Challenges of integration and validation of farm and sensor data for dairy herd management

K. Schodl[1,2], B. Fuerst-Waltl[1], H. Schwarzenbacher[2], F. Steininger[2], M. Suntinger[2], F. Papst[3,4], O. Saukh[3,4], L. Lemmens[5], D4Dairy-Consortium[2] and C. Egger-Danner[2]

[1]University of Natural Resources and Life Sciences, Vienna, Gregor-Mendel-Strasse 33, 1180 Vienna, Austria
[2]ZuchtData EDV-Dienstleistungen GmbH, Dresdner Strasse 89/B1/18, 1200 Vienna, Austria
[3]Graz University of Technology, Rechbauerstrasse 12, 8010 Graz, Austria
[4]Complexity Science Hub, Josefstaedter Strasse 39, 1080 Vienna, Austria
[5]University of Veterinary Medicine Vienna, Veterinaerplatz 1, 1210 Vienna, Austria

Abstract

Precision livestock management has become an integral part of agriculture and dairy farms are increasingly using precision technologies such as sensor systems for herd health monitoring. Various companies exist, offering different sensors, e.g. accelerometers in collars or boli measuring cow activity and rumination, for different purposes such as fertility, health monitoring or feeding management. The D4Dairy project aimed, amongst others, to investigate the potential of sensor data and other farm and cow-specific data for disease prediction and genetic improvement of metabolic, udder and claw health. Results should lay a foundation for herd management tools and genetic health indices, which are expected to work across farms and sensor systems. Prior to any application, validation of sensor measurements is necessary. Some companies validated their sensor technologies in scientific studies by comparing sensor measurements (e.g., rumination time, duration of lying bouts) to a gold standard such as direct or video behavioral observations. Another aspect comprises validation of changes or patterns in sensor variables for a desired outcome such as heat or health alarms. Furthermore, for the implementation of monitoring of any kind, reliability of measurements is another crucial aspect. Erroneous measurements due to hardware or software malfunctioning have to be identified correctly and outliers have to be distinguished from true deviations. The latter is even more difficult for sensor measurements without possibilities for a plausibility check. Activity indices or other dimensionless sensor outputs lack established reference values and may differ even between animals equipped with the same sensor type whereas plausibility of rumen temperature or milk yield can also be assessed based on empirical knowledge. These issues had and partly still have to be overcome in the D4Dairy project and all projects with similar aims. In our contribution we want to present our approaches to sensor data validation, the problems we encountered and how we dealt with them including general recommendations for future studies in this area.

Introduction

Technologies have advanced on dairy farms during the last decades, which has released a potential for precision livestock farming. One of these advances is sensor technology for dairy cows for fertility and herd health management. Various technologies exist measuring for example activity, rumination, or reticular temperature to identify cows in heat for insemination, cows, which are about to give birth or send health alarms. These notifications are based on continuous recording and evaluation of these parameters based on changes in activity, rumination, or temperature patterns in the individual animal. Sensors recording these data are administered to the animal using for example collars, nosebands, foot bands, or rumen boluses. Crucial, of course, is that the sensor device is linked to the individual animal by a unique ID.

Many manufacturers already offer products using sensor technology for detecting cows in heat or at the onset of calving or cows, which may need medical treatment. However, sensor data is often very noisy and before these technologies yield reliable results it is important to clean and validate the (raw) data. There are different ways to approach sensor data validation, depending very much on the purpose and its area of use. The most obvious and sound approach is the comparison to a gold standard such as behavioral observations to assess the sensor's precision (Grinter *et al*., 2019). Due to the high time expenditure of behavioral observations other approaches relied on the comparison against other, already validated devices (Elischer *et al*, 2013) or the agreement between two devices on the same animal (Kok *et al*, 2015). Stygar *et al*. (2021) reviewed if and how various sensor systems offered on the market have been validated.

Aside from the validation of sensor technology for the target customers, these devices also offer a great opportunity to be used for research or other fields of application such as phenotyping for routine genetic evaluation or the development of new decision support tools for farmers. However, it is crucial for any user of the sensor data to know if and how data from these sensors have been validated and how much the data has been pre-processed prior to provision. Commonly, sensor systems available on the market are validated for one or a few specific purposes. Thus, data processing and software algorithms may emphasize some behaviors more than others to generate the most reliable alarms, for example emphasizing mounting activity for heat detection (Elischer *et al*., 2013). Depending on the use of these data for research or other purposes it may be necessary to additionally validate the sensor data or algorithms.

