Impact of microclimatic conditions on dairy cows' milk production in an Automatic Milking System: evaluating trends and predictive modeling

M. Zanchi[1], C.A.M La Porta[1], C. Forte[2], S. Zapperi[3] and L. Ozella[2]
[1]*Departement of Environmental Sciences and Policies, University of Milan, Via Celoria 10, 20133 Milan, Italy*
[2]*Department of Veterinary Sciences, University of Turin, Largo Braccini 2, 10095 Grugliasco (TO), Italy*
[3]*Department of Physics, University of Milan, Via Celoria 16, 20133 Milan, Italy*
laura.ozella@unito.it

Abstract

Heat stress deriving from climate change represents a serious challenge to the dairy industry. Recent advancements in Artificial Intelligence have introduced a variety of tools capable of forecasting future conditions by utilizing past information as inputs, potentially offering valuable resources to address heat stress. This study aims to evaluate and predict the impact of microclimatic conditions on milk production within an Automatic Milking System (AMS). From 2016 to 2023 data on milk production and quality were collected from AMSs on a commercial dairy farm, while meteorological information was obtained from the closest ARPA (Agenzia Regionale per la Protezione Ambientale) station. Temperature and relative humidity data were combined to derive the Temperature Humidity Index (THI). Our analysis revealed a significant negative correlation between the protein content and THI over time, reaching a peak correlation of -0.66 with a lag of five days. This confirms a strong connection between milk quality and the surrounding temperature and humidity conditions. Consequently, understanding both current and historical microclimate conditions could facilitate predictions of future trends in milk-related data. Therefore, we employed a recent Neural Network (NN) model named TSMixer to forecast the protein, fat, milk production and milk temperature for the upcoming two months. This model utilizes input data from the preceding three months. TSMixer was trained and validated using data from 2016 to 2021, and subsequently tested with data from 2022 and 2023. The test outcomes demonstrate the NN's ability to accurately predict linear trends for the upcoming two months across protein, fat and milk production with R^2 values of 0.83, 0.81, 0.80.

Keywords: deep learning, dairy cows, time series forecasting, milking robots, heat stress.

Introduction

Climate change presents significant challenges to the dairy industry by causing heat stress in livestock, a result of various environmental factors including temperature, relative humidity, and airflow (Joksimovic et al, 2011). This combination of factors can negatively affect animal well-being as well as milk production and quality (Tao et al, 2020), thereby having a profound impact on the economic outlook of the global dairy sector (Wheelock et al, 2010). The most acknowledged effect of heat stress on dairy cows is the reduction in milk production and dry matter consumption, which in turn, causes considerable financial detriment to dairy farmers (Bouraoui et al 2002; Gao et al, 2017).

In response to the challenges posed by climate change-induced heat stress on dairy cattle, the utilization of Artificial Intelligence (AI) and Machine Learning (ML) methods has emerged as a promising approach to predict milk production in relation to environmental factors and to identify solutions that dairy farmers can implement to mitigate economic losses. AI and ML methodologies function as potent tools for unravelling the complex and non-linear relationships hidden within extensive datasets, and they have been widely utilized in the field of Precision Livestock Farming (PLF) in recent years (García et al, 2020; Ozella et al, 2023). Through AI-based modelling, the analysis of extensive datasets from robotic dairy operations can be refined, leading to improvements in overall milk productivity and quality. Automatic Milking Systems (AMSs) provide farmers with valuable insights into various milk production parameters, such as production volume, lactation period, and the levels of lactose, fat, and protein. As a result, AI's capacity to accurately forecast milk production enables farmers to develop more effective financial strategies and identify anomalies in production trends, even in relation to heat stress (Bovo et al, 2021; Ji et al, 2022).

In the present study, we investigated how microclimatic variables, as temperature and humidity, affect the milk output of dairy cows by analysing data collected from AMS, including information on each milking session and daily milk yields over a period of six years and seven months. We examined the time-based relationships among milk components such as protein and fat, milk production, and the Temperature Humidity Index (THI). Furthermore, we introduced the application of TSMixer, a novel deep learning technique designed for predicting the future trends of multivariate time series over extended periods. This algorithm was employed to project the future changes in protein levels, fat content, milk production, and milk temperature for the forthcoming two months.

