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

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.

Johannes Kiefl[1], Chiara Cordero[2*], Luca Nicolotti[2], Peter Schieberle[1], Stephen E. Reichenbach[3] and Carlo Bicchi[2]

[1]Deutsche Forschungsanstalt für Lebensmittelchemie, Lise-Meitner-Straße 34, 85354 Freising, Germany

[2]Dipartimento di Scienza e Tecnologia del Farmaco, Università degli Studi di Torino, Via P. Giuria 9, I-10125 Torino, Italy

[3]Computer Science and Engineering Department, University of Nebraska , 1400 R Street,Lincoln, NE 68588-0115, USA

* Address for correspondence:
Dr. Chiara Cordero - Dipartimento di Scienza e Tecnologia del Farmaco, Università degli Studi di Torino, Via P. Giuria 9, I-10125 Torino, Italy – e-mail: chiara.cordero@unito.it ; phone: +39 011 670 7662; fax: +39 011 2367662

Abstract

The continuous interest in non-targeted profiling induced the development of tools for automated cross-sample analysis. Such tools were found to be selective or not comprehensive thus delivering a biased view on the qualitative/quantitative peak distribution across 2D sample chromatograms. Therefore, the performance of non-targeted approaches needs to be critically evaluated. This study focused on the development of a validation procedure for non-targeted, peak-based, GC×GC-MS data profiling. The procedure introduced performance parameters such as specificity, precision, accuracy, and uncertainty for a profiling method known as *Comprehensive Template Matching*. The performance was assessed by applying a three-week validation protocol based on CITAC/EURACHEM guidelines. Optimized ^{1}D and ^{2}D retention times search windows, MS match factor threshold, detection threshold, and template threshold were evolved from two training sets by a semi-automated learning process. The effectiveness of proposed settings to consistently match 2D peak patterns was established by evaluating the rate of mismatched peaks and was expressed in terms of results accuracy. The study utilized 23 different 2D peak patterns providing the chemical fingerprints of raw and roasted hazelnuts (*Corylus avellana* L.) from different geographical origins, of diverse varieties and different roasting degrees. The validation results show that non-targeted peak-based profiling can be reliable with error rates lower than 10 % independent of the degree of analytical variance.
The optimized *Comprehensive Template Matching* procedure was employed to study hazelnut roasting profiles and in particular to find marker compounds strongly dependent on the thermal treatment, and to establish the correlation of potential marker compounds to geographical origin and variety/cultivar and finally to reveal the characteristic release of aroma active compounds.

Keywords: GC×GC-MS; validation; comprehensive template matching fingerprinting; roasted hazelnut; *Corylus avellana* L. volatile fraction; key aroma markers.

1. Introduction

Comprehensive two-dimensional gas chromatography coupled with mass spectrometry (GC×GC-MS) is a powerful tool for targeted and non-targeted analysis of complex mixtures of volatile compounds due to the enhanced peak capacity compared to one-dimensional GC [1-3]. Non-targeted fingerprint analysis can reveal qualitative/quantitative differences in chemical compositions facilitating the identification of potential marker compounds [4,5] and grouping or classification of samples [6,7]. Recent publications have described the development of non-targeted, peak-based fingerprinting tools to exploit the informative content of three dimensional GC×GC-MS data sets characterized by first (^{1}D) and second (^{2}D) dimension retention times and mass fragmentation patterns [5,6,8-12]. Such tools (name of the tool is set in brackets) were developed by Oh et al. (MSort [8]), Wang et al. (DISCO [9]), Kim et al. (mSPA [12]), Castillo et al. (Guineu [11]), Almstetter et al. (INCA [5]), and Leco Corporation (Statistical Compare, St. Joseph, USA). The latter provides commercially available software that was tested by Almstetter et al. [10]. In a recent review, Reichenbach et al. [13] characterized such approaches as *peak feature analysis*. *Peak features* collate the response data of individual analytes across chromatograms; for comparative analysis, analyte peaks must therefore be consistently matched across 2D chromatograms. However, consistent *peak feature analysis* remains challenging because retention times and mass spectra are subject to run-to-run variations due to random or systematic errors depending on sample preparation, injection, chromatographic and mass spectrometric conditions. This process of peak matching is a critical step of data elaboration, because matching errors produce false qualitative and quantitative differences [13], thus complicating the ongoing data interpretation.

The extent of variability, generated during analysis, can be minimized by adopting automated sample preparation techniques, validated protocols for sample processing, and robust instruments. On the other side, software for data processing should reliably match corresponding analyte peaks, within a set of sample chromatograms, by accounting for the remaining variation on, for example, ^{1}D and ^{2}D retention times and fragmentation pattern intensities. This variability can be propagated during data processing, if the *feature* content (i.e., response values of one analyte matched across many chromatograms) is erroneously computed and/or elaborated by the software. Data processing errors can occur at different stages and have partially been addressed: (a) removal of background [14], (b) peak detection

[15], (*c*) recognition of retention time shifts [16], and (*d*) peak alignment [5,8-12,16-18]. For example, peak detection thresholds are used to detect and integrate peaks above a certain signal-to-noise ratio (S/N). Therefore, trace peaks may be detected in some samples, but not in others [13]. This inconsistent peak detection may be propagated by the peak alignment routine by matching just detected peaks while ignoring corresponding analyte peaks with lower S/N. Approaches to eliminate such propagation of errors have scarcely been evaluated, but this issue is critical for the challenging task of a consistent and unbiased *peak feature analysis*. A further example is given by Castillo et al. [11]; the authors compared sixty 2D chromatograms of serum samples getting an aligned data file with 14756 compounds. However, only 1540 compounds were aligned across minimum six 2D chromatograms [11]. This indicates that 90 % of the detected peaks were lost during data processing and these might have included diagnostic analyte peaks.

Non-targeted, peak-based fingerprint analysis should produce reliable data on the qualitative/quantitative distribution of compounds within a sample set. Then, results can be interpreted and research can provide compound identifications, structure elucidation of unknowns, or quantitative information. It is proposed that the quality of such investigations can be assured by applying an appropriate validation procedure. The analytical protocol adopted in the present study was designed and validated in accordance with general protocols (Eurachem/Citac guidelines [19,20]) to systematically investigate method performance parameters (i.e., specificity, repeatability and intermediate precision, accuracy and uncertainty) as a function of the most critical variables (data acquisition and data elaboration levels). An existing alignment tool was adopted using *Comprehensive Template Matching* as introduced by Reichenbach et al. [21]. The validated *Comprehensive Template Matching* procedure was then applied to a food-type intervention study designed to identify marker peaks highly informative for the hazelnuts roasting process. Various raw hazelnuts were subjected to a roasting protocol to induce chemical changes which were recorded with the help of the alignment routine. Hazelnuts are an ideal model system, because different known and unknown formation pathways are generating odorless volatiles and odorants with different kinetics [22,23]. Therefore, non-targeted analysis is useful for studying the biological and/or chemical response after perturbation or technological intervention and provides significant and valid information about interaction-relationships in a food system.

2. Materials and Methods

2.1 Reference compounds and solvents

Pure reference compounds for identity confirmation and *n*-alkanes (*n*-C9 to *n*-C25) were supplied by Sigma-Aldrich (Taufkirchen, Germany) except α/β-thujone (α/β-1-isopropyl-4-methylbicyclo[3.1.0]hexan-3-one, 95/5 weight ratio) supplied by Fluka (Milan, Italy). A standard stock solution of α/β-thujone diluted to 45 ng/mL was prepared in ultrapure water and the solution was stored in a sealed vial at 5 °C. Both, α- and β-thujone, were used as internal standards for peak response normalization (ISTD) adopting the ISTD loading procedure [24,25]. Solvents (cyclohexane, n-hexane, dichloromethane) were all HPLC-grade from Riedel-de Haen (Seelze, Germany). 3-Methyl-4-heptanone was synthesized according to [22].

2.2 Hazelnut samples and roasting

Raw and shelled hazelnuts of *Corylus avellana* L. (harvest 2009, supplied by Marchisio, Cortemilia- CN, Italy) from different geographical origins and varieties/cultivars were roasted at 160°C in a ventilated oven for 7, 12, 17 and 23 minutes. Tonda Gentile Romana (*Romana*, Lazio, Italy), Tonda Gentile delle Langhe (*Gentile*, Piedmont, Italy), Tonda di Giffoni (*Giffoni*, Campania, Italy) were monovarieties and Azerbajian hazelnuts were a blend of different locally grown cultivars. Roasting was conducted every day and the hazelnuts of a uniform dimension (caliber within 12-13 mm) were left at room temperature to cool down. No storage of manually roasted hazelnuts was necessary, thereby avoiding an alteration of the volatile fraction. The hazelnut samples were manually ground prior to vial filling, and the particle size was compared to a ground reference sample.

Standard roasted hazelnuts (harvest 2009, supplied by Soremartec Italia SpA, Alba-CN, Italy) of Romana, Gentile and Giffoni were submitted to roasting in an industrial plant at different time/temperature ratios consistent with their desirable final sensory characteristics. These samples were hermetically sealed under vacuum in non-permeable polypropylene/aluminium/polyethylene packages and stored at -20°C prior to analysis.

