264	0.91	0.01
Hexyl acetate	21.05	1.15	1271	7.05	0.33
(181)	21.06	1.52	1271	0.69	0.27
3-Hydroxy-2-butanone	21.45	0.46	1279	29.36	2.16
2-Octanone	21.50	1.07	1280	0.06	32.51
Octanal	21.70	1.10	1285	0.14	11.56
(124)	22.30	1.42	1298	45.21	0.07
(E)-3-Hexen-1-ol acetate	22.49	1.57	1303	3.39	0.74
(Z)-2-Penten-1-ol	22.50	0.42	1303	43.32	4.36
(E)-2-Penten-1-ol	22.84	0.40	1310	0.02	10.18
(Z)-3-Hexenyl acetate	23.01	1.01	1315	10.98	5.79
(E)-2-Heptenal	23.17	0.86	1318	1.65	0.44
(249)	23.47	1.96	1325	117.35	6.96
Dioxa-1,6-spiro[4.5]decane	23.54	1.44	1326	16.21	3.18
3-Methyl-cyclopentanol	23.68	0.65	1329	9.60	0.64
6-Methyl-5-hepten-2-one	23.80	0.95	1332	18.08	0.25
1-Hexanol	24.30	0.48	1343	18.12	1.36
(E)-3-Hexen-1-ol	24.78	0.47	1354	0.28	0.61
(248)	24.90	1.63	1356	19.75	0.00
(Z)-3-Hexen-1-ol	25.69	0.46	1374	25.57	8.45
Methoxymethyl-benzene	26.05	0.93	1382	261.25	0.47
(E,E)-2,4-Hexadienal	26.30	0.63	1388	15.78	1.75
Nonanal	26.35	1.13	1389	8.59	0.01
(E)-2-Hexen-1-ol	26.55	0.62	1394	17.74	0.66
(Z)-2-Hexen-1-ol	26.62	0.43	1396	1.14	2.91
(E)-2-Octenal	27.70	0.40	1420	35.05	0.77
(137)	28.47	1.09	1438	20.90	1.19
Acetic acid	28.51	0.24	1440	1.19	0.91
(187)	28.73	0.50	1444	0.02	24.48
(E,E)-2,4-Heptadienal	29.20	0.70	1456	15.53	0.50
(186)	29.49	0.52	1463	0.02	14.07
(68)	29.76	1.22	1469	13.72	1.67
(96)	30.15	2.37	1478	91.17	0.24
(125)	30.26	0.56	1480	4.65	0.57
α-Copaene	30.55	2.37	1489	97.87	0.14
(195)	30.83	1.17	1494	6.17	0.21
(134)	31.39	0.32	1508	12.16	0.08
Benzaldehyde	31.59	0.66	1513	20.99	0.11
(246)	32.05	1.58	1524	2.85	0.23
Popanoic acid	32.15	0.26	1526	3.20	8.31
(E)-2-Nonenal	32.26	0.91	1530	32.02	3.15
(191)	32.35	0.43	1531	11.59	0.10
Dimethyl Sulfoxide	32.60	0.49	1537	7.10	1.50
(29)	32.80	0.43	1541	10.39	2.91
1-Octanol	33.01	0.55	1548	1.60	8.54
(88)	33.40	0.26	1557	0.82	31.57
(198)	34.15	0.50	1576	26.88	0.00
5-ethyl-2(5H)-Furanone	34.60	0.47	1587	48.21	0.06
Methyl benzoate	35.66	0.75	1613	132.58	0.02
Butanoic acid	35.75	0.25	1616	6.31	25.45
(117)	35.76	0.50	1616	2.92	4.44
(201)	36.31	0.68	1631	2.01	2.16
(127)	36.60	0.89	1637	15.08	6.85
(203)	36.71	0.21	1643	16.77	1.05
(202)	36.80	0.33	1643	3.42	0.22
(81)	37.42	0.25	1659	0.00	24.58

Table S1: continue

UT feature ID	RT I	RT II	Exp. I ^T	F Value Cultivar	F Value Cultivation
(244)	38.55	1.53	1687	32.45	0.07
(265)	38.92	0.78	1698	8.02	0.99
(86)	39.66	1.84	1719	74.15	0.01
1,4-Cyclohex-2-enedion	39.93	0.53	1725	54.80	4.88
Pentanoic acid	39.93	0.25	1725	1.47	13.30
2,3-Dimethylbenzaldehyde	39.99	0.66	1725	1.63	0.05
(208)	40.48	0.41	1740	6.69	9.39
(E,E)- α -Farnesene	40.60	1.56	1742	22.78	0.02
5-ethyl-2(5H)-Furanone	40.73	0.51	1746	31.10	1.23
(210)	40.87	0.31	1749	0.80	0.88
(264)	41.52	1.49	1768	47.94	8.28
Methyl salicylate	41.53	0.72	1768	1.95	1.87
2-(2-Butoxyethoxy)-ethanol	42.09	0.55	1783	0.09	0.98
(211)	43.37	0.63	1819	0.35	0.68
Hexanoic acid	43.84	0.26	1831	3.88	0.60
(100)	44.35	1.50	1845	12.27	0.96
(85)	44.65	0.96	1854	18.21	2.54
Benzyl alcohol	44.78	0.38	1858	4.28	0.00
(120)	45.60	0.32	1881	0.14	1.11
(263)	45.74	1.36	1885	29.06	0.03
Phenylethyl alcohol	46.02	0.44	1893	2.77	2.21
(245)	46.85	1.31	1917	32.01	0.50
(215)	47.56	0.26	1937	0.22	1.60
(108)	47.59	0.26	1937	0.66	1.61
(E)-2-Hexenoic acid	47.89	0.30	1947	38.33	1.13
(218)	48.14	0.66	1955	16.11	0.92
Phenol	49.10	0.26	1983	5.23	4.19
(212)	49.35	0.74	1990	20.58	0.37
(93)	49.50	1.42	1994	4.01	2.02
(220)	50.64	0.84	2029	5.30	6.84
(266)	50.84	1.52	2035	21.55	0.34
(65)	51.10	0.27	2043	4.18	0.05
(235)	51.12	0.70	2044	29.27	1.32
Methyl 2-methoxybenzoate	51.31	0.66	2049	21.04	0.09
(267)	51.82	0.45	2066	0.30	2.53
(268)	53.99	0.25	2137	2.28	0.59
(234)	54.30	1.35	2142	7.03	1.30
(222)	54.48	0.28	2148	1.71	0.60
(223)	54.79	0.72	2158	8.63	1.10
(228)	55.35	0.26	2175	0.00	0.71
(227)	56.43	0.51	2210	7.06	3.09
(123)	57.69	0.95	2252	0.91	0.44
(224)	57.72	0.29	2254	2.10	0.31
(71)	58.31	1.23	2272	0.61	0.82
(232)	58.80	1.28	2289	6.39	1.84
(225)	59.34	0.54	2307	13.80	0.13
(131)	59.95	0.50	2330	8.74	0.94
(229)	60.95	0.76	2364	2.49	0.08
(230)	62.18	0.94	2406	3.79	0.41
(231)	62.59	0.47	2422	30.67	0.00
(236)	63.05	1.23	2436	0.06	2.14
(45)	63.19	0.86	2442	0.05	0.07
(226)	63.59	0.44	2456	0.97	24.44
(113)	65.49	0.13	2524	34.06	2.28
(237)	65.53	0.88	2525	15.41	0.40
(241)	65.54	0.37	2525	7.44	2.19
(238)	65.88	0.66	2538	5.86	0.03
(109)	66.09	0.14	2543	31.07	2.51
(105)	66.51	0.14	2561	24.63	2.71
(102)	66.96	0.13	2577	36.02	2.25
(239)	67.00	1.11	2577	12.00	1.32
(91)	67.63	0.13	2600	37.33	2.38