One of the aims of the D4Dairy project (https://d4dairy.com/) was to investigate the potential of sensor derived data and their integration with other farm data, such as veterinary records and diagnoses or data from automatic milking systems (AMS) for disease prediction and genetic improvement of metabolic, udder and claw health. For this purpose, farms already using sensor technology were motivated to participate in data collection. The sensor systems used on these farms were the rumen bolus by smaXtec (smaXtec animal care GmbH; 25 farms), the Lely T4C system (Lely International N.V., 35 farms), the SenseHub™ Dairy system (Allflex Livestock Intelligence, MSD Animal Health, 10 farms), the DelPro™ Farm Manager system by DeLaval (14 farms), and CowScout and Rescounter by GEA (9 farms). Additional information was derived from milking systems, veterinary records and diagnoses, national performance recordings, breeding information (e.g., genomic data, estimated breeding values), farm records and information on the operational structures on the farms, management information, climate sensor systems and weather stations, claw trimmings, rapid blood and milk tests for ketosis, and BCS and lameness recordings. Results should lay a foundation for herd management tools and genetic health indices, which are expected to work across farms and sensor systems. However, prior to any further analysis, sensor data had to be inspected and evaluated carefully so results would not be biased by erroneous data due to sensor malfunctioning or measurement errors. In this paper we describe how we approached data validation for the sensor systems by smaXtec, Lely, and Allflex as well as data from AMS and draw conclusions for automatization of data cleaning pipelines for future applications.

Animals, material, and methods

Sensor data were collected between January 2019 and August 2021 on farms with smaXtec sensors, between January 2020 and March 2021 on farms with the system by Lely and between January 2020 and May 2021 on farms with the SenseHub™ sensor system. The smaXtec bolus measured activity and temperature in the cow's reticulum and information was read from the sensors in intervals of ten minutes. The activity data was provided as a dimensionless number prior to any handling by the manufacturer whereas temperature was available in a raw format as well as after correction for temperature drops caused by drinking, which was further used for analysis. The other two sensor systems measured activity and rumination and were provided in a more aggregated format hourly (Allflex, Lely) or every two hours (Lely). Whereas Lely only provided one (dimensionless) number for activity, the data from the Allflex system was available as minutes of activity per hour for different activity levels (resting, medium activity, or high activity) and as an index called 'activity trend'. Rumination was either provided as minutes during the last 24 hours (Lely) or minutes per hour (Allflex).

Furthermore, data of daily milkings was provided by the Austrian central cattle database (Rinderdatenverbund – RDV) for single milkings as well as aggregated over 24 hours over the respective time period.

Approaches to data cleaning and outlier detection

Outliers deviate from the majority of a sample; however, they are difficult to identify in high dimensional data (Paulheim and Meusel, 2015) and it is challenging to clearly distinguish outliers from extreme deviations, which may even be of particular interest for a research question. For the smaXtec data set different approaches for outlier detection were taken up. The first one applied Isolation Forest, an unsupervised machine learning approach for outlier detection building on the co-dependency of data quality and model robustness (Papst *et al*., 2021): outliers can be identified based on an Isolation Forest Score derived from Isolation Trees, which are created during model training. This score indicates how likely it is for a data point to be an outlier (Papst *et al*., 2021). The second approach used classical plausibility checks based on domain expert knowledge and basic statistical approaches to data cleaning. This comprised the exclusion of duplicates and missing values as well as potential measurement errors with measurements occurring multiple times within the regular time window of data retrieval or where time to previous and following measurement exceeded the regular frequency. Whereas further plausibility checks of maximum and minimum values in the data were somewhat possible for temperature values using a priori knowledge on body temperature of cattle, this was not the case for the arbitrary activity value. Temperature values in the data set ranged from -42.8°C to 42.8°C. As negative temperature values are a physiological impossibility in warm-blooded animals this clearly indicated faulty data. However, defining a clear cut-off value for plausible temperature values was not straightforward and thus it was decided to use maximum deviation of three standard deviations from the overall temperature mean as a threshold. This yielded a plausible temperature spectrum between 36.6°C and 42.8°C for the whole data set (Figure 1).

As already mentioned before, there was no reference to assess plausibility of activity values and thus these were validated based on the associated temperature values. Additionally, if activity values were zero during at least twelve in 24 hours, these days were excluded because the sensor was probably not yet administered to the animals. All in all, approximately 5% of the smaXtec sensor data were discarded based on these decision criteria.