148 **2.3 Headspace Solid Phase Microextraction (HS-SPME) devices and sampling conditions**

149 The SPME device and fibers were from Supelco (Bellefonte, PA, USA). A
150 Divinylbenzene/Carboxen/Polydimethylsiloxane (DVB/CAR/PDMS) df 50/30 µm, 2 cm length
151 fiber was chosen and conditioned before use as recommended by the manufacturer.
152 Roasted hazelnuts (1.5 g) were ground, sealed in a 20 mL vial, and equilibrated for 20 min at
153 50°C before sampling. The internal standard (ISTD) loading procedure onto the SPME fiber
154 [24,25] was as follows: the SPME device was manually inserted into a 20 mL sealed vial
155 containing 1 mL of ISTD (α/β-thujone) standard solution at 45 ng/mL, then the fiber was
156 exposed to the headspace at 50°C for 20 min. After ISTD loading, the fiber was exposed to
157 the matrix headspace at 50°C for another 20 min. Just the lower part of the vial, filled with
158 the solid sample, was heated to improve the vapor phase/fiber coating distribution
159 coefficient. The SPME device was directly introduced into the GC injector for thermal
160 desorption for 10 min at 250 °C. Samples for validation purposes were analyzed in triplicate
161 each week. Relative standard deviations (RSD %) for first and second retention times and 2D-
162 peak normalized volume (i.e., cumulative 2D peak area) for 24 identified target analytes are
163 given (**Table 1 and S1).**
164

165 **2.4 GC×GC-qMS instrument set-up**

166 GC×GC analyses were performed on an Agilent 6890 GC unit coupled with an Agilent
167 5975 MS detector operating in the EI mode at 70 eV (Agilent, Little Falls, DE, USA). The
168 transfer line was set at 280°C. A *Standard Tune* was used and the scan range was set at *m/z*
169 35-250 with a scan rate of 10,000 amu/s to obtain a suitable number of data points for each
170 chromatographic peak for reliable identification and quantitation. The system was equipped
171 with a two-stage KT 2004 loop thermal modulator (Zoex Corporation, Houston, TX) cooled
172 with liquid nitrogen and with the hot jet pulse time set at 250 ms with a modulation time of
173 4 s adopted for all experiments. Fused silica capillary loop dimensions were 1.0 m length and
174 100 µm inner diameter. The column set was configured as follows: ^{1}D Carbowax CW20M
175 column (100% polyethylene glycol)(30 m × 0.25 mm i.d., 0.25 µm df) coupled with a ^{2}D
176 OV1701 column (86% polydimethylsiloxane, 7% phenyl, 7% cyanopropyl) (1 m × 0.1 mmi.d.,
177 0.10 µm df). Columns were from Mega (Legnano, Milan, Italy).

178 One microliter of the *n*-alkane sample solution was automatically injected with an
179 Agilent ALS 7683B injection system under the following conditions: split/splitless injector,

split mode, split ratio 1:50, injector temperature 280°C. The HS-SPME sampled analytes were thermally desorbed from the fiber for 10 min into the GC injector using the following conditions: split/splitless in split mode, split ratio 1:20, injector temperature 250°C. The carrier gas was helium at a constant flow of 0.7 mL/min (initial head pressure 260 KPa). The temperature program was 40°C (2 min) to 180°C at 2.5°C/min and to 250°C at 20°C/min (5 min). Data were acquired by Agilent MSD ChemStation ver D.02.00.275 and processed using GC Image GC×GC Software version 2.1b1 (GC Image, LLC Lincoln NE, USA). Statistical analysis was performed with SPSS 14.0 (SPSS Inc. Chicago, Illinois, USA).

2.5 Validation protocol and data elaboration

The validation protocol was designed and applied to characterize the following performance parameters: specificity, repeatability and intermediate precision of the analytical method (sample preparation and separation), and accuracy of results related to the *Comprehensive Template Matching* methodology. The five-days/three-weeks validation scheme is summarized in **Table S2** .

Raw data files were loaded into the software and were background corrected according to [14]. The 2D peaks were automatically detected by fixing a *peak signal-to-noise ratio* (S/N) threshold of 10 and a footprint-area threshold of 10. The peak detection uses the watershed algorithm [15]. A set of 24 target peaks (**Table 1**), selected over the GC×GC patterns to cover homogeneously the chromatographic run, were matched over nine replicates of standard roasted Romana samples and over nine replicates of self-roasted Romana samples to evaluate the inter- and intra-week variability of retention times, 2D peak normalized volumes (normalization was done over the ISTDs 2D peak volume), and mass spectra match factors. Target analysis was performed as described in [3]. Normality of the distribution of normalized 2D peak volumes and normalized S/N values was tested with Kolmogorov-Smirnov and Shapiro-Wilk test with p = 0.05 (Origin 6.1, OriginLab Corporation MA, USA).

To determine optimized peak matching parameters, ^{1}D and ^{2}D retention times search windows were fixed on the basis of the average standard deviations of retention-times data (**Table 1**). The thresholds were set to three times the standard deviation for each dimension giving a retention time window of 5 modulations for the ^{1}D and 0.17 seconds for the ^{2}D dimension. Therefore, the peaks could be expected to elute within this time window with a

high probability (here 99.7 %) considering run-to-run retention time shifts. Based on the reported MS match factors (**Table 1**), a match factor threshold of 600 was used (see below) to confirm target identity, accepting again a run-to-run match factor variability of three times the standard deviation.

Validation was performed *(a)* to assess the fingerprinting method specificity by determining false-positive and false-negative error rates of the peak matching process; *(b)* to assess precision (repeatability and intermediate precision of standard roasted Romana samples) and to estimate the contribution of manual roasting to data dispersion (self-roasted Romana samples); *(c)* to evaluate fingerprinting method accuracy; and finally *(d)* to assess the fingerprinting method uncertainty.

After the validation step, *Comprehensive Template Matching* fingerprinting was adopted to investigate chemical changes on extended data set samples of different geographical origins and roasting conditions. A schematic work-flow of the procedure is reported in **Figure 1** and details are discussed in the next paragraphs. The 23 sample patterns were processed to build a *consensus template* by adding all the detected peaks, exceeding fixed S/N and area thresholds from one randomly selected chromatogram within the set to an empty template. Then, this template was matched with a second chromatogram to add unmatched peaks, and the resulting “updated template” was matched with a third chromatogram of the set and this was repeated with all patterns [6]. An automatic retention time alignment was here used to compensate for pattern shifts due to the intrinsic variation of chromatographic performance [16].

Each non-target peak included in the *consensus template* was saved with its chromatographic descriptors (see above) and every entry was labeled with a unique number to unequivocally identify each non-target peak.

3. Results and Discussion

This section is divided into two parts. The first part describes the routine adopted to assess performance parameters of the fingerprinting method. The second part is dedicated to a practical example where advanced fingerprinting is applied to profile the roasting process of hazelnuts of different origin and variety/cultivar altered by increasing roasting time at constant temperature.

3.1 Assessment of method performance parameters

3.1.1 Specificity

Specificity is strictly related to the chromatographic separation and requires that a peak is correctly assigned to an analyte [19]. From a chromatographic point of view, the separation of hazelnut volatiles presented in this study is the result of an optimization procedure based on systematic columns selection aimed at finding the best orthogonal column combination that minimizes the number of coelutions and maximizes the 2D peak spreading over the chromatographic space [7]. With a view on data elaboration,, peak-based *features* evaluate the analyte responses across many chromatograms, where one feature can contain information of spuriously matched peaks. Specificity was evaluated in three steps: *(a)* optimization of peak matching parameters (i.e., ^{1}D and ^{2}D retention times search windows and MS match factors threshold); *(b)* matching of selected targets reported in **Table 1** across nine replicate samples to establish peak correspondences and to verify peak identities; and (*c*) analysis of blank samples to locate and remove interfering peaks.

The optimum peak matching involves low levels of false-positive and false-negative hits, so that an analyte is consistently matched across several chromatograms and not mismatched with impurity/blank peaks or other interfering analytes. **Figure 2** illustrates the effect of different matching criteria on the peak matching specificity. When a template with 8 peaks (thin circles) is matched by defining just ^{1}D and ^{2}D retention times constraints (**Figure 2a**), analyte peaks are positively matched if the Euclidean Distance between sample and template peaks does not exceed the user-defined retention time window threshold (here of 5 modulations for the ^{1}D and 0.17 seconds for the ^{2}D dimension). In this case, false-positive matches are denoted for peaks 1, 2 and 5-8. When a MS constraint with 600 match factor is added, peak matching gives consistent assignment (**Figure 2b**). However, the average MS match factors and the corresponding relative standard deviations (**Table 1**)

indicate that an increasing MS threshold causes an increasing number of analyte peaks to fail the matching criterion, thus producing false-negative hits. **Figure 2c** illustrates the effect of a too restrictive peak matching obtained by increasing the MS threshold to 774 (**Table 1**): only 3 peaks of 8 template peaks are now correctly matched.

Results reported in **Table 1** confirm that the method is specific for matching an extended set of target analytes containing resolved – unresolved peaks as well as high and low abundant peaks. Peak 15 (nonanal), for example, showed a higher variation of normalized volume, because of its long-tail and consequent difficult peak integration. False-positive and false-negative error rates were zero percent and the true-positive rate was 100 % for 24 target analytes matched across nine replicate chromatograms analyzed over three weeks, supporting the effectiveness of the settings for the selected retention time search window and minimum match factor.