Table S1: continue

UT feature ID	RT I	RT II	Exp. I^T	F Value Cultivar	F Value Cultivation
(84)	68.02	0.13	2615	38.72	2.01
(76)	68.43	0.13	2627	41.26	1.76
(74)	69.09	0.13	2652	36.77	2.03
(73)	69.35	0.13	2663	39.46	1.88
(242)	69.55	0.33	2668	0.42	0.93
Dibutylphthalate (IS dilution solvent)	69.68	0.97	2674	7.48	1.92
(72)	69.74	0.13	2677	34.05	2.17

Table S2. Variable Importance for the Projections (VIPs) values and corresponding standard deviation (SD) according to the PLS-DA classification model for *Picual* and *Hojiblanca* EVO oils conventional vs. organic production.

Variable	<i>Picual</i>		<i>Hojiblanca</i>	
	VIP	SD	VIP	SD
Methoxymethyl-benzene	1.45	0.059	1.27	0.097
(E)-3-Hexen-1-ol acetate	1.40	0.063	1.44	0.218
2-Heptanone	1.35	0.056	1.01	0.154
(5E)-3-Ethyl-1,5-octadiene	1.33	0.044	1.61	0.250
2-Octanone	1.32	0.061	1.02	0.122
(E,E)-3,7-Decadiene	1.28	0.054	-	-
α -Copaene	1.29	0.125	-	-
3,4-Diethyl-1,5-hexadiene (RS+SR)	1.25	0.073	-	-
3-Hexenyl acetate	1.23	0.152	-	-
(5Z)-3-Ethyl-1,5-octadiene	1.22	0.069	-	-
(E,Z)-3,7-Decadiene	1.12	0.096	-	-
Butanoic acid	1.11	0.146	-	-
(E)-2-Penten-1-ol	1.10	0.146	-	-
Heptanal	1.05	0.150	-	-
(Z)-2-Penten-1-ol	1.01	0.171	-	-
(E)-2-Pentenal	1.01	0.157	-	-
(E,E)-3,7-Decadiene	-	-	1.55	0.271
(Z)-3-Hexen-1-ol	-	-	1.40	0.192
1-Hexanol	-	-	1.10	0.233
2-Pentanol	-	-	1.08	0.147
(Z)-1-Methoxy-3-hexene	-	-	1.06	0.291
(E)-3-Hexen-1-ol	-	-	1.06	0.201
(Z)-2-Hexenal	-	-	1.05	0.161

Table S3. Quantification results obtained by MHS-SPME-GC×GC-FID/MS and predicted FID relative response factors in all analyzed samples. Results are expressed in ng/g and accompanied by absolute uncertainty.

accessible at the link: <https://ars.els-cdn.com/content/image/1-s2.0-So889157524010093-mmc2.xlsx>

Fig. S2. Box plots of discriminant compounds (from PLS-DA modeling) between organic and conventional cultivation on both *Picual* and *Hojiblanca* samples. Green boxes are for organic samples (mean/median) while red and pink boxes are for conventional samples (mean/median) data.

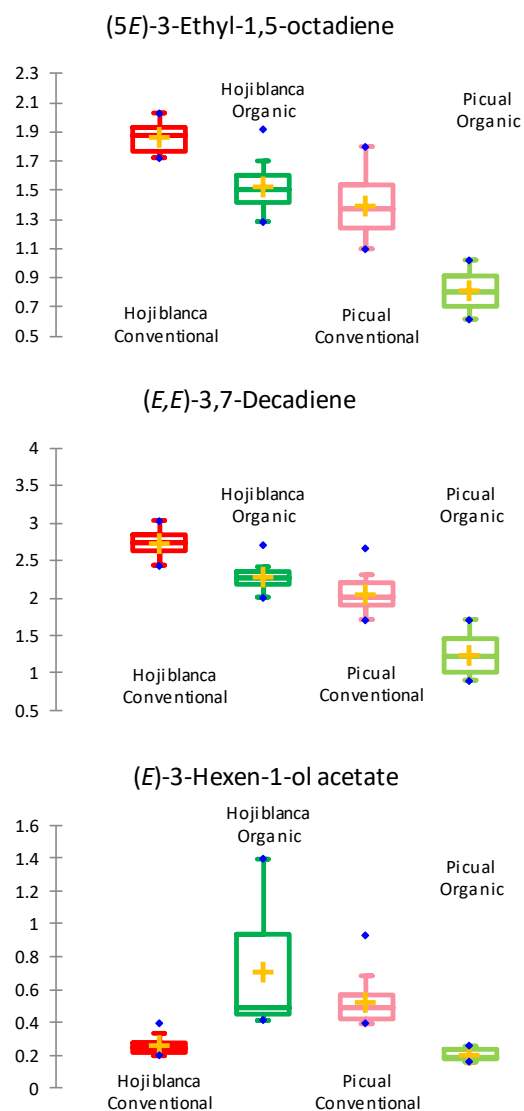


Fig. S3. Box plots of discriminant compounds (from PLS-DA modeling) between organic and conventional cultivation on *Hojiblanca* samples. Green boxes represent organic (mean/median) while red boxes are for conventional samples (mean/median).