Material and methods

Experimental data

This study focuses on analysing data from Holstein Friesian cows obtained from a commercial farm situated in the north-west region of Italy. The data were collected using four milking robots (Astronaut robot, Lely, The Netherlands) from July 2016 to February 2023. The farm maintains an average herd size of 240 Holstein Friesian cows. In particular, the study focuses on the analysis of lactation period of the cow, the milk production by day (kg), the protein content (average percent of protein in milk), fat content (average percent of fat in milk), rumination time (average minutes of rumination by day) and milk temperature (°C) averaged by day. We have performed a data cleaning procedure, retaining only the continuous sequences of AMS measures at daily frequencies, greater than 200 days. The data sequences have been divided for each cow and for each lactation period.

We have also collected meteorological data from an ARPA (Agenzia Regionale per la Protezione Ambientale) station, situated in Bauducchi, Turin, Italy, which is the closest station to the farm. The daily averages of temperature and relative humidity have been gathered from July 2016 to February 2023. The Temperature Humidity Index (THI) has been inferred from this data, according to the equation proposed by (Segnalini et al, 2011):

$$THI = 0.8T + \frac{RH}{100}(T - 14.4) + 46.4 \quad (1)$$

where T is the temperature (°C) and RH the relative humidity (%).

Correlation analysis

The impact of heat stress on milk quality and production has been widely acknowledged in the literature (Becker et al, 2020). We have performed a cross-correlation analysis to verify the presence of this phenomenon in our data. We have averaged at daily frequencies the milk protein, fat, production, and temperature data collected by the milking robots and compared them with the THI temporal variation. The Pearson's correlation coefficient has been employed to estimate the correlation. In addition, we have tried to estimate the time response of the milk quality variations induced by the heat stress. We have introduced a time lag in the correlation analysis. We have then computed the optima of the correlation values between the milk variables and the THI, varying the time lag.

TSMixer neural network

Time-Series Mixer (TSMixer) is a neural network introduced by Chen et al. (2023). It has been designed to forecast long-horizon evolutions of multivariate time-series and it is based on multi-layer perceptrons (MLPs). TSMixer takes as inputs historical observations composed by multiple features and predicts the evolution of a set of output features for a time horizon chosen by the user. TSMixer is able to understand complex patterns in both the time and feature domains, by sharing time-domain MLPs across all features and feature-domain MLPs across all time steps (Chen et al, 2023). TSMixer architecture is composed by a Mixer layer, which shares the MLPs across the time and features domain, and by the temporal projection layer, which is in charge for adapting the shape of the neural network input to the output shape chosen by the user.

We employed TSMixer to predict the next 60-days evolution milk quantity, milk temperature, protein, and fat content. In particular, the historical inputs data are 8 distinct features identified by the past 90-days data of fat, milk production, milk temperature, number of lactation and lactation days, protein, rumination time, and THI (for each cow and lactation period). The neural network has been trained, validated, and tested on the data collected by the milking robots and by the ARPA station. The training set comprises all the data from 2016 to 2021, the validation set comprises all the data 2021 to 2022 and the test set comprises all the data from 2022 to 2023.

We have optimized TSMixer hyper-parameters by a grid-search procedure. The final architecture comprises 8 mixer layers blocks with fully connected layers 64 neurons large. The TSMixer has been trained and validated with an Adam optimizer with a mean squared error (MSE) as loss function. Then the architecture has been tested on the test set using the mean absolute error (MAE) and the determination coefficient R^2, which have been computed separately for each feature.

To further test TSMixer accuracy, we have conducted a comparison of the test absolute error distribution obtained from the neural network with a baseline model. This baseline model reproduces the yearly averaged pattern of the four predicted variables. For each prediction variable, for each day of the year, an average has been computed among all samples for all the years included in the train and validation sets. These averages have been than fitted with a periodic function, defined by a sum of sine and cosine plus a phase:

$$Baseline = A\sin\left(\frac{2\pi}{365}x + \phi_1\right) + B\cos\left(\frac{2\pi}{365}x + \phi_2\right) + C \quad (2)$$

The final fitted function has been then used as a baseline model.

