Rapid analysis of raw milk for monitoring of cow health

Aernouts Ben*, Polshin Evgeny, Lammertyn Jeroen and Saeys Wouter

BIOSYST-MeBioS K.U.Leuven, Kasteelpark Arenberg 30, 3001 Leuven, Belgium.
*Corresponding author: Ben.Aernouts@biw.kuleuven.be

Introduction

Next to the economic importance for the dairy farmer, composition of milk contains valuable information on the metabolic status of the cow, thanks to the intensive interaction between blood circulation and milk production. Therefore, the concentration and total production of different milk components can be used as a diagnostic tool for monitoring individual cow health and nutritional status[1]. Online measurement of milk components (with Vis/NIR spectroscopy) during milking two or more times a day would promote early detection of systemic and local alterations, thus providing a great input for strategic management decisions. Due to the high absorption by water in combination with the strong light-scattering by fat globules in raw milk, sample pathlength for the NIR range is limited to 1 mm and this considerably complicates on-line use of transmission measurements in a milking system[2]. Reflectance spectroscopy, however, does not pose any limitations on the dimensions of the measurement cell[3]. The potential of both measurement modes for predicting milk composition on-line was therefore compared in this study.

Materials and Methods

Samples

All 300 milk samples used in this study were collected in the context of a milk production registration system (MPR). In this system, every four weeks a representative milk sample (27 mL) from every individual cow on the participating farms was collected, covering a wide range of milk compositions and farm-specific effects. Including this variation in the calibration data is essential to develop accurate and robust prediction models. The fat, protein and lactose content of these samples was determined in the laboratory with a Milkoscan FT+ (Foss, Hillerød, Denmark) transmission FTIR instrument.

Near infrared spectroscopy

Reflectance spectra were acquired with a Zeiss Corona 45 VISNIR 1.7 (Carl Zeiss AG, Jena, Germany) diode array spectrometer combining a Si array with a 3.2 nm resolution in the 306.5 to 1135.5 nm range and an InGaAs array with a 6 nm resolution in the 944.5-1710.9 nm range. A built-in OMK measurement head (Carl Zeiss AG, Jena, Germany) was used to acquire diffuse reflectance spectra in a 45° configuration for a 20 mm diameter area on the sample. Since this instrument has no moving components and measures all wavelengths at the same time, it is very interesting for on-line use on the farm. Transmittance spectra were acquired with an ASDI Labspec®Pro (ASDI, Colorado, USA) spectrophotometer combining a Si diode array (1.4 nm resolution in the 350-1000 nm range) with a post-dispersive scanning monochromator consisting of 2 peltier cooled InGaAs detectors (1000-2500 nm range with 2 nm resolution). Due to the high absorption of NIR radiation by water combined with the strong light-scattering by the fat globules (0.1-15 μm diameter), the average transmittance through 1 mm of raw milk was found to be less than 1%. In order to get sufficient light on the detectors, a 1 mm cuvette (Quartz Suprasil® 100-QS, Hellma Benelux, The Hague, The Netherlands) was used in combination with a 200 W halogen light source (Fiber-Lite® DC-950, Dolan-Jenner Industries Inc., Massachusetts, US).

Milk samples were heated to 40°C in a water bath (type 1004, GFL, Burgwedel, Duitland) and stirred prior to the spectroscopic measurements. Transmittance and reflectance measurements were performed randomly per group and the order of the groups was also randomised.

Chemometrics

The acquired spectra and reference analysis values (fat and protein content) were imported into Matlab 7.10 (The MathWorks Inc., Nattick, MA) to be analysed with the PLS toolbox 5.5 (Eigenvector Technologies, Wenatchee, WA). Spectra were truncated to remove noisy regions and regions with insufficient signal (< 0.2% reflectance or transmittance). In order to separate the effect of the sample presentation mode from the effect of the wavelength range, the transmittance spectra were also limited to the 400-1700 nm range.