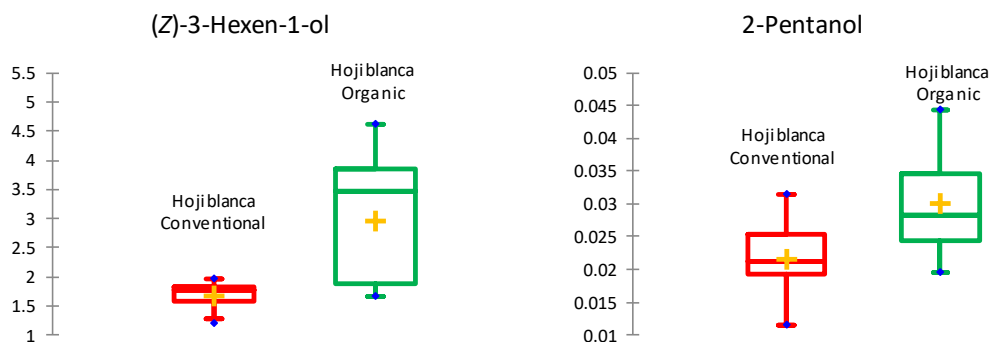
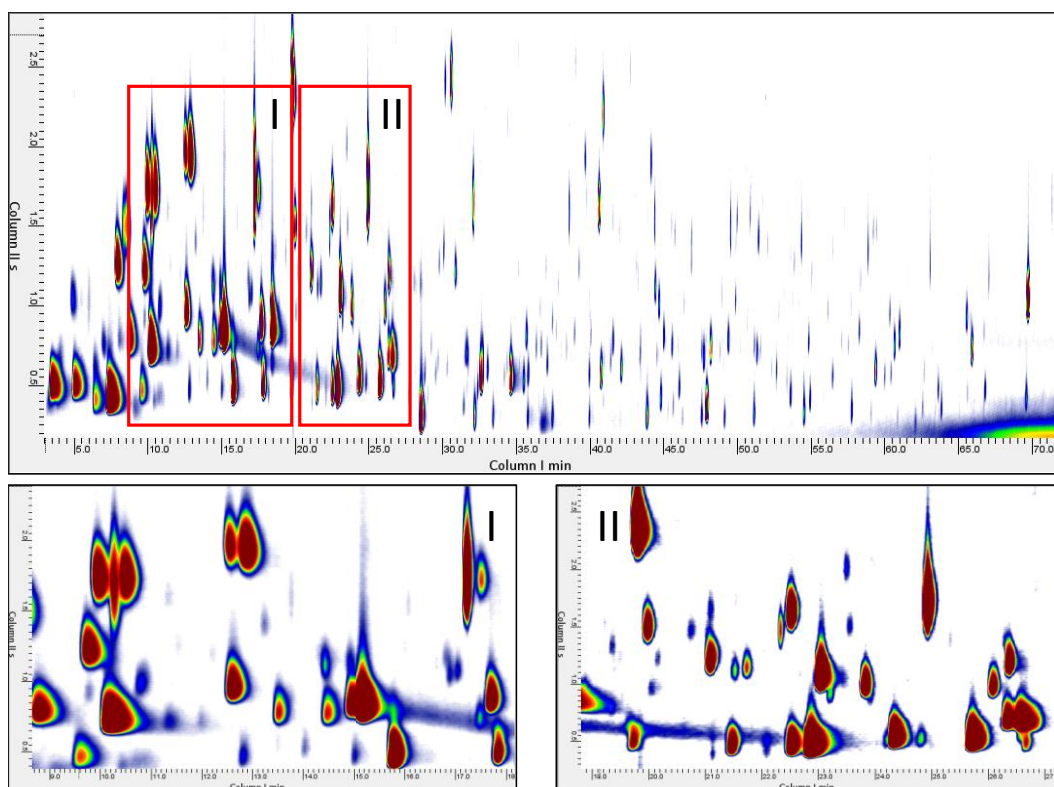


Fig. S4: Contour plot of a *Picual* sample (EVOO_3 conventional cultivation, ripening stage III) detectable volatilome. Enlarged areas highlight elution regions with critical pairs where 1D-GC/FID would fail in providing full/suitable resolution for accurate quantification.