Results and Discussion

We have conducted a cross-correlation analysis between the time series of daily mean milk production, fat, and protein content and milk temperature with daily mean THI over time, from July 2016 to February 2023. The results of the analysis are represented in Figure 1.

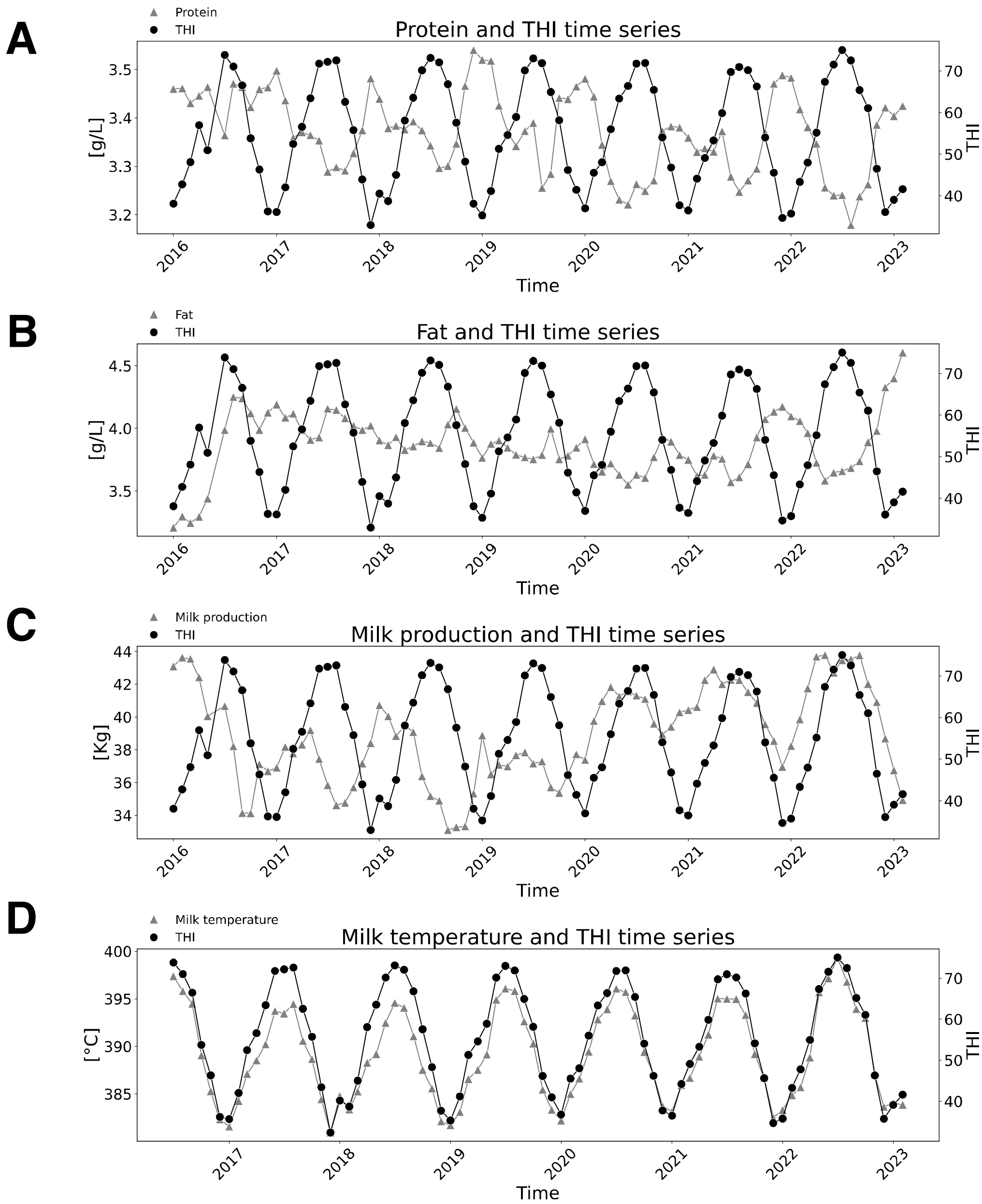

Figure 1: Temporal variation of the average protein, fat, milk production, milk temperature and THI averaged on a monthly base.