The best spectral pre-processing was determined empirically for each measurement method and for each component based on the prediction performance in cross-validation. The reflectance spectra were used as such or they were transformed to absorbance using the log(1/R) transform or the Kubelka-Munck transform.

Reference paper as:
A. Ben, P. Evgeny, L. Jeroen and S. Wouter (2012).Rapid analysis of raw milk for monitoring of cow health, in: Proceedings of the15th International Conference on Near Infrared Spectroscopy, Edited by M. Manley, C.M. McGoverin, D.B. Thomas and G. Downey, Cape Town, South Africa, pp. 324-327.

Secondly, the additive and or multiplicative effects were removed by a scatter correction (baseline correction, mulitiplicative scatter correction, standard normal variate and 1st & 2nd Savitzky-Golay derivatives). As a third pre-processing step, interference removal by orthogonal signal correction (OSC) was tested. Finally, the pre-processed spectra were mean-centered and partial least squares (PLS) regression models developed. For each combination of spectral pre-processing steps, the prediction performance of a PLS calibration model was evaluated based on a groupwise cross-validation where all samples of a group were either in the calibration set or the validation set. The model complexity was optimised by choosing the lowest number of latent variables for which the cumulative prediction error was not significantly worse than the lowest error obtained with up to 20 latent variables[4]. The models built for the different spectral pre-processings were then ranked by increasing root mean square error of cross-validation (RMSECV) and the pre-processing combining the smallest number of latent variables with absolute prediction errors not significantly ($\alpha = 0.05$) worse than those obtained by the model with the lowest RMSECV was chosen as the 'best'.

In a subsequent step, variable selection methods (variable importance in projection, interval PLS and genetic algorithms) were applied to determine whether some wavelength variables could be removed to make the models more parsimonious (fewer latent variables) or better performing (lower RMSECV). The selected variables which led to the most parsimonious model whose prediction performance was not significantly ($\alpha = 0.05$) worse than that of the model with the lowest RMSECV were used to build the final model. These prediction models, optimised based on their prediction performance in groupwise cross-validation on the calibration set, were then used to predict the concentrations in the samples of the test set. In order to determine whether the differences in prediction performance for the different models were significant, a two-way ANOVA test with measurement mode and sample number as the two factors was performed on the absolute prediction errors obtained for the test set[5].

Results and Discussion

The reliability of NIR spectroscopic calibrations is restricted to the range of constituent values and the variation in measurement conditions taken into account during calibration[3]. The basic statistics of the fat, crude protein and lactose content of the calibration and test sets considered are shown in Table 1. These datasets comprise a wide range of milk compositions representative for the variation in individual milk samples which had also been observed in previous research[6].

Table 1. Basic statistics of the milk composition as determined by the reference method for the calibration and test sets.

Component	Mean	Standard deviation	Minimum	Maximum	R^2		
					Fat	Protein	Lactose
a) Calibration set for Vis/NIR spectroscopy (200 samples)							
Fat (% w/w)	4.61	0.83	3.29	6.84	1.00		
Crude protein (% w/w)	3.68	0.44	2.80	4.68	0.37	1.00	
Lactose (% w/w)	4.58	0.33	3.48	5.09	0.07	0.20	1.00
b) Test set for Vis/NIR spectroscopy (100 samples)							
Fat (% w/w)	4.48	0.76	2.72	7.94	1.00		
Crude protein (% w/w)	3.58	0.48	2.65	5.01	0.21	1.00	
Lactose (% w/w)	4.64	0.33	2.92	5.22	0.02	0.16	1.00

The full reflectance spectra for all (300) milk samples after logarithmic transformation to absorbance are shown in Figure 1. Because of the high absorbance of NIR radiation by water, strong absorption bands around 970, 1190, and 1450 nm dominate the spectra. As a result, the characteristic absorption bands of fat, protein and lactose, which are much weaker, are hard to detect.