References

- [1] A. Lioupi *et al.*, “Optimisation of the HS-SPME/GC-MS Approach by Design of Experiments Combined with Chemometrics for the Classification of Cretan Virgin Olive Oils,” *Metabolites*, 2022.
- [2] MAPAMA, “General Subdirectorate for Differentiated Quality and Ecological Agriculture, Ministry of Agriculture and Fisheries, Food and Environment. 2019.” [Online]. Available: <https://www.mapa.gob.es/es/estadistica/temas/publicaciones/anuario-de-estadistica/2019/default.aspx>.
- [3] “European Commission Regulation (EC) Reg. 2568/91,” *Off. J. Eur. Communities*, no. 172, pp. 1–83, 1991.
- [4] L. Cecchi, M. Migliorini, and N. Mulinacci, “Virgin Olive Oil Volatile Compounds: Composition, Sensory Characteristics, Analytical Approaches, Quality Control, and Authentication,” *J. Agric. Food Chem.*, vol. 69, no. 7, pp. 2013–2040, Feb. 2021.
- [5] N. Caporaso, M. Savarese, A. Paduano, G. Guidone, E. De Marco, and R. Sacchi, “Nutritional quality assessment of extra virgin olive oil from the Italian retail market: Do natural antioxidants satisfy EFSA health claims?,” *J. Food Compos. Anal.*, 2015.
- [6] N. Arroyo-Manzanares, F. Gabriel, A. Carpio, and L. Arce, “Use of whole electrophoretic profile and chemometric tools for the differentiation of three olive oil qualities,” *Talanta*, vol. 197, pp. 175–180, May 2019.
- [7] L. Conte *et al.*, “Olive oil quality and authenticity: A review of current EU legislation, standards, relevant methods of analyses, their drawbacks and recommendations for the future,” *Trends Food Sci. Technol.*, vol. 105, pp. 483–493, Nov. 2020.
- [8] M. P. Segura-Borrego *et al.*, “Influence of the ripening chamber’s geographical location on dry-cured Iberian ham’s key odorants,” *Food Res. Int.*, 2022.
- [9] M. P. Segura-Borrego, R. Ríos-Reina, A. J. Puentes-Campos, B. Jiménez-Herrera, and R. M. Callejón, “Influence of the Washing Process and the Time of Fruit Harvesting throughout the Day on Quality and Chemosensory Profile of Organic Extra Virgin Olive Oils,” *Foods*, vol. 11, no. 19, p. 3004, Sep. 2022.
- [10] E. Hong *et al.*, “Modern analytical methods for the detection of food fraud and adulteration by food category,” *Journal of the Science of Food and Agriculture*. 2017.
- [11] L. Cecchi *et al.*, “Is the volatile compounds profile a suitable tool for authentication of virgin olive oils (*Olea europaea* L.) according to cultivars? A study by using HS-SPME-GC-MS and chemometrics,” *Food Control*, vol. 139, p. 109092, Sep. 2022.
- [12] B. Quintanilla-Casas *et al.*, “Virgin olive oil volatile fingerprint and chemometrics: Towards an instrumental screening tool to grade the sensory quality,” *LWT*, vol. 121, p. 108936, 2020.
- [13] G. Beltrán, M. P. Aguilera, C. Del Rio, S. Sanchez, and L. Martinez, “Influence of fruit ripening process on the natural antioxidant content of Hojiblanca virgin olive oils,” *Food Chem.*, vol. 89, no. 2, pp. 207–215, 2005.

- [14] M. Benito, J. M. Lasa, P. Gracia, R. Oria, M. Abenoza, and A. C. Sánchez-Gimeno, "Evolution of phenols and pigments in extra virgin olive oil from irrigated super-intensive orchard," *Eur. J. Lipid Sci. Technol.*, vol. 114, no. 5, pp. 558–567, May 2012.
- [15] B. Jiménez, A. Rivas, M. L. Lorenzo, and A. Sánchez-Ortiz, "Chemosensory characterization of virgin olive oils obtained from organic and conventional practices during fruit ripening," *Flavour Fragr. J.*, 2017.
- [16] F. Mansouri *et al.*, "Phytosterols composition of virgin olive oils from cultivars introduced in eastern Morocco in comparison to Picholine Marocaine," *J. Mater. Environ. Sci.*, 2015.
- [17] J. R. Morelló, M. J. Motilva, M. J. Tovar, and M. P. Romero, "Changes in commercial virgin olive oil (cv Arbequina) during storage, with special emphasis on the phenolic fraction," *Food Chem.*, vol. 85, no. 3, pp. 357–364, 2004.
- [18] M. P. Romero, M. J. Tovar, T. Ramo, and M. J. Motilva, "Effect of crop season on the composition of virgin olive oil with protected designation of origin 'Les Garrigues,'" *JAOCS, J. Am. Oil Chem. Soc.*, vol. 80, no. 5, pp. 423–430, 2003.
- [19] A. M. Vidal, S. Alcalá, A. de Torres, M. Moya, and F. Espínola, "Characterization of Olive Oils from Superintensive Crops with Different Ripening Degree, Irrigation Management, and Cultivar: (Arbequina, Koroneiki, and Arbosana)," *Eur. J. Lipid Sci. Technol.*, 2019.
- [20] B. Jimenez, A. Sánchez-Ortiz, M. L. Lorenzo, and A. Rivas, "Effect of organic cultivation of Picual and Hojiblanca olive varieties on the quality of virgin olive oil at four ripening stages," *Eur. J. Lipid Sci. Technol.*, 2014.
- [21] B. Jimenez, A. Sánchez-Ortiz, M. L. Lorenzo, and A. Rivas, "Effect of Agronomical Practices on the Nutritional Quality of Virgin Olive Oil at Different Ripening Stages," *JAOCS, J. Am. Oil Chem. Soc.*, 2015.
- [22] B. J. Herrera *et al.*, "Influencia del proceso de maduración del fruto en la calidad sensorial de aceites de oliva virgen de las variedades Picual, Hojiblanca y Picudo," *Grasas y Aceites*, 2012.
- [23] R. Jiménez-Herrera, B. Pacheco-López, and J. Peragón, "Water Stress, Irrigation and Concentrations of Pentacyclic Triterpenes and Phenols in *Olea europaea* L. cv. Picual Olive Trees," *Antioxidants*, vol. 8, no. 8, p. 294, Aug. 2019.
- [24] C. Cordero, J. Kiefl, P. Schieberle, S. E. Reichenbach, and C. Bicchi, "Comprehensive two-dimensional gas chromatography and food sensory properties: potential and challenges," *Anal. Bioanal. Chem.*, vol. 407, no. 1, pp. 169–191, Jan. 2015.
- [25] P. J. Marriott and Y. Nolvachai, "Separation Multidimensionality for Improved Sample Characterization—Is It Worth the Effort?," *J. Agric. Food Chem.*, vol. 69, no. 39, pp. 11535–11537, Oct. 2021.
- [26] D. Ryan and P. Marriott, "Comprehensive two-dimensional gas chromatography," *Anal. Bioanal. Chem.*, vol. 376, no. 3, pp. 295–297, 2003.
- [27] G. Purcaro, C. Cordero, E. Liberto, C. Bicchi, and L. S. Conte, "Toward a definition of blueprint of virgin olive oil by comprehensive two-dimensional gas chromatography,"

