"MastiMIR" - A mastitis early warning system based on MIR spectra

L.M. Dale, A. Werner and F. Gollé-Leidreiter

Regional association for performance testing in livestock breeding of Baden-Wuerttemberg (LKV - Baden-Wuerttemberg), Stuttgart, Germany
Corresponding Author: ldale@lkvbw.de

Abstract

At farm level the mastitis disease appearance had decreased the milk production, produced veterinary costs, welfare issues, increased culling rate or caused lower milk payment. Because mastitis is associated with a wide range of characteristics that can be measured in milk and with recent advances in estimation of milk components using mid-infrared spectrometry, it is now possible to have the composition of several additional milk components such as fatty acids, lactoferrin, minerals, negative energy balance, non-esterified fatty acids and b-hydroxybutyrate or citrate, etc. The objective of this study was to build a spectrometric tool, such as MastiMIR for the determination in the milk quality of the animal health status, with the aim to evaluate the diagnosis usability and MIR indicators for the improvement of early mastitis prediction at LKV Baden-Württemberg. All data editing, modelling and calculations were done using the R statistical language and environment. The calibration data set contains around 9082 spectral data from around 1000 GMON herds. The validation approach was first cross validation 10 fold and a lot of 8 farms for an external validation. The 8 farms, chosen for the external validation, were the farms with the highest diagnosis registration and had to cover the important breeds e.g. 3 Holstein farms, 1 Red Holstein farm, 2 Brawn Swiss farms, 2 Simmental farms that are at LKV- Baden-Württemberg registered. To identify animal variables that were positively or negatively associated with mastitis determination, the spectral data set was first pre-processed by Savitzky-Golay first derivative to remove the offset differences between samples for baseline correction, before performimg Legendre polynomial modelling. Then the data was submitted to the combination of lasso regression using the "glmnet" R package. For the non-healthy group the spectral data with mastitis diagnosis for a given cow within 7 days before the new mastitis observation and the editing chosen was just test-day that had more than 400,000 somatic cell count (SCC). What come after the mastitis diagnostic was not taken into account for modelling. For the healthy group only spectra which had no diagnosis associated within ±60 days were used. For "glmnet" model were considered as fix effects the sampling moment, lactation stage and important LKV- Baden-Wtuerttemberg breeds and together with the Legendre polynomial data based on days in milk correction for the 212 OptiMIR selected wavenumbers of spectral data were input variables for MastiMIR model. Our MastiMIR calibration model showed a good accuracy (0.89) and medium prediction accuracy (0.83) we have to underline that was not finding until now any information in the literature of direct use of spectral data to predict the mastitis treat. The model provides four classes of Mastitis warning such as not, moderately, significantly and severely endangered. The moderately endangered class is a signal for the farmer. In that case the farmer would contact the veterinary and a control would be made in order to prevent the mastitis diseases. The MastiMIR model is a complementary tool for the SCC model.

Keywords: mastitis, spectrometry, MIR milk spectral data, dairy cow, cow health

Introduction

The mastitis definition is well known; mastitis is an inflammation of the mammary glands and can be caused by more than 50 different organisms. Usually, mastitis is diagnosed by cell number (SCC) and laboratory diagnostic methods. At farm level the mastitis disease appearance decreases the milk production, produces veterinary costs, welfare issues, and increases culling rate or causes lower milk payment. Mastitis is associated with a wide range of characteristics that can be measured in milk with recent advances in the estimation of milk components using mid-infrared (MIR) spectrometry. Also if a cow has mastitis, the composition of the milk will be affected and with it the MIR-milk-spectrum. The important message from the OptiMIR project was that not only the main components can be analysed with the MIR spectrometer, but also fatty acids (Grelet et al., 2014), minerals, lactoferrin (Soyeurt et al., 2011), BHB, acetate and citrates (Grelet et al., 2015), etc. complex features could also be identified, for example models for ketosis (Grelet etal., 2016), energy deficit (McParland etal., 2011, Smith et al., 2018) and methane emissions (Dehareng et al., 2012) were developed. Nowadays researches such as pregnancy detection (Laine et al., 2017) and mastitis detection tools could help farmers for better the herd management and better production. The objective of this study was to build a spectrometric tool, such as MastiMIR, for the determination of the animal health status from the milk quality, with the aim to evaluate the usability of Mastitis diagnosis in combination with MIR indicators in order to improve early mastitis risk prediction at the milk recording organisation LKV Baden-Württemberg.