On average, 179 2D peaks were detected above S/N 10 in blank sample runs (**Table S2**) and were associated to fiber bleeding, column bleeding or impurities derived from solvents and ISTDs reference material. Templates were matched against the blank runs and matched peaks were removed to obtain pruned templates which were then matched against sample chromatograms.

Literature dealing with peak-based comparative analysis also aimed at improving the specificity of the data elaboration. For example, Oh et al. [8] determined true-positive and false-positive error rates by pair-wise matching the mass spectra of 46 derivatized standards in 16 samples. A standard was positively matched when the Pearson correlation coefficient was near one; a true-positive rate of 92 % and a false-positive error incidence of 11 % was obtained. Retention times, however, were not used as constraints. Wang et al. [9] spiked 6 analytes and correctly matched 5 analytes across 5 replicate chromatograms. Kim et al. [12] used mass spectra and retention times to match corresponding peaks across metabolite samples from rat plasma; a true-positive rate of 70 % was determined. Almstetter et al. [5] spiked 20 standard compounds to *E. coli* extracts and optimized the peak matching procedure until true-positive rate was 100 %. This is in accordance with results reported here, except that our peak matching optimization calculates threshold values using simple and intuitive peak descriptors, i.e., retention time standard deviations and mass spectrum match factor standard deviations, while other approaches adopted iteratively measured threshold values requiring extensive computational work [5,8,9,12]. Thus, optimization of

matching parameters can be done routinely using a training set of sample chromatograms and relative standard deviations of matching parameters calculated from peak features.

3.1.2 Precision

Two levels of precision [26], *repeatability* and *intermediate precision*, were evaluated. The *repeatability* of the separation method was assessed by analyzing standard roasted Romana and self-roasted Romana samples in single weeks, with the same instrument, laboratory, and operator, and the *intermediate precision* was calculated over a period of three weeks. *Repeatability* and *Intermediate precision* were calculated as relative standard deviations (RSD %) for chromatographic peak volumes (normalized volume *vs.* ISTD) and retention times (^{1}D and ^{2}D t_R) for each analyte in each validation week and over all weeks. Results showed a good intermediate precision for retention times, with maximum averaged RSD values of 1.0 % and 3.2 % for ^{1}D and ^{2}D respectively (**Table 1)**.

A higher dispersion is evident for quantitative data (normalized volumes). The normalized volumes for standard roasted Romana samples were submitted to the Analysis of Variance (ANOVA) to evaluate the comparability of data between weeks. The One-Way ANOVA on the nine sample replicates collected over the three-weeks revealed that the null hypothesis, “there is no difference between normalized volume values measured in different weeks“ could be accepted for each of the 22 target compounds (excluding the ISTD) with $p <= 0.05$. The RSD % on normalized volumes showed an average value of 12.4 %, which is an acceptable intermediate precision (**Table 1)**. The highest RSD % value was 47.6 % for peak 15 (nonanal) probably related to its long tailing, as already mentioned above.

The average RSD % of normalized volumes of the self-roasted Romana samples is, however, larger (i.e. 46 %). This increase of dispersion is remarkable and originates from sample processing, and not from the analytical procedure. With the designed validation protocol, it has been possible: *(a)* to estimate the magnitude of this external source of variation and *(b)* to evaluate better fingerprinting accuracy in the cross-comparison of real-world samples. The dispersion registered for the self-roasted Romana samples was considered by assigning a suitable uncertainty interval to fingerprinting results.

3.1.3 Accuracy

The accuracy of the *Comprehensive Template Matching* fingerprinting data was verified [20] on nine standard roasted Romana samples and nine self-roasted Romana samples acquired over the entire validation period and considering all separated and detected 2D peaks. The extension of the validation procedure to the entire chemical pattern (i.e., all 2D peaks above the fixed threshold) is required to evaluate fingerprinting accuracy as a function of the chromatographic performance (specificity and precision) and data elaboration parameters (specificity). Therefore, chromatograms were processed with an optimized procedure to match consistently all detected analyte peaks across GC×GC patterns collected on different days, within three weeks.

First, a *Sample Template* for each chromatogram of the set was created by including all detected peaks exceeding the fixed S/N and area threshold (see section *Validation protocol and data elaboration*). Each non-targeted peak included in each *Sample Template* was saved with its ^{1}D and ^{2}D retention times, detector response, mass fragmentation pattern, match factor threshold and with a unique number. This number was assigned to a non-targeted analyte peak just after successful matching (**Figure 1**).

However, small variations of overall peak intensities were observed within the set. **Figure 3** shows the distribution of S/N values of detected peaks across replicate chromatograms. The histogram evidences that: (*a*) the S/N-threshold for peak detection should be sufficiently low, otherwise most of the peak information is discarded, and (*b*) the S/N values, like 2D peak volumes, are normally dispersed. This means that S/N values of low abundant peaks can scatter around the peak detection limit and hence can sometimes be detected/matched and sometimes not (**Figure 4a**). Thus, only analytes with S/N values always above the peak-detection limit were uploaded into the *Sample Template* to keep a consistent peak-matching. This was automatically done using internal standardization and template thresholds that are higher than the peak detection limit (**Figure 4b**). This procedure enables reduction of the rate of mismatches resulting from template peaks that cannot be matched with the corresponding analyte peaks, because they were not detected. The template threshold was calculated for every chromatogram adopting the S/N of α- and β-thujone (ISTDs). The value was fixed as three times the standard deviation of S/N values from the least intense chromatogram peaks (i.e., those peaks just above the peak detection limit) multiplied by the ISTD S/N abundance. The resulting thresholds are reported in **Table 2**.

The closeness of agreement (i.e., accuracy) for the fingerprinting was verified through a *pair-wise comparison* showing that the method achieves a high similarity rate, which is always within a fixed interval and always above a certain value (**Table 2**) for samples of the same origin and roasting conditions acquired over the entire validation period. The similarity rate is here expressed as percentage of matched peaks between sample pairs: a maximum of 100% is expected for patterns obtained by analyzing the same sample even over an extended time period.

The results show that arbitrarily selected *Sample Templates* reliably match the chromatograms of all replicates (**Table 2**). Accuracy results, expressed as percentage of matched peaks, ranged from a minimum of 91 % (for the sample acquired in week-one, day-two) to a maximum of 95 % with 2 % RSD (standard roasted hazelnut samples). The percentages of matched peaks of self-roasted hazelnut samples ranged from a minimum of 85 % (for the sample acquired in week-three, day-three) to a maximum of 95 % with 3 % RSD. As a consequence, the peak matching performance proved to be robust for samples with high dispersive quantitative values (standard roasted samples 12.4 RSD % and self-roasted samples 46.2 RSD % on normalized peak volumes).

The qualitative differences in peak patterns are important and should consistently be "extractable" by compensating retention time shifts (cf. 3.1.1) and dispersive response values through appropriate template threshold values. On the contrary, problems with consistent peak matching are often reported in literature. Oh et al. [8] compared eight replicate chromatograms of a derivatized fatty acid/organic acid mixture with eight replicate chromatograms of a derivatized fatty acid/amino acid mixture and resulted in a list with 8683 features, while only 46 were expected to be generated from standard compounds. Performing a metabolite profiling on wild-type *vs.* double mutant *E. coli* strains, Almstetter et al. [5] obtained a list of 2259 features from nine replicates per sample group (a peak detection threshold of 500 was fixed). To limit the number of entries, the authors excluded those features that could not be matched in at least 9 of 18 samples resulting in 398 peaks. Castillo et al. [11] studied the metabolic profile of 60 human serum samples, from which 14756 features were extracted, but only 1013 features were found to be useful. Kim et al. [12] aligned 5 replicate samples of rat plasma with an average of 446 analyte peaks per 2D chromatogram, but just 146 peaks could be matched throughout all replicates. Although these works might have handled differently complex peak patterns (depending on sample

composition, sample preparation and injection technique), the evaluation and optimization of fingerprinting accuracy could help to reliably extract their true qualitative differences.

In our study, analyte peaks (on average 166 for each standard roasted sample chromatogram) could successfully be extracted from the larger number of detected peaks (in average 387 for standard roasted samples) after a systematic optimization of data elaboration parameters. As a consequence, analyte peaks could reliably be matched across replicate chromatograms acquired over three weeks with an average of 94 % of true-positive peak matches. Manual investigation of the missing peak matches revealed that especially coeluting compounds were not reliably detected by the software (**Figure 2**). This limit might be overcome by adopting a suitable deconvolution/unmixing algorithm for unresolved peaks.

3.1.4 Uncertainty

The goal of the fingerprinting is to reveal qualitative and quantitative differences within a set of samples. As a consequence, the uncertainty [20,21] should account for the dispersion of the quantitative data, mainly influenced by sample preparation and chromatographic separation, and the consistency of the qualitative data (% of matched peaks) mostly influenced by method specificity and accuracy.

The combined standard uncertainty can be calculated through the classical equation [20]:

$$[u(y)/y]^2 = [u(y)_A/y]^2 + [u(c)_B/c]^2 + [u(d)_B/d]^2 + [u(f)_B/f]^2 + [u(p)_B/p]^2$$

Equation 1

where $u(y)/y$ is the combined standard uncertainty for the measurand y; $u(y)_A/y$ is the uncertainty referred to as repeatability and intermediate precision data; $u(c)_B/c$ is the uncertainty derived from calibration data; $u(d)_B/d$ is the uncertainty derived from dilution; $u(f)_B/f$ is the uncertainty derived from the efficiency of the method; and $u(p)_B/p$ is the uncertainty derived from errors on weight.