J. Chromatogr. A, vol. 1334, pp. 101–111, 2014.

- [28] I. Lukić, S. Carlin, I. Horvat, and U. Vrhovsek, “Combined targeted and untargeted profiling of volatile aroma compounds with comprehensive two-dimensional gas chromatography for differentiation of virgin olive oils according to variety and geographical origin,” *Food Chem.*, vol. 270, pp. 403–414, Jan. 2019.
- [29] F. Magagna *et al.*, “Combined untargeted and targeted fingerprinting with comprehensive two-dimensional chromatography for volatiles and ripening indicators in olive oil,” *Anal. Chim. Acta*, vol. 936, pp. 245–258, Sep. 2016.
- [30] Y. Y. Broza, P. Mochalski, V. Ruzsanyi, A. Amann, and H. Haick, “Hybrid volatilomics and disease detection,” *Angew. Chem. Int. Ed. Engl.*, vol. 54, no. 38, pp. 11036–11048, Sep. 2015.
- [31] M. Phillips *et al.*, “Detection of an extended human volatome with comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry,” *PLoS One*, vol. 8, no. 9, p. e75274, 2013.
- [32] F. Stilo, C. Bicchi, S. E. Reichenbach, and C. Cordero, “Comprehensive two-dimensional gas chromatography as a boosting technology in food-omic investigations,” *J. Sep. Sci.*, vol. 44, no. 8, pp. 1592–1611, Apr. 2021.
- [33] A. Amann *et al.*, “The human volatilome: volatile organic compounds (VOCs) in exhaled breath, skin emanations, urine, feces and saliva,” *J. Breath Res.*, vol. 8, no. 3, p. 34001, Sep. 2014.
- [34] F. STILO *et al.*, “Delineating the extra-virgin olive oil aroma blueprint by multiple headspace solid phase microextraction and differential-flow modulated comprehensive two-dimensional gas chromatography,” *J. Chromatogr. A*, vol. 1650, p. 462232, Aug. 2021.
- [35] F. Stilo *et al.*, “Delineating unique and discriminant chemical traits in Brazilian and Italian extra-virgin olive oils by quantitative 2D-fingerprinting and pattern recognition algorithms,” *J. Food Compos. Anal.*, vol. 115, no. September 2022, p. 104899, Jan. 2023.
- [36] S. E. Reichenbach, C. A. Zini, K. P. Nicolli, J. E. Welke, C. Cordero, and Q. Tao, “Benchmarking machine learning methods for comprehensive chemical fingerprinting and pattern recognition,” *J. Chromatogr. A*, vol. 1595, pp. 158–167, Jun. 2019.
- [37] M. Granvogl and P. Schieberle, “The sensomics approach: A useful tool to unravel the genuine aroma blueprint of foods and aroma changes during food processing,” in *Comprehensive Analytical Chemistry*, 1st ed., vol. 96, Elsevier B.V., 2022, pp. 41–68.
- [38] “Regulation (EU) 2018/848 of the European Parliament and of the Council of 30 May 2018 on organic production and labelling of organic products and repealing Council Regulation (EC) No 834/2007,” *Off. J. Eur. Union*, vol. 150, pp. 1–134, 2018.
- [39] M. Uceda and L. Frias, “Seasons of harvest. Changes on fruit oil content, oil composition and oil quality,” in *Proceedings of the II Seminario Oleicola International*, 1975, pp. 25–46.
- [40] R. Ríos-Reina, F. Camacho, M. L. Morales, B. Jiménez-Herrera, and R. M. Callejón, “Influence of Irrigation Modalities (Irrigation Management and Dryland), Fruit Ripening, and Cultivation Modality (Organic and Conventional) on Quality and

Chemosensory Profile of Hojiblanca and Picual Extra Virgin Olive Oils,” *Eur. J. Lipid Sci. Technol.*, 2021.

- [41] A. Neugebauer, M. Granvogl, and P. Schieberle, “Characterization of the Key Odorants in High-Quality Extra Virgin Olive Oils and Certified Off-Flavor Oils to Elucidate Aroma Compounds Causing a Rancid Off-Flavor,” *J. Agric. Food Chem.*, vol. 68, no. 21, pp. 5927–5937, Apr. 2020.
- [42] L. J. Van Gemert, “Odour thresholds,” *Compil. odour Threshold values air, water other media*, 2003.
- [43] G. Luna, M. T. Morales, and R. Aparicio, “Characterisation of 39 varietal virgin olive oils by their volatile compositions,” *Food Chem.*, 2006.
- [44] S. Squara *et al.*, “Artificial Intelligence decision-making tools based on comprehensive two-dimensional gas chromatography data: the challenge of quantitative volatilomics in food quality assessment,” *J. Chromatogr. A*, vol. 1700, p. 464041, Jul. 2023.
- [45] F. Stilo, E. Liberto, S. E. Reichenbach, Q. Tao, C. Bicchi, and C. Cordero, “Untargeted and Targeted Fingerprinting of Extra Virgin Olive Oil Volatiles by Comprehensive Two-Dimensional Gas Chromatography with Mass Spectrometry: Challenges in Long-Term Studies,” *J. Agric. Food Chem.*, vol. 67, no. 18, pp. 5289–5302, May 2019.
- [46] D. Bressanello *et al.*, “Combined untargeted and targeted fingerprinting by comprehensive two-dimensional gas chromatography: revealing fructose-induced changes in mice urinary metabolic signatures,” *Anal. Bioanal. Chem.*, vol. 410, no. 11, pp. 2723–2737, 2018.
- [47] S. Squara *et al.*, “Extending the breadth of saliva metabolome fingerprinting by smart template strategies and effective pattern realignment on comprehensive two-dimensional gas chromatographic data,” *Anal. Bioanal. Chem.*, no. 0123456789, Jan. 2023.
- [48] N. I. of S. and Technology, *NIST/EPA/NIH Mass Spectral Library with Search Program: (Data Version: NIST 08, Software Version 2.0f)*, vol. 2010, no. 1/22. Gaithersburg MD: National Institute of Standards and Technology (NIST), 2005.
- [49] S. E. Reichenbach, X. Tian, A. A. Boateng, C. A. Mullen, C. Cordero, and Q. Tao, “Reliable peak selection for multisample analysis with comprehensive two-dimensional chromatography,” *Anal. Chem.*, vol. 85, no. 10, pp. 4974–4981, 2013.
- [50] M. T. Morales, G. Luna, and R. Aparicio, “Comparative study of virgin olive oil sensory defects,” *Food Chem.*, vol. 91, no. 2, pp. 293–301, 2005.
- [51] D. Melucci *et al.*, “Rapid direct analysis to discriminate geographic origin of extra virgin olive oils by flash gas chromatography electronic nose and chemometrics,” *Food Chem.*, vol. 204, pp. 263–273, 2016.
- [52] F. Stilo, E. Liberto, S. E. Reichenbach, Q. Tao, C. Bicchi, and C. Cordero, “Exploring the Extra-Virgin Olive Oil Volatilome by Adding Extra Dimensions to Comprehensive Two-Dimensional Gas Chromatography and Time-of-Flight Mass Spectrometry Featuring Tandem Ionization: Validation of Ripening Markers in Headspace Linearity Condition,” *J. AOAC Int.*, vol. 104, no. 2, pp. 274–287, 2021.