Material and methods

Due to the health monitoring Baden-Württemberg (GMON cattle BW) which started in the beginning of 2010 diagnoses of approx. 1200 farms can be used for research and MastiMIR model. The diagnoses were documented by the veterinarian with the help of 86-part diagnostic keys. The gold standards to create the MastiMIR model were the mastitis diagnoses together with the spectral data. The diagnoses used for the model were: chronic, acute and subclinical mastitis, as well as coli mastitis. The model is based purely on standardized spectral data since all spectra registered at the MRO LKV- Baden-Württemberg level have been standardised starting from January 2012, due to the OptiMIR project participation. All data editing, modelling and calculations were done using the R statistical language and environment.

To identify animal variables that were positively or negatively associated with mastitis determination, the spectral data set was first pre-processed by Savitzky-Golay first derivative in order to remove the offset differences between samples for baseline correction, before performing Legendre polynomial transformation based on days in milk. Then the data was submitted to logistic regression in combination with LASSO variable selection and regularization and 10 fold cross validation using the "glmnet" R package. For the non-healthy group the spectral data with mastitis diagnosis for a given cow within 7 days before the new mastitis observation and the editing chosen was just test-day that had more than 400,000 somatic cell count (SCC). What comes after the mastitis diagnostic was not taken into account for modelling. For the healthy group only spectra which had no diagnosis associated within ±60 days were used. For “glmnet” model were considered as fix effects the sampling moment (with three variants: standard, morning and evening), lactation stage (if lactation number was greater than 5 it was taken as 5) and important LKV- Baden-Wtuerttemberg breeds (Holstein, Brown-Swiss and Simmental) and together with the Legendre polynomial data based on days in milk correction for the 212 OptiMIR selected wavenumbers of pre-processed spectral data were input variables for MastiMIR model. The calibration data set contained around 9,082 spectral data from around 1200 GMON herds. The first validation approach was based on a random split of data, 70% of data was used for calibration model and 30% for validation model. The second validation model was based on a lot of 8 farms for an external validation in order to exclude animal and farm effects. These 8 farms were the farms with the highest diagnosis registration rate and had to cover the important breeds e.g. 3 Holstein farms, 1 Red Holstein farm, 2 Brown Swiss farms and 2 Simmental farms from LKV-Baden-Württemberg were registered. For this two validation models the same data cleaning approach as for calibration model was used. Due to the external validation with the extreme values diagnosis cases, a third validation model is proposed with production data from a whole production year. Data from 1st October 2017 till end of September 2018 in combination with diagnosis data was aimed to verify if the proposal model could be afterwards used or not in production. From a research and statistical point of view, production data approach could show what in the reality exists and if the model will be working in reality. Statistical methods such as cox event time analysis were performed in order to define the mastitis risk/danger. The class limits were determined by using statistical methods such as cumulative probability and, the class size was negatively correlated with the mastitis class.

Results and discussions

Mastitis can only be predicted to a limited extent via the number of cells. Therefore a model based on spectral data, animal parameters, and mastitis diagnoses such as MastiMIR has been developed. After modelling with GLMNET in R, a sensitivity (the percentage of sick cows that were correctly identified as having the condition) of more than 85% in calibration and 75% for the validation and external validation model could be obtained. The specificity (the percentage of healthy cows that were correctly identified as not having the condition) is more than 90% for calibration model and 1st validation model and 83% for the external validation model, 2nd validation model.

The MastiMIR calibration model showed a good accuracy (0.89) and medium prediction accuracy (0.83). It can be underlined that until now no information of direct use of spectral data to predict the mastitis treat has been found in the literature. Regarding the 3rd validation model with production data, it can be seen that the sensitivity is just 63% while the specificity is 70%. This can be explained by the probable presence of untreated Mastitis cases, subclinical mastitis and missing registration of diagnosis events in the production data. The idea was to cover this group of data by means of a mastitis risk probability provided by a presumed logistic-linear relationship (S-curve) between MastiMIR probability and the mastitis danger. This model allowed by using different thresholds to distinguish four different risk/danger classes. The class limits were negatively correlated with the mastitis class and were statistical supported by cox event time analysis.

