Assessment of productive anomalies in dairy cows

M. Agrusti[1], S. Foroushani[2], M. Ceccarelli[1], M. Bovo[1], D. Torreggiani[1], P. Tassinari[1], T. Amon[2], S. Benni[1]

[1]*Department of Agricultural and Food Science, Alma Mater Studiorum Università di Bologna, Viale Fanin 48, 40127 Bologna, Italy*

[2]*Department of Engineering for Livestock Management, Leibniz Institute for Agricultural Engineering and Bioeconomy, Max-Eyth-Allee 100, Postdam, Germany*

stefano.benni@unibo.it

Abstract

In the panorama of precision livestock farming animal welfare, which the quality of products inevitably depends on, is increasingly important. Nowadays, automatic milking systems allow a more detailed monitoring of individual animals and the customized modelling of the productivity trend of each cow, as well as of the herd. It is already known that a warm, humid environment is the main cause of heat stress for dairy cows, and this is becoming even more serious due to climate change. Data from environmental sensors in the barns together with productivity and activity data enable the study and assessment of production loss due to heat stress. In this work, a new method for identifying production anomalies by modelling the lactation curve is presented. The model allows us, on the one hand, to study the residuals (difference between the observed data and the corresponding prediction) and, on the other hand, to examine the production deficit at the end of the lactation cycle. Furthermore, the use of a machine learning model applied to the data obtained from the first analysis shows it is possible to predict the milk yield loss due to heat stress. The training of the model on several animals in similar conditions (e.g. lactation, age) can be a valuable support for the farmer to predict the potential milk yield losses of the herd and to introduce the necessary preventive or mitigative measures.

Keywords: Precision livestock farming; Data analysis; Agricultural engineering; Animal welfare; Numerical model

Introduction

Heat stress is one of the most critical issues jeopardizing animal welfare and productivity in dairy farms, with consequences for both milk quantity and quality, and for the efficiency of the use of natural resources (Benni et al., 2020), as well as and for the energy required for milk production (Strpić et al., 2020). In this regard, the global trend of higher average temperatures coupled with more frequent temperature peaks reduces the lactation efficiency and thus indirectly increases the negative environmental impact of dairy cattle farming (Heinicke et al., 2019). Moreover, prolonged stresses may harm the cow's health, and the consequent treatment of animals is a burden for the overall economic balance of the farm and, whereas it includes the use of antibiotics, this entails the well-known drawbacks on the environment and on human health.

In the last decades, many devices have been introduced in livestock farms to monitor and control environmental conditions, animal behavior and production parameters (Bonora et al., 2018). Available data types usually include on-farm sensors, providing

detailed milk composition (monthly on a cow level, bi-weekly on a herd level), other performance data (e.g., fertility), weather and environmental data (Mbuthia et al., 2022). Additionally, milk quality reports at dairies and labs are often logged for long-term analysis. Nevertheless, such data processing practices do not allow a daily monitoring of the trends so the farmer can take prompt action (Bovo et al., 2021). Moreover, the information from different sources are almost never crossed among different systems and different levels. A proper numerical modelling for early diagnosis and the identification of optimal prevention strategies requires good data management, in terms of acquisition, processing and harmonization, particularly in fields where very heterogeneous data may be collected (Bovo et al., 2020), and agreed protocols have not been yet standardized. Essentially, tools to effectively interpret the already available information are lacking.

In this study, a new method for identifying production anomalies by modelling the lactation cycle is presented. The model allows both to study the residuals and to examine the production deficit at the end of the cycle. In particular, the term "residual" stands for the difference between the observed data and the corresponding prediction. Furthermore, the use of a machine learning model applied to the data obtained from the first analysis has shown it is possible to predict the milk yield loss due to heat stress. The training of the model on several animals in similar conditions (e.g. lactation, age) can be a plausible support for the farmer in order to predict the potential milk yield losses of the heard and to introduce the necessary preventive or mitigative measures.

Material and methods

The numerical models

The study was conducted using data collected in an experimental farm in Groß Kreutz, Germany. The barn is equipped with a Lely™ "Astronaut" automatic milking system (AMS). More information about the farm is available in earlier publications, e.g. (Hempel et al., 2018)in situations where monitoring in the direct vicinity of the animals is not possible, we collected long-term data in two naturally ventilated dairy barns in Germany between March 2015 and April 2016 (horizontal and vertical profiles with 10 to 5 min temporal resolution.