The contribution to the combined uncertainty of the method of sample weighing and dilution were negligible being respectively 1.0E-6 and 1.0E-4. Calibration was not included in this analytical procedure; thus, the uncertainty range, to be associated to normalized volumes after successful peak matching, was expressed through $u(y)_A/y$ (i.e., *Repeatability* of

the method). The relative uncertainty associated with the quantitative results was calculated as +/-12.4 %.

On the other hand, the standard uncertainty for the *Comprehensive Template Matching* fingerprinting to be associated with the result of the cross-comparison of samples and expressed as percent of matched peaks, was referred to as the accuracy of the data (i.e., % error) of the peak-based fingerprinting. The relative uncertainty associated with the fingerprinting results was +/-6 %.

3.2 Application of the validated pair-wise Comprehensive Template Matching fingerprinting in profiling the roasting process of hazelnuts.

Comprehensive Template Matching fingerprinting has been demonstrated to be a general tool to compare sample fingerprints without any knowledge of sample composition. To be truly comprehensive, the fingerprinting procedure should evaluate the complete peak information and, as a consequence, all 2D analyte peaks of all sample chromatograms should accurately be aligned.

This is demonstrated by studying the effect of roasting at different time-temperature profiles on hazelnut volatile development. In view of this, the validated *pair-wise* comparison method was applied to extract temperature-sensitive *features* from a representative sample subset (here Tonda Gentile). The resulting *consensus template* then was used to extract qualitative/quantitative data on analytes in all 23 chromatograms. This was done to verify the independence of the potential marker compounds from the geographical origin and variety/cultivar of the hazelnuts. The *consensus template* was created according to the procedure described in *validation protocol and data elaboration* section (**Figure 1**).

Marker compounds indicating roasting should be detectable at an early stage of heat treatment. The potential marker compounds are detectable within 7 minutes of roasting, making it reasonable to create a *consensus template* for the Tonda Gentile with samples of 0 and 7 minutes roasting time. The c*onsensus template* of the sample pair was next matched with all chromatograms of the Tonda Gentile subset. The average matching rate was 88 % (*n*=5, i.e., 0, 7, 12, 17 and 23 minutes roasting) across all chromatograms. Peak data were automatically aligned. The resulting list, containing the respective normalized peak volume values, was sieved to define significant *features* (**Table 3**). The fold changes of increasing

response values were calculated relative to the peak responses of the 7 min roasted hazelnut sample. A two-fold increase (corresponds to minimum +200%; **Table 1**) was required to define a significant *feature* for self-roasted hazelnut samples. Finally, 24 of the 83 features were regarded as "significant indicators" for the roasting process of Tonda Gentile samples. Ongoing data reduction addressed potential roasting markers that were independent of geographical origin or variety. Thus, normalized peak volume values of the selected 24 supposed markers were extracted from all chromatograms of the entire sample set, using specific quantifier ions (**Table S3**), and submitted to a Principal Component Analysis (PCA)

PCA was used for an unsupervised analysis and was performed initially on each variety/origin independently (**Figure S1**) to show the degree of correlation between potential marker compounds and roasting time. The first principal component (roasting degree) explained on average 71 % of the total variance. Several marker *features* had loadings >+/-0.6 on component 2, indicating dependence on an unknown factor. *Features* with loadings of more than 0.6 for component 1 and loadings of <+/-0.6 for component 2 were deemed strong markers for "roasting degree" thus enabling the removal of visually recognizable "outliers". The number of potential markers was thereby reduced from 24 to 11 (**Table 3**). Linear regression analysis on normalized volume values of the proposed 11 marker compounds for all hazelnut samples (**Figure 5**) revealed a nearly proportional relationship between increased roasting time and increased normalized peak volume values, with R^2 of 0.8147. A look backward into raw data (**Figure 3**) revealed that normalized volumes of markers, e.g., 2,3-pentandione and unknown compound **81**, steadily increased with roasting time, whereas 5-methyl-(E)-2-hepten-4-one and unknown compound **79**, excluded on the basis of the PCA screening, reached their maximum after 12 minutes (**Figure S1, S3**).

This *Comprehensive Template Matching* fingerprinting method combined with a food-type intervention study has proven to be reliable and straightforward to select significant peak data from a set of 23 unique sample files. The simple linear model derived from marker compounds, for example, allows an interpolation and prediction of roasting time, thereby facilitating automated food processing. With manual roasting, the marker compound responses considerably vary (Figure 5) limiting the possibility to distinguish between smaller time intervals. These results also provide deep insights into the formation

of volatile compounds during roasting and differences emerging from different varieties of different origin. Alasalvar et al. reported on the formation of 71 volatile compounds from hazelnuts from Tombul (Turkey) [27]. Burdack-Freitag and Schieberle [22] recently investigated the formation of key aroma compounds during roasting of Tonda Romana (Italy) hazelnuts, showing that 2-methyl-butanal (**8**, malty), 2,3-pentandione (**19**, buttery), phenylacetaldehyde (**77**, flowery), and 5-methyl-(E)-2-hepten-4-one (**52**, filbertone, nutty) have an important impact on the aroma of roasted hazelnuts. However, GC×GC-qMS has been used in this study to explore systematically the effect of roasting at different time intervals on the formation of odor-active as well as odorless volatile compounds. Concentrations of identified odorants **8**, **19**, **77** (2-methylbutanal, 2,3-pentandione, phenylacetaldehyde) increase almost linearly with roasting time independently on variety suggesting them as suitable marker compounds to predict the degree of roasting. Future studies could investigate whether these odorants can evoke different distinct aromas despite their similar release profiles across different hazelnuts. 1-Methyl-pyrrole (**29**), pyridine (**35**), 3-hydroxy-2-butanone (**50**), 2-ethyl-pyrazine (**54**), furfural (**66**) and the three unknowns **23**, **73**, **81** belong to the group of proposed roasting markers (**Table 3**). Some of them, e.g., furfural and pyrazines are known Mailliard reaction products [28,29]. It is assumed that the reaction of amino compounds and reducing sugars is mediated by heat and fits well with our observation of almost linear increase with increasing roasting time. In contrast, a second group of temperature-sensitive compounds has been identified (**13**, **22**, **36**, **41**, **45**, **52**, **58**, **63**, **71**, **74**, **78**, **79**, **80**). They are not roasting marker compounds, because they lack in the typical release profile at least in one variety (**Table 3**). For example, 5-methyl-(E)-2-hepten-4-one is similarly released in the Gentile and Giffoni samples, whereas its level in Azerbaijan sample remains low (**Figure S3**). According to this, compound *79* tracks approximately the same profile for the Gentile, Giffoni and Romana samples, but not for the Azerbaijan hazelnuts (**Figure S3**). The release profiles of these components show that their normalized volume values rapidly increase and remain constant or decrease with longer roasting times (17-23 min). The formation of these ketones is still unclear, although **52** was suggested to originate from a yet unknown precursor [23].

4.Conclusions

This study presents a systematic approach to evaluate the *fitness for purpose* of a peak-based fingerprinting method, Comprehensive Template Matching fingerprinting. The reliability of the proposed method was confirmed by employing performance parameters such as specificity, precision, accuracy, and uncertainty [30], and following a general validation protocol based on Eurachem/Citac guidelines [19,20].

The results of the comparative 2D data analysis were improved by properly compensating the dispersion of detector response values through an appropriate tuning of the main data elaboration parameters (i.e., ^{1}D and ^{2}D retention times search windows, MS match factor threshold, detection threshold, and template threshold) with two sets of training samples showing different degrees of analytical variance (i.e., industrially and manually prepared sample material). Optimization was done by non-iterative standardized procedures that could be fully automated by the GC-Image software.

The validated non-targeted, peak-based fingerprinting method has successfully been applied to elucidate the generation of volatile compounds during roasting in a set of 23 hazelnut samples, where 11 roasting markers were identified, and to study the release of key aroma compounds showing specific profiles as a function of variety/origin of hazelnut samples.

Further investigations are under way: a) to study the effect of data elaboration settings on the result quality when, for example, high density data from high frequency MS detectors or high resolution MS are studied and b) to investigate a possible correlation between concentration changes of character-impact compounds with overall odor impressions, to identify further unknown significant *features* and to clarify formation pathways explaining observed trends.

Acknowledgments

The results reported in this chapter regarding work in the authors' laboratory were obtained within the "ITACA" and "ECOFOOD" projects of the POR-FESR "Competitività regionale e occupazione" 2007/2013, Asse 1, Misura I.1.1, "Piattaforme innovative" of the Piedmont Region (Italy).