Sensor data by Lely and Allflex have been pretreated more intensively prior to provision and thus it was not known if and based on which criteria data have already been removed or altered beforehand. Although the SenseHub™ system also provided an arbitrary activity index, all other parameters on activity and rumination were available

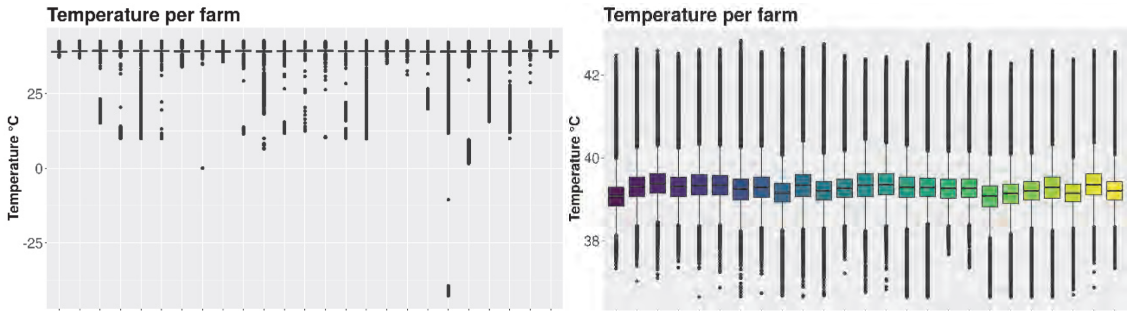

Figure 1. Reticular temperature per farm before (left side) and after (right side) data cleaning.

as minutes of this behavior per hour. Thus, if adding up all sensor variables yielded 60 minutes of behavior in one hour, data were considered correct. Furthermore, data were discarded if daily sums for rumination, eating and activity were zero for at least 24 hours suggesting that sensors were not administered, lost, or cows were removed from the herd and the sensor was not deactivated. The first day after sensor installation was also removed from the data set.

Lely sensor data either provided activity and rumination data in two-hour intervals and eating and rumination data hourly, respectively. Whereas rumination and eating were indicated in minutes of the last 24 hours, activity was again provided as dimensionless number for each two-hour interval. According to informations from the manufacturer sensors needed seven days to build a history for the animal and to reliably generate heat alarms. Thus, the first seven days of data after sensor installation were discarded. Due to the aggregation of rumination and eating time it was not easy to identify periods of potential sensor malfunctioning shorter than 24 hours. Thus, it was decided to remove data 24 hours before and after records indicating less than 10 minutes of ruminating or eating, respectively. This resulted in ranges of 10 to 893 minutes and 10 to 808 minutes for rumination and eating time, respectively. Furthermore, days with less than 11 and 22 measurements per day, respectively, were also removed from the data to avoid bias when performing further aggregation steps. Finally, four percent of the data had to be discarded based on these criteria.

Last but not least, data retrieved from AMS were validated and potential outliers were flagged including the reason. Criteria for outlier flags were the first milking of the lactation, milking intervals lower than 60 minutes and exceeding 24 hours, single milkings below one kilogram, and an hourly milk yield 50% above the ±10-day average (based on Hogeveen *et al*., 2001). Approximately 2% of the data did not fulfil these criteria. Furthermore, AMS data was matched with calving dates retrieved from the RDV to cross-validate lactation start and days in milk.

Discussion

Using commercially available sensor technology for research purposes enabled the inclusion of many farms and data due to their great availability compared to customized sensor technology in research settings. However, some limitations have to be considered such as the lack of knowledge about data processing by the sensor manufacturers prior to data provision due to trade secrets (Papst *et al*., 2019). Within the D4Dairy project this was not an issue because sensor data were not used to measure and interpret cow behavior itself. Rather, it was aimed at investigating how these sensor data may be helpful for early detection of diseases or their potential as auxiliary traits in breeding without interpretation of any physiological or behavioral relationships. Still, data had to be inspected and validated regarding measurement errors or other sources of faulty data (e.g., cows losing a sensor or devices being removed without being deactivated in the system).

All sensor types in this study were measuring activity, which was represented as a number of an undefined unit in all cases. Apart from noticeably long periods of zero activity, which indicated that the sensor must have been detached from the animal, there was no reference available for plausibility assessment. However, by using a priori knowledge about ruminal temperature as well as rumination and feeding time in cows, decisions for plausibility of a whole sensor record (activity and temperature and rumination or feeding time, respectively) were made based on the plausibility of these parameters.