The dataset covers the period 2015-2020 and entails:

— time series of milk yields of 189 cows, from the AMS;
— time series of temperature and relative humidity from multiple sensors inside the barn.

The packages "*matplotlib*" (Hunter, 2007) and "*seaborn*" (Waskom, 2021) were used to display results and intermediate plots, while other Python (Van Rossum & Drake, 2009) libraries like "*lmfit*", "*numpy*" and "*pandas*" proved to be useful for handling data. The Wood model (Wood, 1967) was used to obtain a curve representing the "expected" yield of each animal during the lactation cycle. As proved by the Random Sample Consensus (RANSAC) approach (Fischler & Bolles, 1981)Random Sample Consensus (RANSAC, the sampling operation followed by a multiple fit, can lead to a solid result and is capable of interpreting/ smoothing data containing a significant percentage of gross error.

Description of the pipeline

The datasets related to the AMS and to the microclimatic data were first read using the *pandas* library and subsequently joined in order to obtain a dataset containing all the information needed for the analysis.

$$THI = 0.8T + RH(T - 14.4) + 46.4$$

The Temperature-Humidity Index (THI) was calculated in the following way (National Research Council, 1971):

For each animal and for each lactation, it is possible to fit the Wood model to obtain the parameters *a, b, c*:

$$MY(DIM) = a\mathrm{DIM}^{b}e^{-c\mathrm{DIM}}$$

Where: MY is the daily milk yield [kg/d] and DIM is days in milk.

A filter was applied to select only the data where the daily THI did not exceed a threshold of 65, used as a predicator of potential heat stress.

A more robust statistics can be obtained by randomly sampling the original amount of data and producing a different Wood fit for each sample.

This way, a collection of Wood models is obtained:

$$\mathrm{MY}_k(DIM) = a_k \cdot \mathrm{DIM}^{b_k} \cdot e^{-c_k \cdot DIM}$$

where are the parameters of the k^{th} curve.

The sampling and fitting process can be repeated *N* times, selecting each time a fixed fraction *f* of the original data. Here, we used *N*=500 and *f*=1/10. The obtained family of curves and the corresponding parameters can then be used to define a representative median curve:

$$\mathrm{MY}_{\mathrm{median}}(\mathrm{DIM}) = A \cdot \mathrm{DIM}^{B} \cdot e^{-C \cdot DIM}$$

where:

$$\begin{cases} A = median(a_k) \\ B = median(b_k) \\ C = median(c_k) \end{cases}$$

The median values of the parameters have been assumed instead of the mean values since they are not affected by outlier values. In some cases, unacceptable curves are obtained, e.g., entailing negative or infinite values or unrealistic trends. For this reason, it has become necessary to perform a selection of only the meaningful curves. Then, residuals were calculated as the difference between the actual milk yield data and values of the median curve. The beam of such curves is then filtered selecting only curves with an initial positive trend. The dispersion of the different values obtained in correspondence of the different curves can be used to define a criterion for the detection of anomaly values, for instance by selecting a proper multiple value of the standard deviation value of defining a confidence interval. This because the values more distant from the median curve would be considered in the method as anomaly points.

Production anomalies

The method proposed here allows the introduction of a lactation model that is robust with respect to statistical fluctuations and automatically creates an acceptability range linked to the dispersion of the curves belonging to the beam.

The standard deviation and its multiples can be used to find a threshold for the residuals in order to determine whether any given value is an anomaly. All points out of the beam were considered "anomalies", Note that positive residuals were also considered. While positive residuals cannot be attributed to heat stress or other adverse effects, they can be informative and could serve as indicators of change in the physiological condition of a cow.

Results and Discussion

The use of a multiple fit on different partial datasets obtained after the sampling operation, led to 500 Wood curves for each lactation. In Figure 1, the curves deemed physically acceptable are shown for a lactation cycle of one sample animal.