Table 1. MastiMIR calibration and validation statistics

MastiMIR Model	Sensitivity	Specificity
Calibration	85.6%	90.3%
1st Validation	74.9%	90.4%
2nd Validation	75.6%	83.3%
3rd Validation	63.9%	70.7%

It can be seen in the distributions of the MastiMIR and the SCC classes over the lactation week, that the mastitis class distribution has the shape of the lactation curve on both models. The MastiMIR class distribution on whole population from GMON cattle BW for the year 2017 is more pronounced than SCC class. Regarding the animals with MastiMIR danger or risk it can be pointed out that mastitis can occur also when the cows have less SCC. Animals with higher SCC may still have other diseases. There is also a difference between the healthy classes and moderately endangered and also significantly endangered. The size of the group decreases as with the SCC classes with the danger. The Cox event time analysis improved the classification. If an animal has mastitis risk/danger diagnostic, it can be seen earlier with the MastiMIR model than with the SCC class model. The transition from significantly endangered to severely endangered was better separated (differentiated). The transition from healthy to moderately endangered class was displayed earlier. If a cow has health problems due to mastitis, not only does it have a lower amount of milk or higher SCC but it also reacts with a change of the main milk components: the lactose content is negatively correlated with mastitis and the protein content and the fat-lactose ratio are positively correlated. A positive correlation also applies to the milk fine components sodium, Lactoferrin and BHB, as the literature has already confirmed.

Conclusions

Until now was not publish in the literature the development of a model based on the spectral data and veterinary diagnosis. The MastiMIR model is going from the fact that the animal is already diagnose by veterinary doctor and it is trying to find in the spectral data the finger print for mastitis warning. MastiMIR could help furthers the farmer to identify the early mastitis in order to have a better herd management. The MastiMIR model provides four classes of mastitis warning such as not, moderately, significantly and severely endangered. MastiMIR can be a good warning tool to prevent mastitis. The moderately endangered class is a signal for the farmer. In that case the farmer would contact the veterinary and a control would be made in order to prevent the mastitis diseases. Compared to the SCC model, the MastiMIR model shows an earlier occurrence of the 'slightly at risk' classification. The MastiMIR model is a complementary tool for the SCC model, MastiMIR can supplement the SCC classes. An evaluation in the field within the framework of the ELENA project is currently being prepared.

List of references

Dehareng, F., C. Delfosse, E. Froidmont, H. Soyeurt, C. Martin, N. Gengler, A. Vanlierde and P. Dardenne. 2012. Potential use of milk mid-infrared spectra to predict individual methane emission of dairy cows. Animal, 6(10): 1694-1701.

Grelet, C., J.A. Fernández-Pierna, H. Soyeurt, F. Dehareng, N. Gengler, and P. Dardenne. 2014. Creation of universal MIR calibration by standardization of milk spectra: example of fatty acids. EAAP - 65th Annual Meeting, Copenhagen. p.108.

Grelet, C., C. Bastin, M. Gelé, J.B. Davière, R. Reding, A. Werner, C. Darimont, F. Dehareng, N. Gengler and P. Dardenne. 2015. Milk biomarkers to detect ketosis and negative energy balance using MIR spectrometry. In Book of Abstracts of the 66th EAAP-Annual Meeting (p. 354). Wageningen Academic Publishers.

Grelet, C., C. Bastin, M. Gelé, J.B. Davière, M. Johan, A. Werner, R. Reding, , J.A. Fernandez-Pierna, F.G. Colinet, P. Dardenne, and N. Gengler. 2016. Development of Fourier transform mid-infrared calibrations to predict acetone, β-hydroxybutyrate, and citrate contents in bovine milk through a European dairy network. J. Dairy Sci., 99(6): 4816-4825.

Lainé, A., C. Bastin, C. Grelet, H. Hammami, F.G. Colinet, L.M. Dale, A. Gillon, J. Vandenplas, F. Dehareng, and N. Gengler. 2017. Assessing the effect of pregnancy stage on milk composition of dairy cows using mid-infrared spectra. J. Dairy Sci. 100:2863-2876.

McParland, S., G. Banos, E. Wall, M.P. Coffey, H. Soyeurt, R.F. Veerkamp, and D.P. Berry. 2011. The use of mid-infrared spectrometry to predict body energy status of Holstein cows. J. Dairy Sci. 94:1222-1239.

Smith, S. L., M. P. Coffey, and E. Wall. 2018. Association of milk MIR-derived body energy traits with fertility parameters in cows. Proc. 11^{th} WCGALP: 1121.

Soyeurt, H., C. Bastin, F.G. Colinet,V.R. Arnould, D.P. Berry, E. Wall, F. Dehareng, H.N. Nguyen, P. Dardenne, J. Schefers, and J. Vandenplas. 2012. Mid-infrared prediction of lactoferrin content in bovine milk: potential indicator of mastitis. Animal. 6(11):1830-8.