- [53] F. Angerosa, L. Camera, N. D'Alessandro, and G. Mellerio, "Characterization of Seven New Hydrocarbon Compounds Present in the Aroma of Virgin Olive Oils," *J. Agric. Food Chem.*, vol. 46, no. 2, pp. 648–653, 1998.
- [54] S. Schöneich, G. S. Ochoa, C. M. Monzón, and R. E. Synovec, "Minimum variance optimized Fisher ratio analysis of comprehensive two-dimensional gas chromatography / mass spectrometry data: Study of the pacu fish metabolome," *J. Chromatogr. A*, vol. 1667, 2022.
- [55] L. C. Lee, C. Y. Liong, and A. A. Jemain, "Partial least squares-discriminant analysis (PLS-DA) for classification of high-dimensional (HD) data: A review of contemporary practice strategies and knowledge gaps," *Analyst*, vol. 143, no. 15. Royal Society of Chemistry, pp. 3526–3539, 2018.
- [56] M. D. R. G. da Silva, A. M. Costa Freitas, M. J. B. Cabrita, and R. Garci, "Olive Oil Composition: Volatile Compounds," in *Olive Oil - Constituents, Quality, Health Properties and Bioconversions*, InTech, 2012.
- [57] I. S. Kosma, M. G. Kontominas, and A. V. Badeka, "The Application of Chemometrics to Volatile Compound Analysis for the Recognition of Specific Markers for Cultivar Differentiation of Greek Virgin Olive Oil Samples," *Foods*, vol. 9, no. 11, p. 1672, Nov. 2020.
- [58] N. Jurado-Campos *et al.*, "Exploration of the potential of different analytical techniques to authenticate organic vs. conventional olives and olive oils from two varieties using untargated fingerprinting approaches," *Food Control*, 2021.
- [59] A. Guinda, A. Lanzón, and T. Albi, "Differences in hydrocarbons of virgin olive oils obtained from several olive varieties," *J. Agric. Food Chem.*, 1996.
- [60] R. Bortolomeazzi, P. Berno, L. Pizzale, and L. S. Conte, "Sesquiterpene, Alkene, and Alkane Hydrocarbons in Virgin Olive Oils of Different Varieties and Geographical Origins," *J. Agric. Food Chem.*, vol. 49, no. 7, pp. 3278–3283, Jul. 2001.
- [61] P. Zunin, R. Boggia, P. Salvadeo, and F. Evangelisti, "Geographical traceability of West Liguria extravirgin olive oils by the analysis of volatile terpenoid hydrocarbons," *J. Chromatogr. A*, 2005.
- [62] I. Lukić *et al.*, "Towards understanding the varietal typicity of virgin olive oil by correlating sensory and compositional analysis data: a case study," *Food Res. Int.*, 2018.
- [63] K. B. Bubola *et al.*, "Morphological and molecular characterization of Bova olive cultivar and aroma fingerprint of its oil," *Food Technol. Biotechnol.*, no. 52, pp. 342–350, 2014.
- [64] F. Angerosa, "Influence of volatile compounds on virgin olive oil quality evaluated by analytical approaches and sensor panels," *European Journal of Lipid Science and Technology*, vol. 104, no. 9–10. pp. 639–660, 2002.
- [65] L. Cecchi, M. Migliorini, E. Giambanelli, A. Rossetti, A. Cane, and N. Mulinacci, "New Volatile Molecular Markers of Rancidity in Virgin Olive Oils under Nonaccelerated Oxidative Storage Conditions," *J. Agric. Food Chem.*, vol. 67, no. 47, pp. 13150–13163, Nov. 2019.
- [66] H. H. Jeleń and M. N. Wieczorek, "Commentary: 'Quantitative' vs quantitative

Headspace Solid-Phase Microextraction (HS-SPME) in food volatile and flavor compounds analysis,” *J. Food Compos. Anal.*, vol. 115, no. September 2022, p. 104955, Jan. 2023.