Reference paper as:
A. Ben, P. Evgeny, L. Jeroen and S. Wouter (2012).Rapid analysis of raw milk for monitoring of cow health, in: Proceedings of the15th International Conference on Near Infrared Spectroscopy, Edited by M. Manley, C.M. McGoverin, D.B. Thomas and G. Downey, Cape Town, South Africa, pp. 324-327.

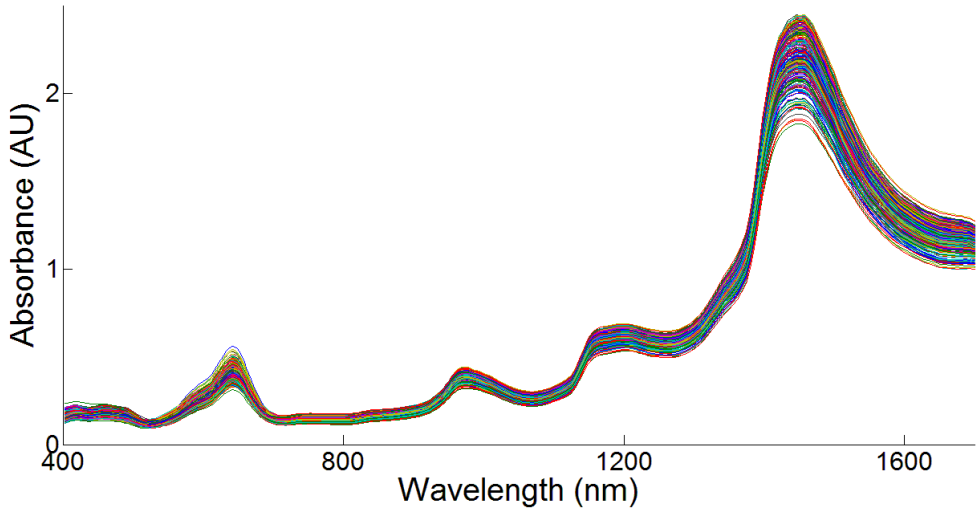

Figure 1. Absorbance spectra of all (300) milk samples (measured in reflectance).

In Table 2, statistics for the 'best' prediction models for the fat, crude protein and lactose content based on Vis/NIR spectra measured in reflectance and transmittance mode are summarised. Although there was no significant difference observed between transmittance and reflectance for predicting the fat, crude protein and lactose content, Vis/NIR reflectance spectroscopy performed better for fat and protein prediction (R^2 = 0.996 and 0.947). This is probably due to a combination of absorption and scattering of the light by fat globules and casein micelles in this spectral range. Higher fat and crude protein contents result in more scattering moieties (fat globules and casein micelles), presumably reflecting more light in the direction of the light source.

The best predictions for lactose were obtained using the transmittance spectra (R^2 = 0.883) while those based on the reflectance spectra were far less accurate (R^2 = 0.708). A possible explanation for this observation might be an insufficient interaction of the Vis/NIR radiation reflected back in the direction of the light source by the fat globules and casein micelles and the lactose which is diluted in the milk serum.

Table 2. Model statistics for fat, crude protein and lactose content prediction with Vis/NIR spectroscopy.

Component	Measurement mode	Pre-processing	Variable selection	LV	Cross-validation		Test set prediction		
					RMSECV[1]	R^2	RMSEP[1]	R^2	Bias
Fat[1]	Reflectance 400 - 1700 nm	Reflectance SG2D(45,2) OSC	FiPLS	3	0.042	0.997	0.052	0.996	0.017
	Transmittance 400 - 1700 nm	Absorbance SG1D(27,2) OSC	RiPLS	5	0.075	0.991	0.100	0.982	0.026
Crude Protein[1]	Reflectance 400 - 1700 nm	Reflectance SG2D(45,2) OSC	RiPLS	3	0.097	0.955	0.113	0.947	0.021
	Transmittance 400 - 1700 nm	Absorbance SG2D(51,2) OSC	RiPLS	3	0.131	0.919	0.162	0.890	0.003
Lactose[1]	Reflectance 400 - 1700 nm	Reflectance SG2D(51,2) OSC	FiPLS	2	0.166	0.720	0.182	0.708	0.059
	Transmittance 400 - 1700 nm	Absorbance SG1D(27,2) OSC	RiPLS	9	0.091	0.916	0.115	0.883	-0.043