The time series of protein presents an evident negative correlation with respect to the THI, suggesting a decline in milk protein content when heat-stress conditions verify. After the introduction of a lag between the two timeseries, we have found a minimum of the correlation coefficient of -0.66 with a lag of 5 days, suggesting a potential time lag between the decline in protein and the heat-stress effect. The timeseries of fat presents an evident negative correlation after the year 2020, with a minimum for the lagged correlation of -0.15 without any lag. The findings regarding protein and fat behaviour are consistent with other studies conducted in Mediterranean climates. An increase of the average THI, during spring and summer, generates a corresponding decrease in the protein and fat contents of milk (Bernabucci et al, 2010; Kadzere et al, 2010). Furthermore, the time lag observed in the correlation between protein and THI may be explained by the fact that the physiological response of cows to an increase in THI is not immediate and may require several days to manifest, as suggested by Collier et al. (2006). The milk production exhibits a different pattern depending on the period considered. A negative correlation is observed in the first three years, from 2016 to 2019, while a positive one is observed from 2021 to 2023. It is widely recognized that heat stress has a detrimental effect on total milk production. Reductions in milk yield due to heat stress are typically observed when the daily average temperature-humidity index exceeds 68 (Tao et al., 2020). To better understand the patterns found in our study, further investigations should be conducted, focusing on the THI levels that predominantly determine the reduction in milk production. The milk-temperature shows a positive correlation with the THI, reaching a maximum correlation of 0.68 without any time lag. Previous studies (e.g., Igono et al., 1988; Toghdory et al, 2022) have found that as the environmental temperature increases, so does the temperature of the milk produced by cows, indicating a direct relationship.

We applied the TSMixer neural network to predict the evolution of milk-related data. The TSMixer model takes as input the past 90 days of data collected from a single cow, including fat, milk production, milk temperature, number of lactations and lactation days, protein, rumination time, and THI. It then processes this information to forecast the next 60 days' evolution of protein, fat, milk production, and milk temperature. We have tested the accuracy of TSMixer predictions by analysing for each output feature the MAE and R^2. The distribution of the MAE is represented in Figure 2.