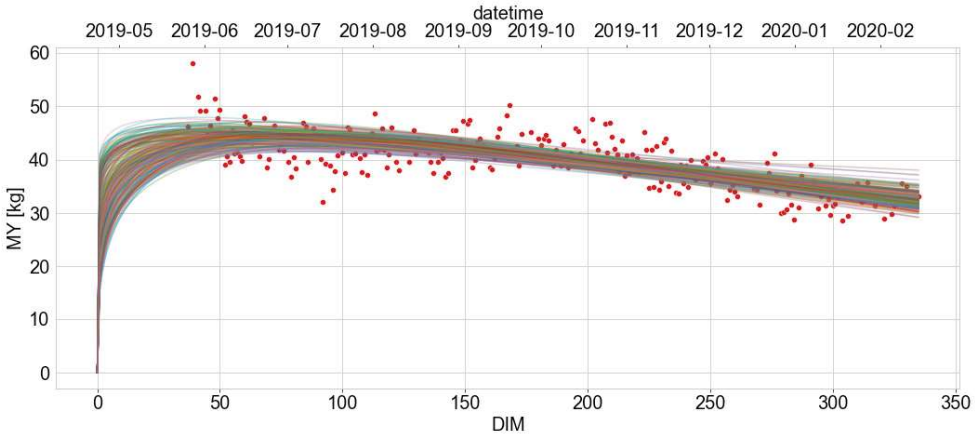

Figure 1: Wood curves with zootechnical significance for one lactation of a cow considered in the study.

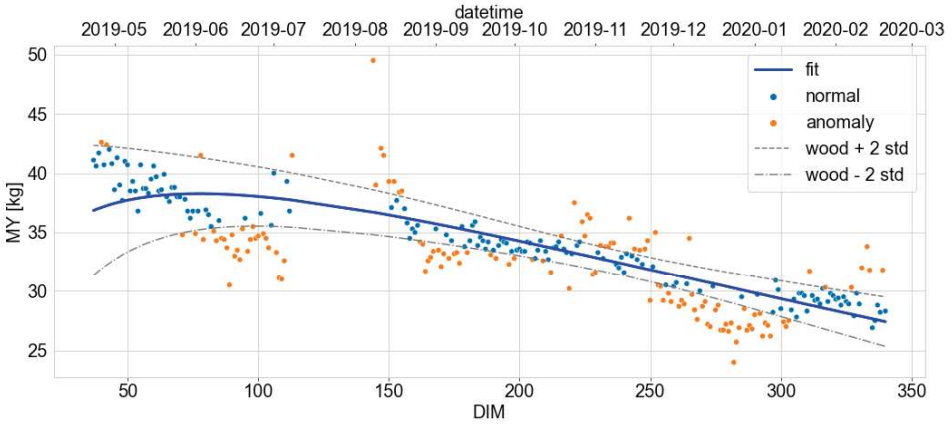

Figure 2: Anomaly detection in the plane DIM Vs MY.

Figure 2 shows, in solid blue, the median curve obtained by computing the median values of the parameters of the fit curves of Figure 1. Then it shows the 95% confidence interval expressed by two standard deviations from the median curve. Blue points are considered within normal range (could be attributed to the expected normal dispersion), while orange points, outside the 95% range have been considered as anomaly points.

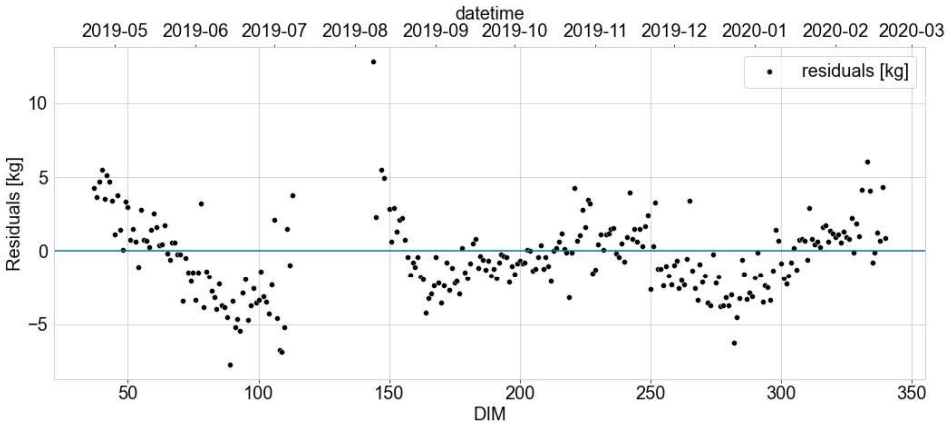

Figure 3: Scatterplot of the residuals of a fit curve.