[1] expressed in % w/w. SG1D(x) and SG2D(x) = 1st and 2nd Savitzky-Golay Derivatives with 2nd polynomial order and with a window size of x nm, OSC = Orthogonal Signal Correction, FiPLS and RiPLS = Forward and Reverse interval Partial Least Squares, LV = number of Latent Variables selected, RMSECV = Root Mean Square Error of Cross-Validation, RMSEP = Root Mean Square Error of Prediction.

In most cases, the best spectral pre-processing method was a combination of a Savitzky-Golay derivative (both 1st and 2nd derivatives with different numbers of smoothing points) followed by orthogonal signal correction. The iPLS variable selection method was always selected either in the forward or the backward mode. More results are shown and discussed in.[7]

Reference paper as:
A. Ben, P. Evgeny, L. Jeroen and S. Wouter (2012).Rapid analysis of raw milk for monitoring of cow health, in: Proceedings of the15th International Conference on Near Infrared Spectroscopy, Edited by M. Manley, C.M. McGoverin, D.B. Thomas and G. Downey, Cape Town, South Africa, pp. 324-327.

Conclusion

Vis/NIR reflectance spectroscopy combines a very high accuracy for fat and crude protein prediction with a simple measurement configuration which is highly suitable for on-line measurements of fluids; it gave poor results for lactose prediction. If the lactose concentration also has to be monitored, transmittance measurement is the preferred sample presentation. A more sophisticated sampling device would then be required to obtain representative milk samples of 1 mm thickness.

Acknowledgements

The authors would like to thank Milk Control Centre Flanders (MCC –Vlaanderen) for providing the milk samples and their reference analyses. Ben Aernouts and Wouter Saeys are respectively funded as aspirant and postdoctoral fellow of the Research Foundation – Flanders. Evgeny Polshin is funded through a grant from the Erasmus Mundus External Cooperation Window mobility scheme with Russia. Ben Aernouts would like to thank ICNIRS for the travel award to attend the NIR-2011 conference in Cape Town, South Africa.

References

1. N.C. Friggens, C. Ridder and P. Lovendahl, *J. Dairy Sci.* **90,** 5453-5467 (2007).
2. R.Tsenkova, S. Atanassova, K. Toyoda, Y. Ozaki, K. Itoh and T. Fearn, *J. Dairy Sci.* **82,** 2344-2352 (1999).
3. W. Saeys, A.M. Mouazen and Ramon, H. *Biosyst. Eng.* **91,** 393-402 (2005).
4. D.M. Haaland and E.V. Thomas, *Anal. Chem.* **60,** 1193-1202 (1988).
5. H.R. Cederkvist, A.H.Aastveit and T. Naes, *J. Chemometr.* **19,** 500-509 (2005).
6. P. Walstra, T.J. Geurts, A. Noomen, A. Jellema and M.A.J.S. van Boekel, *Dairy Technology, Principles of Milk Properties and Processes.* Marcel Dekker Inc., New York, USA (1999).
7. B. Aernouts, E. Polshin, J. Lammertyn and W. Saeys, *J. Dairy Sci.* **94,** 5315-5329 (2011).

Reference paper as:
A. Ben, P. Evgeny, L. Jeroen and S. Wouter (2012).Rapid analysis of raw milk for monitoring of cow health, in: Proceedings of the15th International Conference on Near Infrared Spectroscopy, Edited by M. Manley, C.M. McGoverin, D.B. Thomas and G. Downey, Cape Town, South Africa, pp. 324-327.