- [67] B. Sgorbini, C. Cagliero, E. Liberto, P. Rubiolo, C. Bicchi, and C. Cordero, “Strategies for Accurate Quantitation of Volatiles from Foods and Plant-Origin Materials: A Challenging Task,” *J. Agric. Food Chem.*, vol. 67, no. 6, pp. 1619–1630, Feb. 2019.
- [68] L. Nicolotti, V. Mall, and P. Schieberle, “Characterization of Key Aroma Compounds in a Commercial Rum and an Australian Red Wine by Means of a New Sensomics-Based Expert System (SEBES) - An Approach to Use Artificial Intelligence in Determining Food Odor Codes,” *J. Agric. Food Chem.*, vol. 67, no. 14, pp. 4011–4022, 2019.
- [69] International Olive Oil Council, “International trade standard applying to olive oils and olive-pomace oils international trade standard applying to olive oils and olive-pomace oils,” *CoI/T.15/Nc*, no. NO 3/REV.12, p. 17, 2009.
- [70] International Olive Oil Council, “Sensory analysis of olive oil - Method for the organoleptic assessment of virgin olive oil,” *COI/T.20/DOC.15*, no. Rev.10, pp. 1–20, 2018.
- [71] P. Reboredo-Rodríguez, C. González-Barreiro, B. Cancho-Grande, and J. Simal-Gándara, “Concentrations of aroma compounds and odor activity values of odorant series in different olive cultivars and their oils,” *J. Agric. Food Chem.*, vol. 61, no. 22, pp. 5252–5259, Jun. 2013.
- [72] B. Quintanilla-Casas *et al.*, “Profiling versus fingerprinting analysis of sesquiterpene hydrocarbons for the geographical authentication of extra virgin olive oils,” *Food Chem.*, vol. 307, no. March 2019, 2020.
- [73] B. Quintanilla-Casas *et al.*, “Supporting the sensory panel to grade virgin olive oils: an in-house-validated screening tool by volatile fingerprinting and chemometrics,” *Foods*, vol. 9, no. 10, pp. 1–14, 2020.
- [74] B. Kolb and L. S. Ettre, *Static headspace-gas chromatography : theory and practice*. New York: Wiley-VCH, 2006.
- [75] J. Y. De Saint Laumer *et al.*, “Prediction of response factors for gas chromatography with flame ionization detection: Algorithm improvement, extension to silylated compounds, and application to the quantification of metabolites,” *J. Sep. Sci.*, vol. 38, no. 18, pp. 3209–3217, Jul. 2015.
- [76] T. Cachet *et al.*, “IOFI recommended practice for the use of predicted relative-response factors for the rapid quantification of volatile flavouring compounds by GC-FID,” *Flavour Fragr. J.*, 2016.
- [77] C. M. Kalua, M. S. Allen, D. R. Bedgood, A. G. Bishop, P. D. Prenzler, and K. Robards, “Olive oil volatile compounds, flavour development and quality: A critical review,” *Food Chem.*, vol. 100, no. 1, pp. 273–286, 2007.
- [78] R. Aparicio, M. T. Morales, R. Aparicio-Ruiz, N. Tena, and D. L. García-González, “Authenticity of olive oil: Mapping and comparing official methods and promising alternatives,” *Food Res. Int.*, vol. 54, no. 2, pp. 2025–2038, 2013.

Conclusive Remarks

The aim of this PhD Project was the development of advanced and innovative analytical strategies, with a particular focus on the transition to greener methodologies. The project specifically addresses the application of these strategies to the food chemistry field, aiming to improve the sustainability, efficiency, and accuracy of analytical techniques in food quality control. By exploring novel approaches in the field, the thesis contributes to the advancement of environmentally friendly practices, ensuring both high performance and minimal environmental impact. To achieve these objectives, the presented research in this thesis was structured around three distinct research lines, each addressing specific challenges while sharing the common goal of promoting the green transition in analytical chemistry. The first research line focused on the development of greener methodologies for assessing hazelnut quality, emphasizing sustainability without compromising analytical accuracy. The second line explored the quantification of health-hazardous volatile organic compounds (VOCs) in water samples, integrating greener approaches to enhance environmental safety. Lastly, the third research line aimed to establish a more general, yet efficient, green workflow for the qualification, traceability, and authenticity of various food products, ensuring reliability and sustainability across diverse application.

Referring to the first research line, the paper “*Validation of a High-Throughput Method for the Accurate Quantification of Secondary Products of Lipid Oxidation in High-Quality Hazelnuts (Corylus avellana L.): A Robust Tool for Quality Assessment*” validated a high-throughput procedure combining MHS-SPME with conventional and fast-GC-MS for hazelnut quality and safety assessment. The fast-GC setup offered higher throughput with acceptable performance, supporting industrial needs and aligning with greener analytical practices by reducing analysis time. By quantifying four aldehydes, this method surpasses traditional approaches in accuracy and information capacity. It provides an objective rancidity benchmark, aiding decision-making for hazelnut screening and storage management while supporting ISO 17025 accreditation.

Thanks to the development of this method, the next step involved approaching the design of a low-cost, compact, and portable system, carried out in collaboration with the EVOQUE project. Preliminary data from this three-year project, which is still in its first year and presented in this thesis, provide promising insights into the successful realization of these innovative portable systems. Such advancements are fully aligned with the goals of GAC, demonstrating a commitment to sustainability and efficiency in analytical methodologies.

According to the second research line, the potential of renewable gases, such as nitrogen and hydrogen, as sustainable alternatives to helium in analytical workflows was

evaluated. Unlike helium, which is non-renewable and involves high extraction costs, nitrogen and hydrogen can be generated in laboratories using specialized generators, offering a low-cost and renewable solution. This shift significantly reduces operational expenses while aligning with the principles of GAC. For instance, the use of nitrogen as a purge gas in P&T-GC-MS workflows not only meets the performance standards required by EPA Method 8260 and DLGS152/06 but also achieves comparable precision, accuracy, and detection limits to helium.

Similarly, hydrogen, when used as a carrier gas, demonstrates strong potential in targeted applications, providing results consistent with helium in many cases. The introduction of the HydroInert™ source has been fundamental in mitigating challenges such as sensitivities issues and secondary effects, ensuring compatibility with hydrogen for GC-MS analyses. While certain untargeted approaches may still require further refinement to avoid unconventional fragmentation, the results support hydrogen as a viable alternative in many analytical contexts.

The quantification of target compounds using these alternative gases fully complies with regulatory requirements, ensuring reliability and robustness. As a future perspective, validating these methods under EPA 8260 standards with renewable gases could further solidify their application, paving the way for widespread adoption in routine analytical practices.