References

[1] M. Adahchour, J. Beens, U.A.T. Brinkman, J. Chromatogr., A 1186 (2008) 67.
[2] K.M. Pierce, J.C. Hoggard, R.E. Mohler, R.E. Synovec, J. Chromatogr., A 1184 (2008) 341.
[3] C. Cordero, E. Liberto, C. Bicchi, P. Rubiolo, S.E. Reichenbach, X. Tian, Q. Tao, J. Chromatogr. Sci. 48 (2010) 251.
[4] X. Li, Z. Xu, X. Lu, X. Yang, P. Yin, H. Kong, Y. Yu, G. Xu, Anal. Chim. Acta 633 (2009) 257.
[5] M.F. Almstetter, I.J. Appel, M.A. Gruber, C. Lottaz, B. Timischl, R. Spang, K. Dettmer, P.J. Oefner, Anal. Chem. (Washington, DC, U. S.) 81 (2009) 5731.
[6] C. Cordero, E. Liberto, C. Bicchi, P. Rubiolo, P. Schieberle, S.E. Reichenbach, Q. Tao, J. Chromatogr., A 1217 (2010) 5848.
[7] C. Cordero, C. Bicchi, P. Rubiolo, J. Agric. Food Chem. 56 (2008) 7655.
[8] C. Oh, X. Huang, F.E. Regnier, C. Buck, X. Zhang, J. Chromatogr., A 1179 (2008) 205.
[9] B. Wang, A. Fang, J. Heim, B. Bogdanov, S. Pugh, M. Libardoni, X. Zhang, Anal. Chem. (Washington, DC, U. S.) 82 (2010) 5069.
[10] M.F. Almstetter, I.J. Appel, K. Dettmer, M.A. Gruber, P.J. Oefner, J. Chromatogr., A 1218 (2011) 7031.
[11] S. Castillo, I. Mattila, J. Miettinen, M. Oresic, T. Hyotylainen, Anal. Chem. (Washington, DC, U. S.) 83 (2011) 3058.
[12] S. Kim, A. Fang, B. Wang, J. Jeong, X. Zhang, Bioinformatics 27 (2011) 1660.
[13] S.E. Reichenbach, X. Tian, C. Cordero, Q. Tao, Journal of Chromatography A, In press:DOI: 10.1016/j.chroma.2011.07.046 (2011).
[14] S.E. Reichenbach, M. Ni, D. Zhang, E.B. Ledford, J. Chromatogr., A 985 (2003) 47.
[15] I. Latha, S.E. Reichenbach, Q. Tao, J. Chromatogr., A 1218 (2011) 6792.
[16] M. Ni, S.E. Reichenbach, A. Visvanathan, J. TerMaat, E.B. Ledford, J. Chromatogr., A 1086 (2005) 165.
[17] C.G. Fraga, B.J. Prazen, R.E. Synovec, Anal. Chem. 73 (2001) 5833.
[18] D. Zhang, X. Huang, F.E. Regnier, M. Zhang, Anal. Chem. (Washington, DC, U. S.) 80 (2008) 2664.
[19] The fitness for purpose of analytical methods. A laboratory guide to method validation and related topics., Eurachem Guide, 1998, p. http://www.eurachem.org/guides/pdf/valid.pdf.
[20] Guide to Quality in Analytical Chemistry, CITAC/Eurachem Guide, 2002, p. http://www.eurachem.org/guides/pdf/CITAC%20EURACHEM%20GUIDE.pdf.
[21] S.E. Reichenbach, P.W. Carr, D.R. Stoll, Q. Tao, J. Chromatogr., A 1216 (2009) 3458.
[22] A. Burdack-Freitag, P. Schieberle, J. Agric. Food Chem. 58 (2010) 6351.
[23] W. Silberzahn, R. Tressl, Prog. Flavour Precursor Stud. Proc. Int. Conf. (1993) 147.
[24] Y. Wang, J. O'Reilly, Y. Chen, J. Pawliszyn, J. Chromatogr., A 1072 (2005) 13.
[25] L. Setkova, S. Risticevic, C.M. Linton, G. Ouyang, L.M. Bragg, J. Pawliszyn, Anal. Chim. Acta 581 (2007) 221.
[26] ISO 3534-1, Statistics - Vocabulary and Symbols, International Organization for Standardization, 1998.
[27] C. Alasalvar, F. Shahidi, K.R. Cadwallader, J. Agric. Food Chem. 51 (2003) 5067.
[28] C. Cerny, W. Grosch, Z. Lebensm.-Unters. Forsch. 198 (1994) 210.
[29] H.-D. Belitz, W. Grosch, P. Schieberle, Food Chemistry, Springer, Berlin Heidelberg, 2004.
[30] C. Cordero, C. Bicchi, D. Joulain, P. Rubiolo, J. Chromatogr., A 1150 (2007) 37.

Figure legends

Figure 1: Work flow adopted for the *Comprehensive Template Matching* fingerprinting of hazelnut samples under study.

Figure 2: Effect of user-defined retention time window and MS match factor thresholds on consistent peak matching. 2a) Only a retention time window constraint was defined: 2 of 8 overlaid template peaks (thin circles) were correctly matched and 6 were erroneously matched; 2b) The retention time window and a MS match factor threshold (600) were defined: all analyte peaks were correctly matched; 2c) Restrictive rules have been applied (retention time window and MS match factor threshold of 774): only 3 of 8 peaks were correctly matched.

Figure 3: Histogram reporting the dispersion of signal-to-noise ratios (S/N) of all detected peaks of six replicate sample chromatograms of Standard Roasted Romana hazelnuts acquired over the three weeks of validation.

Figure 4: a) Signal-to-noise-ratio (S/N) of a low abundant peak measured on nine days in three weeks. A peak can be consistently matched across the nine 2D chromatograms, if the peak is always detected, which means that the peak's S/N is always above the user-defined peak detection limit (here S/N 10 indicated by the lower line). Such peaks are recognized using template thresholds for Sample Template development (indicated by upper line at S/N 30) . Asterisks label peaks that would be detected and matched, if the peak detection limit would be set to S/N 30 (inconsistent peak matching indicating false marker compounds); b) S/N of the internal standard α-thujone was used for calculating the template threshold for each 2D chromatogram.

Figure 5: Linear correlation of 11 potential marker compounds with roasting time (0-7-12-17-23 min, 11x4 data points) from four different hazelnut varieties/origins.

Captions to Tables

Table 1: Target analytes adopted for method performance evaluation on Standard Roasted and Self-roasted Romana samples over a period of three weeks*.

Table 2: Results of the pair-wise comparison of replicate analyses by *Comprehensive Template Matching*. Replicates were acquired during the three-week long validation period at different days in the week (1-3).

Table 3: Primary output of the non-targeted analysis of differently roasted hazelnuts (7-12-17-23 minutes) using *Comprehensive Template Matching* on Gentile sample subset.

Supplementary files

Table S1: detailed RSD % data (week 1, week 2, week 3) according to table 1.

Table S2: Experimental design giving the number of replicates per sample and total number of analyses runs. Standard and Self-roasted Romana samples as well as blank and alkane runs were acquired for validation purposes.

Table S3: secondary data output from non-targeted analysis. Peak features were identified and markers were confirmed using selective quantifier ions.

Figure S1: PCA-plots of four different hazelnut samples a) Azerbaijan b) Tonda Gentile from Piedmont c) Tonda Romana from Lazio and d) Tonda Giffoni from Campania. The 2D Normalized Volume values of compounds (numbered rectangular) showing increasing concentrations during hazelnut roasting were analyzed. Some compounds show consistently high correlation with the factor “roasting degree” (component 1) despite the variety/origin.

Figure S2: Ten most temperature-sensitive roasting markers (cf. table 3 for peak numbering) of Gentile hazelnuts. Fold changes were calculated relative to 7 min normalized volume value (7 min = 1).

Figure S3: Release profiles of potential roasting markers 2,3-pentanedione and unknown 81. 5-Methyl-(E)-2-hepten-4-one and unknown compound 79 were excluded from marker screening, because releases were inconsistent across all hazelnut samples (cf. 3.2).

Table 1: Target analytes[#] adopted for method performance evaluation on Standard Roasted and Self Roasted Romana samples over a period of three weeks*.