As seen in Figure 3, residuals can be negative or positive, meaning in the latter case that milk yield can exceed the expected value. Therefore, in order to detect net production deficit in consecutive intervals, the daily residuals must be accumulated. In Figure 4 the cumulative curves of the expected and real milk yield trends are shown and overlapped to the trend line of their differences (the solid blue line). It is interesting to note that this differences (solid blue line in Figure 4) reach the value in correspondence with a DIM value equal to 90 days, about corresponding to the days with a production peak in the lactation curve. This is a recurrent condition with reference to the group of cows analysed, meaning the model was able to predict with high accuracy the cumulative milk yield in the first 90 days corresponding to the most productive stage of the lactation.

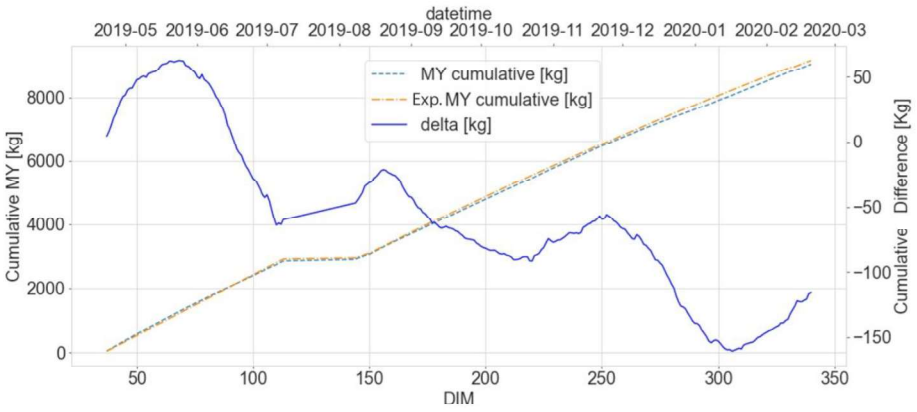

Figure 4: Cumulative curves of the expected and real Milk Yield (dashed and dash-dotted lines, respectively) and the corresponding differences (continuous blue line). In the legend delta represents the curve of the differences between the real and expected values.

Figure 5 shows the trend of the average difference between the expected and real cumulated milk yield for the entire group. As above anticipated the minimum values can be observed in correspondence to DIM ranges characterized by high values of the lactation curve (see also Fig. 2). This analysis has been performed on all the cows counting at least 100 consecutive days in milking and at least 8 valid fit curves.

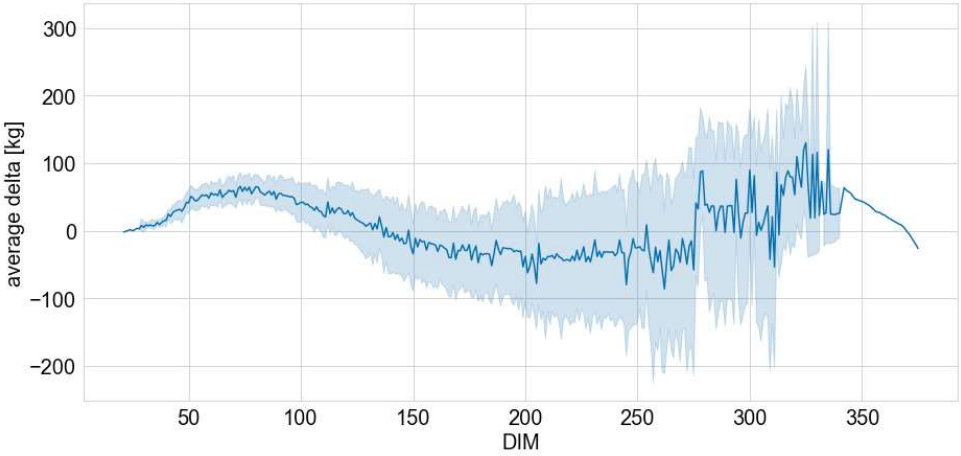

Figure 5: Average delta between expected and real cumulated milk yield. The light-blue coloured interval represents the 95% confidence interval.