The last pillar refers to the innovative approach pursued in GAC, which connects to the analysis conducted in the last research part. By employing a high-information-capacity technique, it was possible to monitor and identify numerous analytes turning into the possibility of defining simultaneously multiple samples' traits related to quality and authenticity. This achievement highlights the power of multitarget methodologies, which align with the 8th principle of GAC, emphasizing the need for methods capable of delivering comprehensive information in a single analysis. Additionally, the adoption and development of advanced tools and approaches for data processing and data mining were integral to this work, facilitating the efficient extraction and interpretation of complex datasets. This combination of high-throughput analysis and innovative data handling underscores the commitment to sustainability and efficiency in modern analytical workflows.

In particular, the paper entitled *“Validation of a High-Throughput Method for the Accurate Quantification of Secondary Products of Lipid Oxidation in High-Quality Hazelnuts (Corylus avellana L.): A Robust Tool for Quality Assessment”* introduces the use of chromatographic fingerprinting and multi-target quantification for characterizing high-quality hazelnuts, laying the foundation for robust AI decision-making tools. The enhanced separation power of GC×GC improves specificity and accuracy in quantifying key quality markers, enabling consistent data transfer across time and labs. This approach provides objective assessments of aroma, storage quality, harvest origin, and spoilage, supporting strategic decisions in the food industry, especially in the context of climate change and market fluctuations, guiding warehousing, storage, and breeding strategies.

By *“Augmented Visualization by Computer Vision and Chromatographic Fingerprinting on Comprehensive Two-dimensional Gas Chromatographic Patterns: Unraveling Diagnostic*

Signatures in Food Volatilome” it was demonstrated the effectiveness of combining CV with chromatographic fingerprinting for Augmented Visualization, improving access to compositional data and the interpretation of complex phenomena. The proposed workflow realigns and compares chromatograms, revealing analytes with informative changes in butter production. It is flexible, reliable, and compensates for temporal misalignments. Validation confirms its consistency, supporting its application in food quality and other fields, with AI algorithms enhancing decision-making.

In the manuscript *“Translating 2D-Chromatographic Fingerprinting to Quantitative Volatilomics: Unrevealing Compositional Changes in Maize Silage Volatilome for Robust Markers Discovery”* quantitative volatilomics was successfully applied to analyze the complex volatilome of maize silage. The translation of a GC×GC-TOF MS method to a differential-flow modulation platform preserved key analytical features, enabling accurate quantification. Both methods were equally effective in discriminating silage quality and stability markers, providing valuable insights into silage fermentation and supporting the identification of key volatiles linked to product quality. This approach offers a robust tool for assessing silage and its fermentation process.

The study *“Beyond Current Quality Indices: Quantitative Volatilomics Unrevealed Cultivar Traits, Harvesting Practices Impact, and Aroma Blueprint of Extra-Virgin Olive Oils”* applies the concept of AI smelling to predict aromatic traits in extra-virgin olive oil (EVOO) using advanced chromatographic techniques. Without quantification, aroma prediction would not have been possible, and accurate quantification was achieved through the enhanced separation power of 2D-GC. This technique allowed precise identification of key volatiles, making it a powerful tool for predicting and controlling EVOO’s aromatic profile. The integration of AI-driven analysis further enhances the accuracy and reliability of these predictions.

Acknowledgements

A conclusione di questo elaborato, desidero menzionare tutte le persone, senza le quali questo lavoro di tesi non esisterebbe nemmeno. Il primo ringraziamento lo dedico alla mia supervisor Prof.ssa Chiara Cordero. La sua competenza, le sue qualità, la sua tenacia mi hanno accompagnato, supportato, spronato e, soprattutto, fatto crescere personalmente e professionalmente sin dal primo giorno, quando entrai a svolgere il mio periodo di tesi sperimentale per il completamento del mio percorso di studi. Da quel momento un po' di primavera sono sbocciate, le peculiarità prima citate sono rimaste sempre le stesse ed io ho potuto sempre più accrescere la mia persona e la mia cultura sollecitato da questi stimoli.

Grazie alle professoressa Cagliero, Liberto, Rubiolo e Sgorbini per essere state un punto di appoggio e di riferimento nei momenti di difficoltà e per avermi fatto sperimentare la bellezza di lavorare in un team unito in cui nessuno viene lasciato al suo destino.

Un ringraziamento al Prof. Bicchi, per il suo grande carisma. Credo che, non solo per me, sia stato l'impersonificazione della figura del ricercatore per essenza che, desideroso di trovare risposta ad un quesito che si è posto innanzi, assetato e vorace di conoscenza trova la risposta nel libro della sua *scientia* e *sapientia*. È stato per me quel personaggio a cui tutti noi dovremmo tendere per la riuscita delle nostre piccole e grandi imprese.

Un ringraziamento particolare ai miei colleghi di laboratorio, per avermi consigliato, seguito ed aiutato durante questo percorso ed introdotto con pazienza nel mondo della ricerca.

Un ringraziamento al Dott. Miliazza e al Dott. Stani per il supporto e la fiducia datomi durante i periodi svolti in azienda e non solo. Un grazie doveroso va ai team dei laboratori di SRA Instrument Italia e Francia che con il loro aiuto e la loro conoscenza mi hanno fatto scoprire ed apprezzare la faccia della medaglia del mondo industriale.

Un grazie a tutti i miei amici che hanno condiviso con me questi bellissimi anni e con i quali sono cresciuto. Grazie per tutti i bei momenti passati assieme e per tutti quelli che ancora ci aspettano.

Il più grande ringraziamento va alla mia famiglia, in primis ai miei genitori per avermi cresciuto, amato e concesso ogni possibilità. Un grazie ai miei cugini e i miei zii per essere stati quei porti sicuri nel mare in tempesta.

Infine un grazie ai miei nonni per essere quegli esempi di vita, quei fari di luce nella notte, quegli anelli di congiungimento tra il passato e il futuro della nostra famiglia.