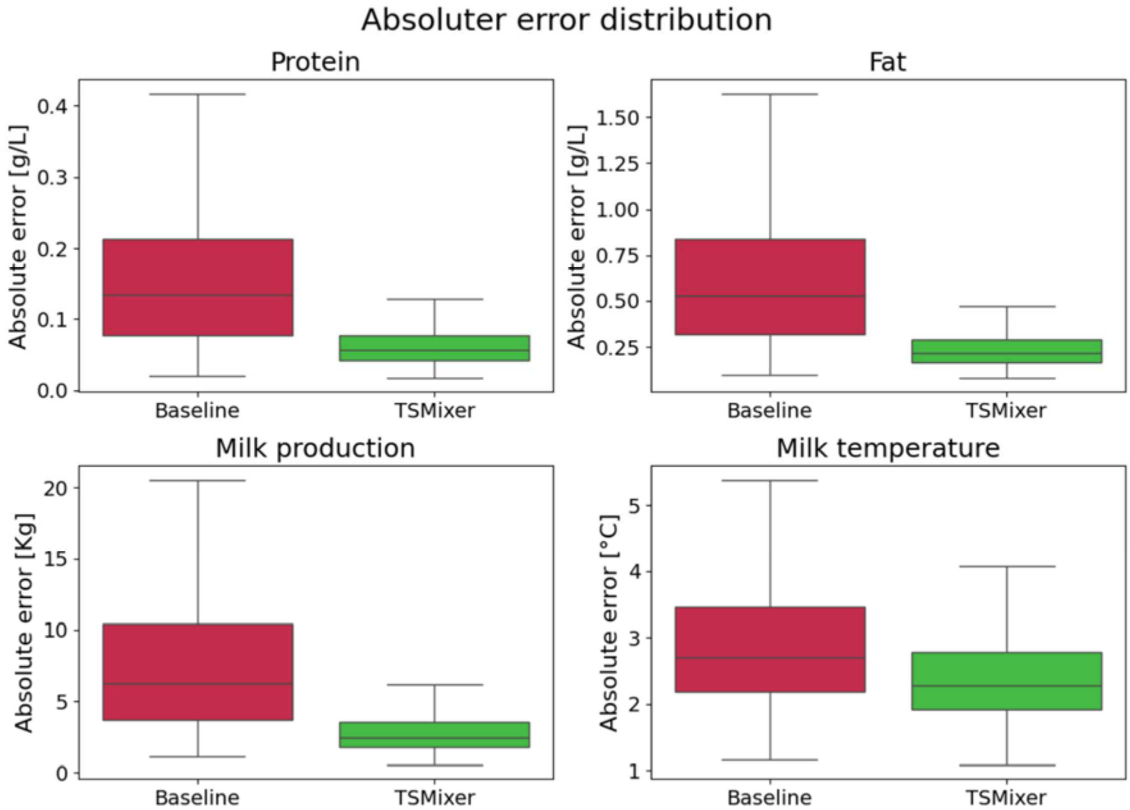

Figure 2: Absolute error distributions obtained for each predicted feature on the test set.

The TSMixer has achieved respectively R^2 values for protein, fat, milk production, and milk temperature of 0.83, 0.81, 0.80 and 0.63. The first three values highlight a good agreement between the true values and the predictions. While the TSMixer seems to perform poorly in milk temperature predictions.
To support the application of the TSMixer model for predicting the evolution of milk-related data, we compared its Mean Absolute Error (MAE) against that of a baseline model, which accounts for the seasonal variations caused by annual temperature fluctuations. We calculated the MAE for the baseline model and illustrated these findings in Figure 2. Upon comparing the baseline model's boxplots with those of the TSMixer, it's evident that the TSMixer model surpasses the baseline in accuracy, demonstrating lower mean absolute errors. To further assess this result, we have computed the Wilcoxon statistical test for each feature to compare the errors distribution obtained from the baseline and the TSMixer model. For all features the TSMixer distribution is statistically lower than the one obtained by the baseline with a Pvalue less than 0.01. This superiority indicates that the TSMixer model's efficacy extends beyond merely replicating predictable seasonal patterns, thereby offering additional insights. The only instance where the TSMixer does not excel over the baseline involves milk temperature predictions, where both models' boxplots are similar. This similarity likely arises from the strong correlation between milk temperature and the THI observed in Figure 1D, suggesting that the predictions in this domain are predominantly influenced by annual temperature shifts.

Conclusions

Our research presents an innovative method designed to deepen the understanding of both current and historical microclimate conditions, thereby improving the accuracy of predictions for future milk production trends based on data from AMS. By conducting a cross-correlation analysis between the time series data of daily average milk yield, fat, and protein concentrations, and the daily average THI, we observed a notable trend. Specifically, an increase in average THI is associated with a decline in milk fat and protein levels, exhibiting a delay of 5 days for protein adjustments. This indicates that heat stress affecting dairy cows has a tangible impact on milk composition. Furthermore, we propose employing TSMixer, a cutting-edge deep learning algorithm tailored for extended forecasts of multivariate time series, to predict the future dynamics of protein and fat concentrations, milk production, and milk temperature. TSMixer has shown superior performance compared to a baseline model, demonstrating its effectiveness despite its streamlined architecture. This model can be seamlessly integrated into automated systems for immediate forecasting, offering a valuable resource for informed decision-making. The deployment of TSMixer provides farmers with two principal benefits. Firstly, it grants them foresight into potential future trends, enabling proactive measures in herd management and resource distribution. Secondly, the implementation of TSMixer in real-time automated systems streamlines the decision-making process, allowing for swift action in response to evolving circumstances.

Acknowledgements

The work presented in this paper was made possible in part through the financial support provided by DNDG Srl through the doctoral scholarship of Marco Zanchi.