Conclusions

Currently there is no systematic and statistically robust method for detecting production deviations from caused by various factors, such as environmental stress. The statistical method developed in this paper offers a robust way to identify production anomalies in the lactation period of individual cows on the basis of a multiple fit. Moreover, the use of a multiple of the standard deviation to define the acceptability range of the daily milk yield can leads to the introduction of a variable threshold, which could be used for production anomaly detection.

The anomalies identified with this method can be both positive and negative with respect to the range of acceptable values. This feature makes it clearly visible a further prospective of the use of this method: its application to an even greater number of cows and lactations will allow to collect an increasing number of anomalies. This can help a machine learning model, which is the subject of an ongoing study, in its training phase, making its forecasting performances more stable. This approach can be also used to classify daily production data as “normal” or “abnormal”.

Acknowledgements

This paper is part of the research project PRIN 2017 “Smart dairy farming: innovative solutions to improve herd productivity” funded by the Italian Ministry of Education, University and Research [20178AN8NC].

The study was possible thanks to the Marco Polo Scholarship of the University of Bologna, funding a three-month research period at the Department of Engineering for Livestock Management, Leibniz Institute for Agricultural Engineering and Bioeconomy for Dr. Miki Agrusti, PhD student.

References

Benni, S., Pastell, M., Bonora, F., Tassinari, P., & Torreggiani, D. (2020). A generalised addictive model to characterise dairy cows' responses to heat stress. *ANIMAL*, *14*, 418–424. https://doi.org/10.1017/S1751731119001721

Bonora, F., Benni, S., Barbaresi, A., Tassinari, P., & Torreggiani, D. (2018). A cluster-graph model for herd characterisation in dairy farms equipped with an automatic milking system. *Biosystems Engineering*, *167*, 1–7. https://doi.org/10.1016/j.biosystemseng.2017.12.007

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). https://doi.org/10.3390/ani11051305

Bovo, M., Benni, S., Barbaresi, A., Santolini, E., Agrusti, M., Torreggiani, D., & Tassinari, P. (2020). A Smart Monitoring System for a Future Smarter Dairy Farming. *2020 IEEE International Workshop on Metrology for Agriculture and Forestry, MetroAgriFor 2020 - Proceedings*. https://doi.org/10.1109/MetroAgriFor50201.2020.9277547

Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications of the ACM*, *24*(6). https://doi.org/10.1145/358669.358692

Heinicke, J., Ibscher, S., Belik, V., & Amon, T. (2019). Cow individual activity response to the accumulation of heat load duration. *Journal of Thermal Biology*. https://doi.org/10.1016/j.jtherbio.2019.03.011

Hempel, S., König, M., Menz, C., Janke, D., Amon, B., Banhazi, T. M., Estellés, F., & Amon, T. (2018). Uncertainty in the measurement of indoor temperature and humidity in naturally ventilated dairy buildings as influenced by measurement technique and data variability. *Biosystems Engineering*, *166*, 58–75. https://doi.org/10.1016/J.BIOSYSTEMSENG.2017.11.004

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science \& Engineering*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

Mbuthia, J. M., Mayer, M., & Reinsch, N. (2022). A review of methods for improving resolution of milk production data and weather information for measuring heat stress in dairy cattle. *Livestock Science*, *255*, 104794. https://doi.org/10.1016/J.LIVSCI.2021.104794

National Research Council. (1971). *A guide to environmental research on animals*. National Academy of Sciences.

Strpić, K., Barbaresi, A., Tinti, F., Bovo, M., Benni, S., Torreggiani, D., Macini, P., & Tassinari, P. (2020). Application of ground heat exchangers in cow barns to enhance milk cooling and water heating and storage. *Energy and Buildings*, *224*, 110213. https://doi.org/10.1016/j.enbuild.2020.110213

Van Rossum, G., & Drake, F. L. (2009). *Python 3 Reference Manual*. CreateSpace.

Waskom, M. L. (2021). seaborn: statistical data visualization. *Journal of Open Source Software*, *6*(60), 3021. https://doi.org/10.21105/joss.03021

Wood, P. D. P. (1967). Algebraic Model of the Lactation Curve in Cattle. *Nature*, *216*(5111), 164–165. https://doi.org/10.1038/216164a0