Only few studies validating sensor devices for recording of rumination or using sensor measured rumination for heat, calving or disease detection provide precise information on data cleaning and outlier detection for sensor data. Reith *et al*. (2014) investigated heat detection using similar sensor systems for rumination recording like in this study and excluded values below 180 minutes and above 660 minutes per day. In her review, Beauchemin (2018) concluded that rumination and eating times range between 2.5 – 10.5 and 2.4 – 8.5 hours per day, respectively. In the present study these time ranges comprise more extreme values. However, whereas Reith *et al*. (2014) were interested in heat detection, the studies in D4Dairy aim at disease detection and lower rumination or feeding times may indicate a physiological response of unhealthy cows. The high values of rumination time observed in this sample may be due to the sensor type. According to Beauchemin (2018), substantially higher rumination times were recorded by acoustic sensors, which were also used in the present study, compared to other technologies. Furthermore, validation studies for acoustic rumination recording systems yielded good overall accordance to behavioral observations, but varied considerably between individual animals due to e.g., muscle or skin thickness or interference by background sounds (Beauchemin, 2018). Thus, this sensor system may have limited value for studies aiming at investigating rumination behavior itself, whereas disease detection based on the assessment of relative changes of patterns in individual animals may benefit a lot from this technology.

Data validation based on Isolation Forest is a promising method for identifying potential outliers. By assigning a score according to the likelihood of a data point to be an outlier, the final decision of excluding or including data may still be taken by the user, if intended. The main intention of this concept builds on the co-dependency of input data quality and model robustness to assess performance of predictive models given distribution shifts in incoming data (Papst *et al*., 2021). Whereas this data-driven approach focuses more on an application in the field, the second validation approach based on domain expert knowledge may be more suitable for 'upstream' research work on model formulation and feature definition.

Conclusion

Data validation and quality assurance is a crucial aspect when analyzing high dimensional data such as data from dairy cattle sensor systems. Even more so, steps of data cleaning should be comprehensive and made transparent if used for research purposes, which is not often the case in scientific literature using sensor data for device validation or further detection purposes. Limitations of data, which has been altered prior to data provision and which has been validated for specific purposes, should be taken into account in particular when using commercially available devices where processing of raw data is not known due to trade secrets. However, when being aware of these limitations these data offer a huge potential for the use in research for disease detection as well as the development of applications such as decision support tools or phenotyping strategies for auxiliary traits. Finally, integrating sensor data with other farm- and cow-specific data enables cross-validation between data sets and thus may help to additionally refine data for implausible values.

Acknowledgement

This work was conducted within the COMET-Project D4Dairy (Digitalization, Data integration, Detection and Decision support in Dairying; Project number 872039), which is supported by BMK, BMDW and the provinces of Lower Austria and Vienna in the framework of COMET – Competence Centers for Excellent Technologies. The Comet program is handled by the FFG grant number 872039.

References

Beauchemin, K.A. 2018. Invited review: Current perspectives on eating and rumination activity in dairy cows. J. Dairy Sci. 101:4762–4784.

Elischer, M.F., M.E. Arceo, E.L. Larcher and J.M. Siegford, 2013. Validating the accuracy of activity and rumination monitor data from dairy cows housed in a pasture-based automatic milkling system. J. Dairy Sci. 96:6412-6422.

Grinter, L.N., M.R. Campler and J.H.C. Costa, 2019. Technical note: Validation of a behavior-monitoring collar's precision and accuracy to measure rumination, feeding, and resting time of lactating dairy cows. J. Dairy Sci. 102:3487-3494.

Hogeveen, H., W. Ouweltjes, C.J.A.M. de Koning and K. Stelwagen, 2001. Milking interval, milk production and milk flow-rate in an automatic milking system. Livest. Prod. Sci. 72:157-167.

Kok, A., A.T.M. van Knegsel, C.E. van Middelaar, H. Hogeveen, B. Kemp and I.J.M. de Boer, 2015. Technical note: Validation of sensor-recorded lying bouts in lactating dairy cows using a 2-sensor approach. J. Dairy Sci. 98:7911-7916.

Papst, F., O.Saukh, K. Roemer, F. Grandl, J. Jakovljevic, F. Steininger, M. Mayerhofer, J. Duda and C. Egger-Danner, 2019. Embracing opportunities of livestock big data integration with privacy constraints. Proceedings of the 9th International Conference on the Internet of Things (IoT). October 2019. 27:1-4.

Papst, F., K. Schodl and O. Saukh, 2021. Exploring co-dependency of IoT data quality and model robustness in precision cattle farming. SensSys '21: Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems. November 2021:433-138.

Paulheim, H. and R. Meusel, 2015. A decomposition of the outlier detection problem into a set of supervised learning methods. Mach. Learn. 100:50-531.

Reith, S., H. Brandt and S. Hoy, 2014. Simultaneous analysis of activity and rumination time, based on collar-mounted sensor technology, of dairy cows over the per-estrus period. Livest. Sci. 170:219-227.

Stygar, A., Y. Gómez, G.V. Berteselli, E. Dalla Costa, E. Canali, J.K. Niemi, P. Llonch and M. Pastell, 2021. A systematic review on commercially available and validated sensor technologies for welfare assessment of dairy cattle. Fron.t Vet. Sci. 8:634338.