References

Becker, C. A., Collier, R. J., & Stone, A. E. 2020. Invited review: Physiological and behavioral effects of heat stress in dairy cows. *Journal of dairy science*, *103*(8), 6751-6770.

Bernabucci, U., Basiricò, L., Morera, P., Dipasquale, D., Vitali, A., Cappelli, F. P., & Calamari, L. U. I. G. I. 2015. Effect of summer season on milk protein fractions in Holstein cows. *Journal of dairy science*, *98*(3), 1815-1827.

Bouraoui, R., Lahmar, M., Majdoub, A., & Belyea, R. 2002. The relationship of temperature-humidity index with milk production of dairy cows in a Mediterranean climate. *Animal Research*, *51*(6), 479-491.

Bovo, M., Agrusti, M., Benni, S., Torreggiani, D., & Tassinari, P. 2021. Random forest modelling of milk yield of dairy cows under heat stress conditions. *Animals*, *11*(5), 1305.

Chen, S. A., Li, C. L., Yoder, N., Arik, S. O., & Pfister, T. 2023. Tsmixer: An all-mlp architecture for time series forecasting. *arXiv preprint arXiv:2303.06053*.

Collier, R. J., Dahl, G. E., & VanBaale, M. J. 2006. Major advances associated with environmental effects on dairy cattle. *Journal of dairy science*, *89*(4), 1244-1253.

Gao, S. T., Guo, J., Quan, S. Y., Nan, X. M., Fernandez, M. S., Baumgard, L. H., & Bu, D. P. 2017. The effects of heat stress on protein metabolism in lactating Holstein cows. *Journal of dairy science*, *100*(6), 5040-5049.

García, R., Aguilar, J., Toro, M., Pinto, A., & Rodríguez, P. 2020. A systematic literature review on the use of machine learning in precision livestock farming. *Computers and Electronics in Agriculture*, 179, 105826.

Igono, M. O., Johnson, H. D., Steevens, B. J., Hainen, W. A., & Shanklin, M. D. 1988. Effect of season on milk temperature, milk growth hormone, prolactin, and somatic cell counts of lactating cattle. *International Journal of Biometeorology*, 32, 194-200.

Ji, B., Banhazi, T., Phillips, C. J., Wang, C., & Li, B. 2022. A machine learning framework to predict the next month's daily milk yield, milk composition and milking frequency for cows in a robotic dairy farm. *biosystems engineering*, *216*, 186-197.

Joksimović-Todorović, M., Davidović, V., Hristov, S., & Stanković, B. 2011. Effect of heat stress on milk production in dairy cows. *Biotechnology in Animal Husbandry*, *27*(3), 1017-1023.

Kadzere, C. T., Murphy, M. R., Silanikove, N., & Maltz, E. 2002. Heat stress in lactating dairy cows: a review. *Livestock production science*, *77*(1), 59-91.

Ozella, L., Brotto Rebuli, K., Forte, C., & Giacobini, M. 2023. A Literature Review of Modeling Approaches Applied to Data Collected in Automatic Milking Systems. *Animals*, *13*(12), 1916.

Segnalini, M., Nardone, A., Bernabucci, U., Vitali, A., Ronchi, B., & Lacetera, N. 2011. Dynamics of the temperature-humidity index in the Mediterranean basin. *International journal of biometeorology*, *55*, 253-263.

Tao, S., Rivas, R. M. O., Marins, T. N., Chen, Y. C., Gao, J., & Bernard, J. K. 2020. Impact of heat stress on lactational performance of dairy cows. *Theriogenology*, *150*, 437-444.

Toghdory, A., Ghoorchi, T., Asadi, M., Bokharaeian, M., Najafi, M., & Ghassemi Nejad, J. 2022. Effects of environmental temperature and humidity on milk composition, microbial load, and somatic cells in milk of Holstein dairy cows in the northeast regions of Iran. *Animals,* 12(18), 2484.

Wheelock, J. B., Rhoads, R. P., VanBaale, M. J., Sanders, S. R., & Baumgard, L. H. 2010. Effects of heat stress on energetic metabolism in lactating Holstein cows. *Journal of dairy science*, *93*(2), 644-655.