			absolute retention times				normalized response		match factor	
ID	compound	I^T_S	$^1D\ t_r$ (min)	week 1-3 (RSD %)	$^2D\ t_r$ (s)	week 1-3 (RSD %)	peak volume	week 1-3 (RSD %)	MS spectra	week 1-3 (RSD %)
1	2,3-pentanedione[#]	1041	8.6	1.2	1.1	2	0.4	9.1	767	5
2	2,2-dimethyl-3-hexanone	1107	10.5	1	1.8	1.9	0.09	17.2	771	14
3	(E)-3-penten-2-one	1115	10.8	0.9	1.3	2.4	1.29	7.6	765	17
4	pyridine	1169	13	0.9	1.2	3.2	0.17	15.6	781	3
5	5-methyl-(Z)-2-hepten-4-one[#]	1182	13.5	0.8	2.2	1.4	0.13	11.7	848	6
6	3-methyl-3-penten-2-one	1185	13.7	0.8	1.6	1.6	0.14	13.9	771	12
7	1-pentanol	1238	16	0.8	1.2	3.7	0.11	9	839	2
8	2-methylpyrazine	1252	16.6	0.7	1.3	2.9	0.96	7.6	871	2
9	5-methyl-(E)- 2-hepten-4-one[#]	1280	17.9	0.6	2.3	2.1	1.44	12	867	6
10	2,5-dimethylpyrazine[#]	1311	19.3	0.7	1.6	2.8	0.55	8.1	816	13
11	2-ethylpyrazine	1321	19.8	0.6	1.6	3	0.37	10.2	829	2
12	2,3-dimethylpyrazine	1334	20.4	0.6	1.6	3.7	0.08	7.5	813	3
13	1-hydroxy-2-butanone	1362	21.8	0.6	1.2	4	0.09	8.2	861	5
14	2-ethyl-5-methyl-pyrazine	1373	22.3	0.6	1.9	2.9	0.15	11.9	838	2
15	nonanal[#]	1383	22.8	0.5	2.9	1.9	0.28	47.6	779	11
16	2,3,5-trimethylpyrazine[#]	1393	23.3	0.5	1.8	3.2	0.15	7.6	680	3
17	3-ethyl-2,5-dimethyl-pyrazine[#]	1434	25.3	0.5	2.1	2.9	0.11	11.1	804	3
18	furan-2-carbaldehyde	1451	26.1	0.5	1.2	5	1.6	12.5	901	6
19	benzaldehyde[#]	1509	28.9	0.4	1.5	3.9	0.18	17.7	680	10
20	2-phenylacetaldehyde[#]	1628	34.4	0.5	1.6	4.1	0.14	17.4	730	3
21	2-furanmethanol	1649	35.4	0.4	1.1	5.4	0.78	13.6	766	10
22	phenylmethanol	1859	44.5	0.4	1.4	7.4	0.03	12.3	722	5
23	α-thujone[#]	1409	24	0.5	3	2.7	1	0	880	2
24	β-thujone[#]	1428	24.9	0.5	2.9	2.2	0.07	7.5	848	8
	α-thujone raw peak volume							*23.5*		
	β-thujone raw peak volume							*29.7*		
	average of standard roasted Romana			0.6		3.2		12.4	801	6
	average of self roasted Romana			1		2.8		46.2	774	9

[#]: Analytes were identified with authentic standards (linear retention index (I^T_S), ^2D absolute retention time, EI mass spectrum); all other analytes were tentatively identified (linear retention indices and mass spectra identical to data reported in literature).

* Detailed RSD % data on each single week (w1, w2 and w3) are shown in Supplementary Information Table 1.

Table 2: Results of the pair-wise comparison of replicate analyses by *Comprehensive Template Matching*. Replicates were acquired during the three-week long validation period at different days in the week (1-3).

Validation Week	**1**	**1**	**1**	**2**	**2**	**2**	**3**	**3**	**3**	
Validation Day	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**	**RSD %**
Replicate analysis of standard roasted Romana samples										
Number of peaks above peak detection limit[a]	487	395	470	431	321	325	357	342	354	
Template threshold [S/N][b]	54	30	51	31	33	25	30	33	30	29
Number of template peaks[c]	172	168	162	205	150	159	156	155	165	10
Percentage of matched peaks [%][d]	95	91	95	94	92	93	95	94	93	2
Error [%][e]	5	9	5	6	8	7	5	6	7	
Replicate analysis of self roasted Romana samples										
Number of peaks above peak detection limit[a]	478	640	436	579	552	624	605	618	501	
Template threshold [S/N][b]	47	144	30	157	85	93	132	90	98	43
Number of template peaks[c]	68	50	67	64	72	72	55	66	48	15
Percentage of matched peaks [%][d]	92	92	86	95	89	89	90	91	85	3
Error [%][e]	8	8	14	5	11	11	11	9	15	

a: Peak detection limit was fixed at S/N > 10.

b: Template threshold values were calculated relative to S/N of the ISTDs.

c: Number of template peaks, i.e. peaks exceeding the template threshold were loaded into the template after blank sample removal.

d: Percentage of matched peaks is the arithmetic mean of the pair-wise comparison of arbitrarily chosen templates with the other eight 2D chromatograms of the set (n=3).

e: Error % represents fingerprinting accuracy.

Table 3: Primary output of the non-targeted analysis of differently roasted hazelnuts (7-12-17-23 minutes) using *Comprehensive Template Matching* on Gentile sample subset.

no.[a]	1D t_r (min)	2D t_r (s)	fold change[b]				no.[a]	1D t_r (min)	2D t_r (s)	fold change[b]				no.[a]	1D t_r (min)	2D t_r (s)	fold change[b]			
			7	12	17	23				7	12	17	23				7	12	17	23
1	4.2	0.97	1.0	1.2	0.5	0.4	30	11.4	1.47	1.0	1.2	1.4	1.7	59	22.2	1.47	1.0	0.7	0.8	1.2
2	4.7	1.47	1.0	1.3	0.6	0.6	31	11.5	1.94	1.0	1.0	0.6	1.2	60	22.6	1.73	1.0	0.7	0.6	0.8
3	4.8	0.67	1.0	0.4	0.9	1.1	32	11.8	2.44	1.0	1.2	1.1	1.2	61	23.7	1.73	1.0	0.8	0.6	0.7
4	5.1	1.81	1.0	1.4	0.6	0.6	33	12.4	1.56	1.0	1.8	1.1	1.4	62	25.3	1.81	1.0	2.6	2.0	2.0
5	5.2	0.76	1.0	1.2	2.6	2.7	34	12.5	1.81	1.0	0.9	1.4	1.2	63	25.6	1.81	1.0	2.4	2.0	3.5
6	5.2	1.05	1.0	0.7	1.4	2.6	*35	13.1	1.18	1.0	2.3	3.2	5.5	64	25.7	0.88	1.0	2.9	2.3	2.8
7	5.4	1.64	1.0	1.4	0.7	0.6	36	13.6	2.15	1.0	11.4	4.5	8.9	65	25.8	1.60	1.0	1.0	2.3	1.7
*8	5.8	0.97	1.0	5.2	4.3	6.5	37	13.8	1.60	1.0	2.3	1.4	2.3	*66	26.0	1.22	1.0	6.5	8.4	11.6
9	6.3	1.01	1.0	2.1	2.8	2.0	38	13.8	2.23	1.0	1.9	2.7	2.6	67	26.8	1.81	1.0	1.5	1.3	2.6
10	6.6	1.56	1.0	1.7	1.2	1.6	39	14.0	1.85	1.0	1.5	2.0	2.2	68	27.1	1.98	1.0	1.3	1.4	1.2
11	6.8	1.14	1.0	0.5	0.4	0.5	40	14.1	1.14	1.0	1.2	0.5	1.2	69	27.4	1.94	1.0	0.3	0.7	0.9
12	7.4	1.47	1.0	1.3	1.6	2.4	41	14.4	2.15	1.0	6.8	2.6	4.3	70	27.6	1.73	1.0	1.4	0.9	1.1
13	7.6	1.05	1.0	1.5	3.7	3.8	42	14.5	1.64	1.0	0.6	0.2	0.2	71	27.8	1.98	1.0	2.8	3.4	2.6
14	7.6	1.39	1.0	1.7	1.5	2.6	43	15.8	1.94	1.0	1.8	1.7	2.3	72	29.2	1.64	1.0	1.0	1.0	0.6
15	7.7	0.93	1.0	0.6	0.5	0.3	44	15.9	1.09	1.0	1.8	1.4	1.4	*73	29.5	1.89	1.0	5.6	3.9	3.9
16	7.9	1.39	1.0	1.3	1.8	1.1	45	16.0	1.22	1.0	1.7	1.8	3.0	74	32.1	1.98	1.0	1.4	1.4	3.2
17	8.2	0.63	1.0	1.8	0.8	0.8	46	16.2	2.44	1.0	1.4	0.7	1.8	75	32.8	1.68	1.0	1.8	2.7	2.9
18	8.6	1.43	1.0	0.3	0.2	0.3	47	16.3	1.85	1.0	2.0	1.7	2.9	76	33.4	2.06	1.0	1.8	0.8	2.1
*19	8.7	1.05	1.0	2.3	3.3	3.5	48	17.0	1.98	1.0	1.1	1.5	1.4	*77	34.2	1.64	1.0	22.8	27.1	38.5
20	9.0	1.85	1.0	0.7	0.7	0.9	49	17.4	1.52	1.0	0.9	0.5	2.0	78	34.9	1.39	1.0	3.5	2.9	2.5
21	9.0	1.35	1.0	0.9	0.9	0.9	*50	17.6	1.14	1.0	2.7	2.9	3.9	79	37.1	1.52	1.0	4.9	1.4	4.0
22	9.8	1.26	1.0	2.0	2.0	3.8	51	17.6	1.64	1.0	0.8	0.8	1.1	80	37.2	1.22	1.0	1.4	2.1	3.9
*23	10.4	1.64	1.0	2.1	3.2	3.9	52	17.9	2.31	1.0	4.0	2.9	3.9	*81	38.0	1.26	1.0	2.0	0.9	6.2
24	10.6	1.73	1.0	1.2	1.7	1.6	53	18.4	2.02	1.0	2.2	2.4	2.2	82	44.6	1.09	1.0	0.6	0.6	0.8
25	10.7	1.05	1.0	0.3	0.3	0.5	*54	19.8	1.60	1.0	6.8	16.2	26.7	83	45.8	1.56	1.0	2.3	1.7	2.5
26	10.9	2.02	1.0	0.8	0.5	0.8	55	20.2	1.26	1.0	0.7	1.0	1.3							
27	11.0	1.26	1.0	1.5	1.5	1.6	56	20.9	1.43	1.0	1.4	2.1	1.5							
28	11.0	1.64	1.0	1.9	1.7	1.7	57	21.0	1.26	1.0	0.8	0.9	1.3							
*29	11.3	1.18	1.0	1.5	3.5	10.7	58	21.2	1.77	1.0	2.2	1.7	3.5							

Table 3 continued:

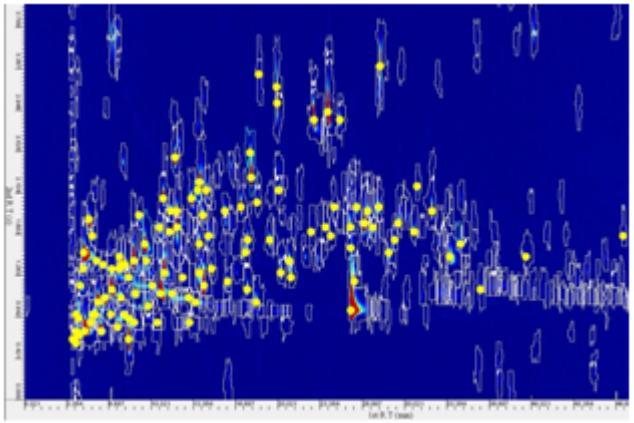
a: Features are reported together with the template peak numbering. A feature characterizes an analyte peak with ^{1}D and ^{2}D retention times and EI mass spectra.
b: Relative increase of Normalized Peak Volumes of 7 minutes roasted sample versus 12, 17 and 23 minutes roasted samples.
* An asterisk indicates analytes which were identified as roasting markers.

Figure 1

GC Image™
Background substraction
Blob detection above certain S/N ratio

Consensus template development [3,6],
every template peak gets an unique number

-> remove template peaks of blank samples
-> remove multiple entries



GC Project™
Consensus template is consecutivley matched with all chromatograms,
template peak number is assigned to positive matched analyte

-> apply correct 1st, 2nd rentention time window for peak pattern matching
-> apply correct MS match factor threshold
-> apply correct template threshold

Peak tables are automatically aligned according to the unique numbers,
response data of same analytes are sorted in one row

Figure 2

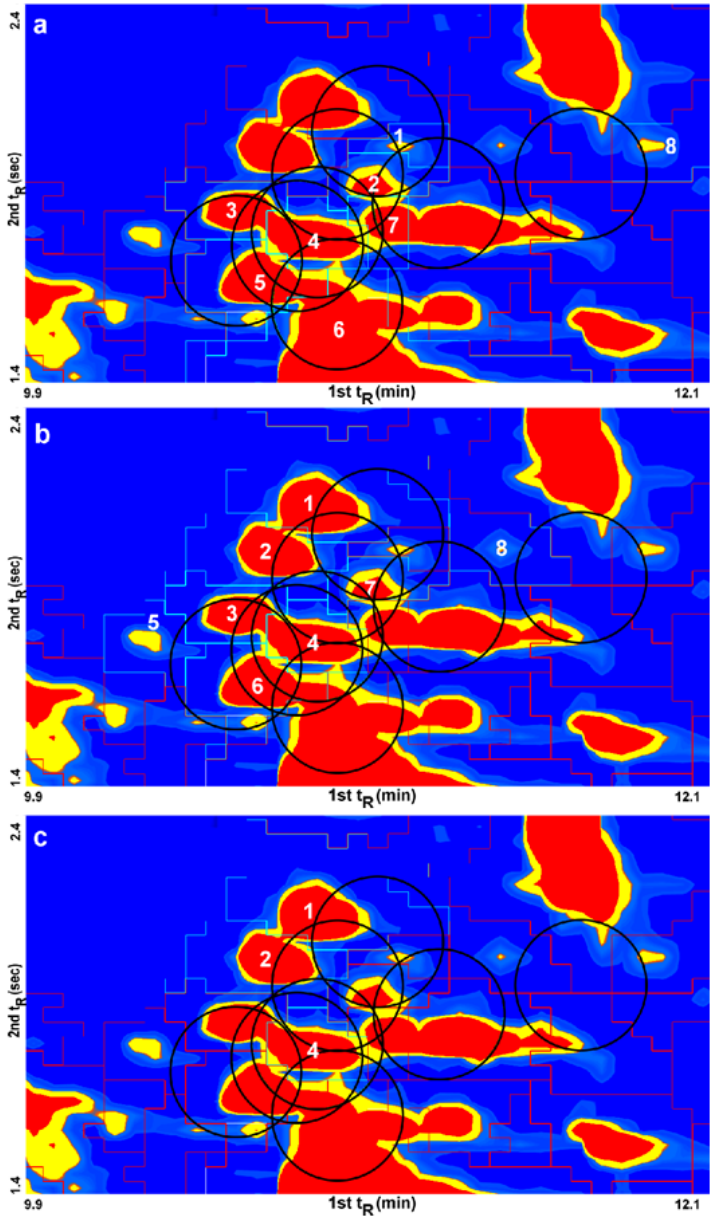


Figure 3

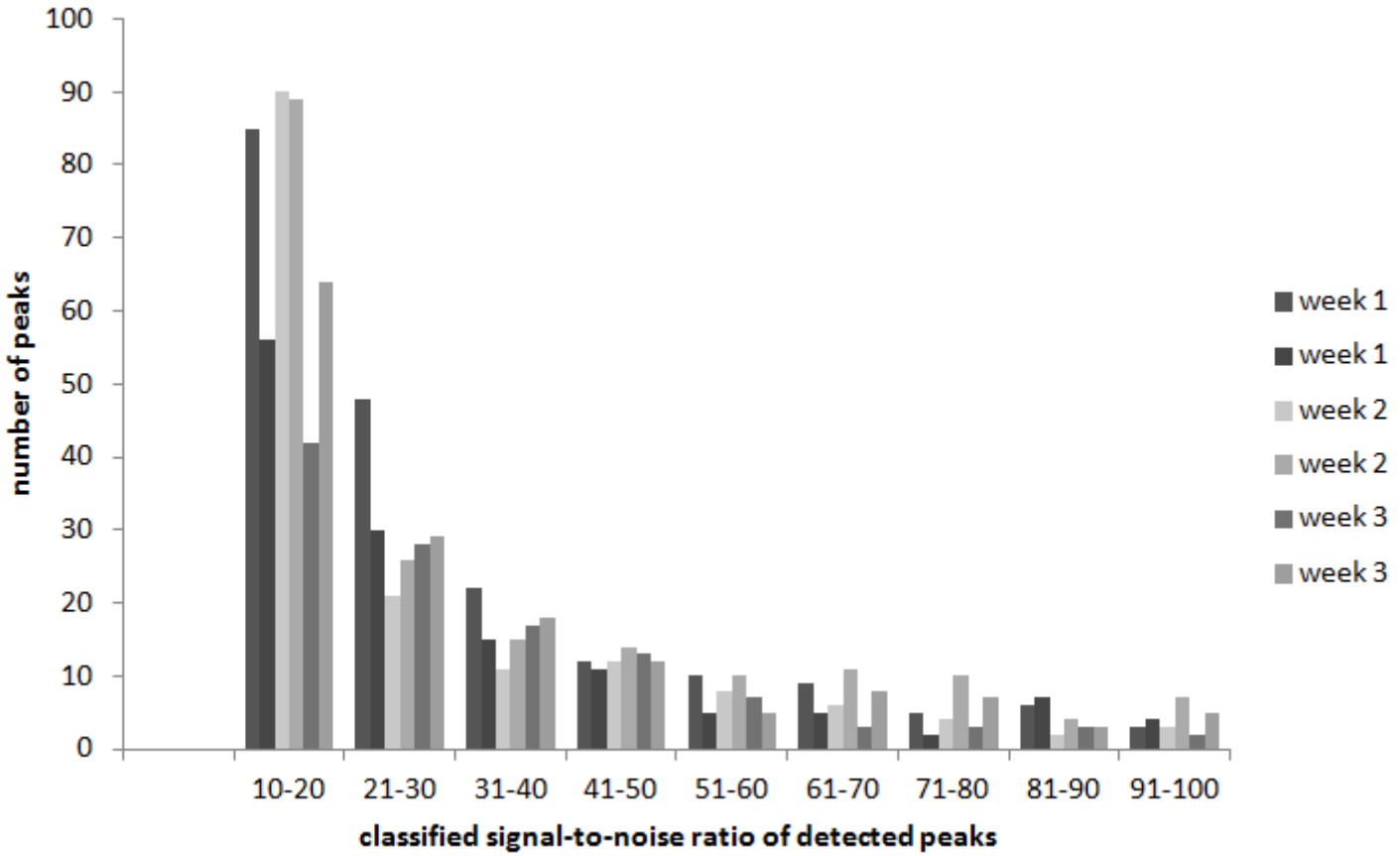
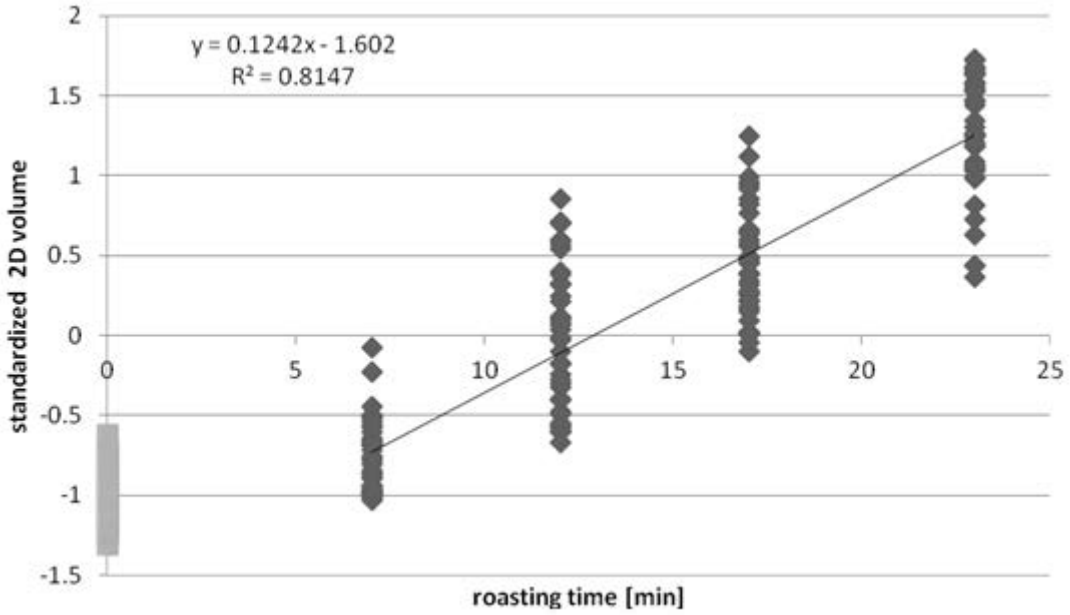
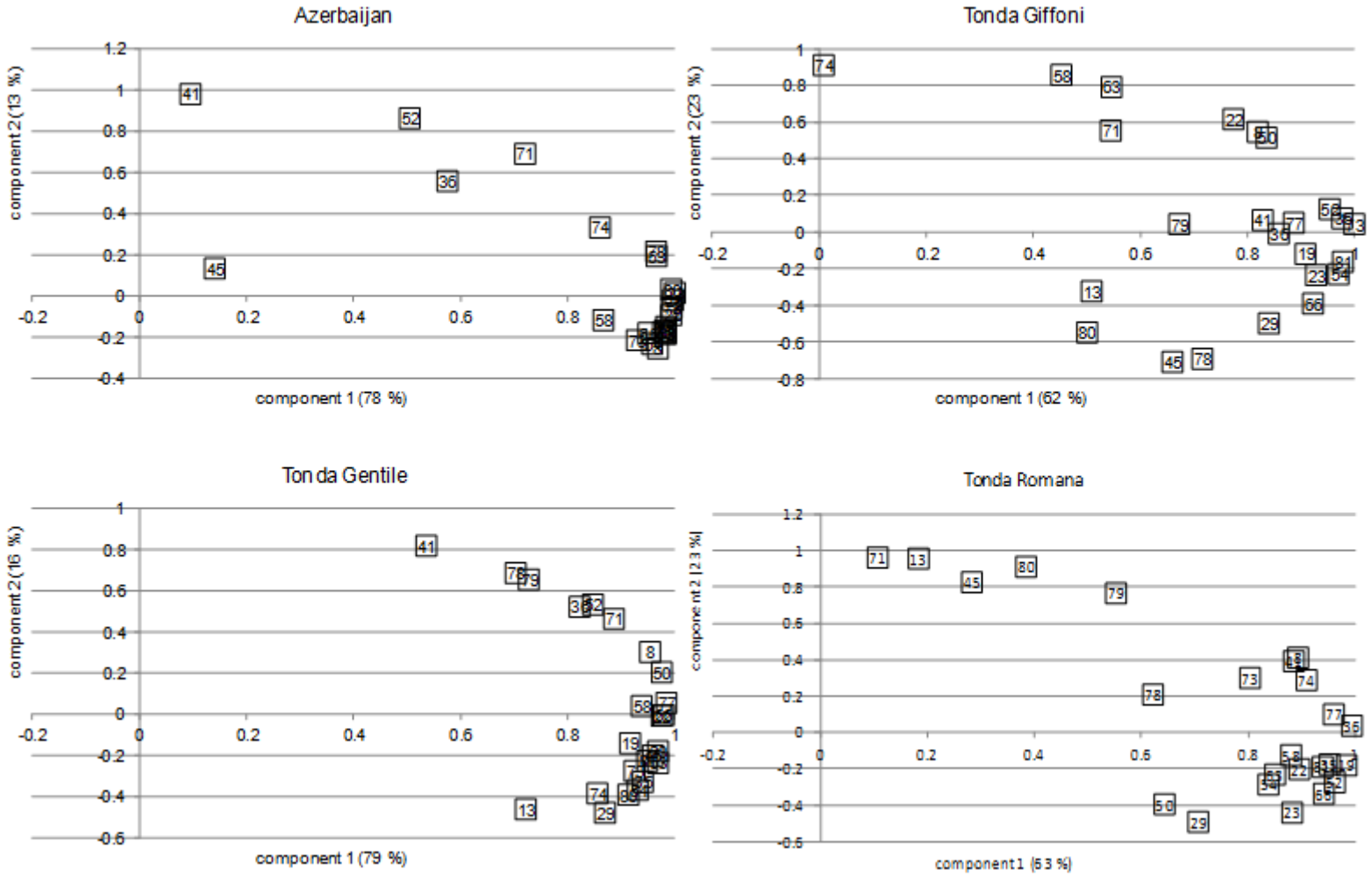

100
90
80
70
60
50
40
30
20
10
0
number of peaks
10-20
21-30
31-40
41-50
51-60
61-70
71-80
81-90
91-100
classified signal-to-noise ratio of detected peaks
week 1
week 1
week 2
week 2
week 3
week 3

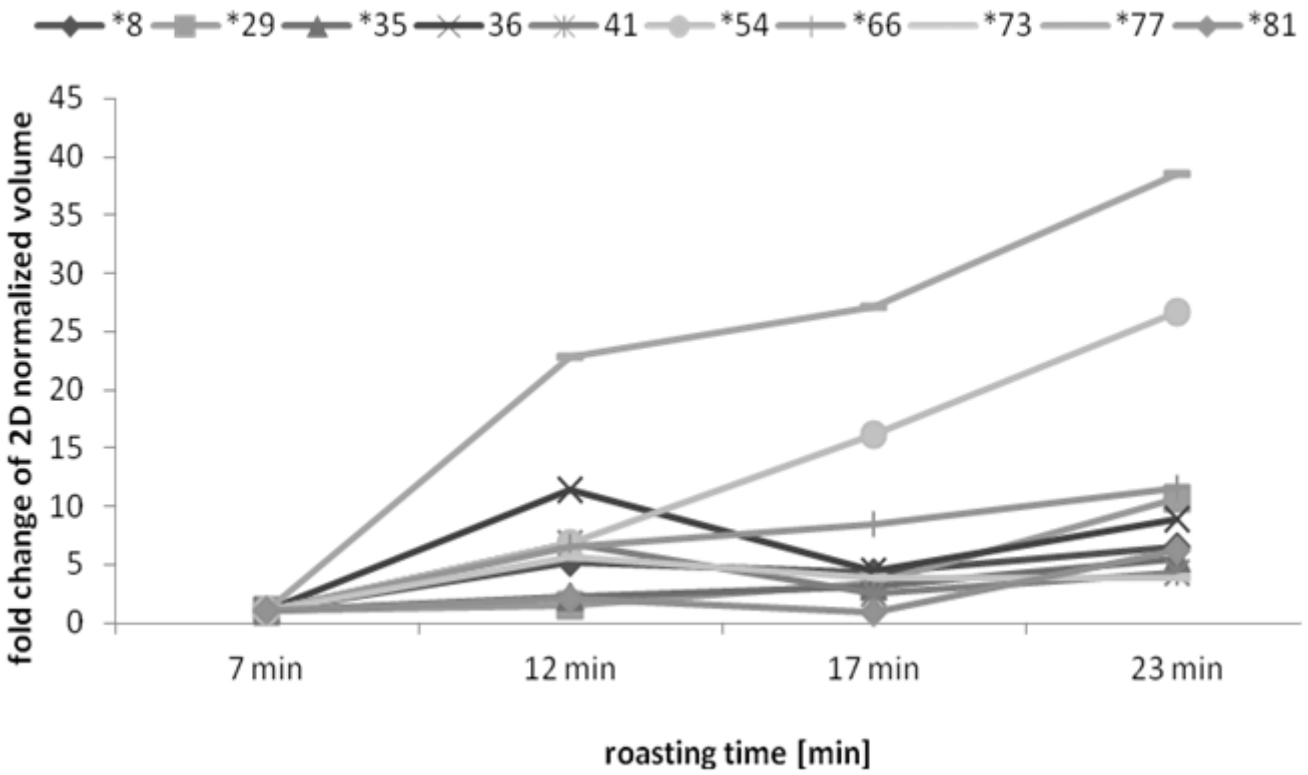
Figure 4


y = 0.1242x - 1.602
R² = 0.8147
standardized 2D volume
roasting time [min]
2
1.5
1
0.5
0
-0.5
-1
-1.5
0
5
10
15
20
25

Supplementary Figure 1



Supplementary Figure 2



Supplementary Figure 3