Robustness of near-infrared calibration models for the prediction of milk constituents during the milking process

Andreas Melfsen*, Eberhard Hartung and Angelika Haeussermann

Christian-Albrechts-University Kiel, Institute of Agricultural Engineering, 24098, Kiel, Germany

The robustness of in-line raw milk analysis with near-infrared spectroscopy (NIRS) was tested with respect to the prediction of the raw milk contents fat, protein and lactose. Near-infrared (NIR) spectra of raw milk ($n = 3119$) were acquired on three different farms during the milking process of 354 milkings over a period of six months. Calibration models were calculated for: a random data set of each farm (fully random internal calibration); first two thirds of the visits per farm (internal calibration); whole datasets of two of the three farms (external calibration), and combinations of external and internal datasets. Validation was done either on the remaining data set per farm (internal validation) or on data of the remaining farms (external validation). Excellent calibration results were obtained when fully randomised internal calibration sets were used for milk analysis. In this case, RPD values of around ten, five and three for the prediction of fat, protein and lactose content, respectively, were achieved. Farm internal calibrations achieved much poorer prediction results especially for the prediction of protein and lactose with RPD values of around two and one respectively. The prediction accuracy improved when validation was done on spectra of an external farm, mainly due to the higher sample variation in external calibration sets in terms of feeding diets and individual cow effects. The results showed that further improvements were achieved when additional farm information was added to the calibration set. One of the main requirements towards a robust calibration model is the ability to predict milk constituents in unknown future milk samples. The robustness and quality of prediction increases with increasing variation of, e.g., feeding and cow individual milk composition in the calibration model.

Keywords: On-line milk analysis, raw milk quality control, diffuse reflectance NIRS, calibration model robustness.

Introduction

Milk composition and its variation during lactation of cows can indicate imbalances in health or nutrition. Particularly changes of fat to protein ratio, milk urea nitrogen content, and concentration of ketone bodies provide suitable information on energy, protein and crude fibre supply, and on metabolic imbalances in dairy cows. Likewise, somatic cell count (SCC) and e.g. concentration of minerals and lactose in raw milk are indicators for udder diseases and helpful tools for monitoring of udder health (Friggens et al. 2007; Brandt et al. 2010). Since monitoring of animal health and metabolic indicators are increasingly part of herd management in dairy production, the knowledge of daily milk composition changes can assist this process. An analytical tool based on near-infrared spectroscopy (NIRS)

with the possibility to analyse milk during daily milking routine could provide that information. The robustness of the near-infrared (NIR) equipment makes this technology in general suitable for in-line application in milking parlours, with the aim to analyse individual cow's milk during milking. Analysing raw milk with NIRS was the topic in several studies either in a static measurement setup (Tsenkova et al. 1999, 2001, 2006, Schmilovitch et al. 2000; Chen et al. 2002; Saranwong & Kawano, 2008; Aernouts et al. 2011) or in an in-line application (Kawasaki et al. 2005, 2008). Our own investigations resulted in a good accuracy in predicting fat, protein, lactose and urea content as well as SCC in milk with NIRS both in a static measurement setup (Melfsen et al. 2012a) and in an in-line application (Melfsen et al. 2012b). These and results in the literature of the last decade suggest that NIRS is a promising tool to predict concentration of some milk constituents. However, in most studies on the use of NIRS in milk the prediction was done on randomised test sets closely

*For correspondence; e-mail: amelfsen@ilv.uni-kiel.de

connected with the calibration set. In practical farm application, the calibration model has to be applied to samples from farms, cows or seasonal circumstances that are either not part of the calibration or meanwhile have changed (independent test set). Chemometric models should possess a high robustness towards external variations to achieve sufficient prediction accuracy (Wang et al. 1991; Naes et al. 2004). In contrast to manufactured products, it is rather difficult to include all future variability in calibrations for natural animal products. The composition of raw milk is influenced by within-cow variability (cow age, lactation number, lactation state, breed, health status, oestrus and gestation status), within-nutrition variability (cows diet rationing, energy and protein supply, feeding time, pasture management), within-farm variability (number of milkings per day) as well as climatic and seasonal variability, such as temperature and daylight hours (Neitz & Robertson, 1991). Additional influences of the individual animal on NIR spectra are based to a large extent on scattering effects of fat globules (Tsenkova et al. 2000, 2009; Cattaneo et al. 2009). Furthermore, variations in measuring instruments and environment must be included in calibration models to achieve accurate prediction results. This is of particular importance when the spectroscopic signal from the compound of interest in milk is rather small (Thomas & Ge, 2000). Most of these effects are missing in the multivariate calibration models on milk composition analysis described in literature. Robust calibration models with the inclusion of large amounts of biological variability exist in literature mainly for herbal products such as wheat (Delwiche & Norris, 1993), kiwi fruits (McGlone & Kawano, 1998) and apples (Peirs et al. 2001, 2003).

The aim of this study is to evaluate and compare the robustness and prediction accuracy of raw milk constituents with NIRS by means of different multivariate models comprising internal and external effects.

Material and Methods

NIR spectra of raw cow milk were acquired during the milking process. At the same time, corresponding subsamples of each milking were taken during the milking process intended for reference analysis in the laboratory. Measurements were done on three different farms, each applying a different feeding diet, over a time period of six month.

NIR spectra acquisition

An experimental measuring setup has been designed, consisting of a commercially available milk meter (LactoCorder, WMB AG, Balgach, Switzerland), a measuring cell, designed as a flow-through chamber, a sample bottle, a contact reflection sensor head (PSS-H-B01) and a diode array spectrometer system with an InGaAs (Indium Gallium Arsenide) detector (PSS-1720; both Polytec GmbH,

Waldbronn Germany). The setup was installed in the long milk tube of one milking place per farm.

The measuring cell and thereupon the designed bypass system were attached to the sampling valve of the milk meter as described in Melfsen et al. (2012b). The valve diverted a representative sample of 6·25% of the raw milk during the milking process to the measuring cell, which was designed in such a way that the milk flow after the cell can be directed either into a sample bottle or back to the milk tube. The contact reflection sensor head was attached to the measuring cell and connected to the spectrometer. Near-infrared spectra were acquired in the measuring cell in diffuse reflection mode in the wavelength region of 851–1649 nm during the milking process every 500 ms while the milk was flowing through the cell. Each 2 kg of milking, spectra were averaged and subsamples of raw milk in the bypass system were collected in a sample bottle for reference analysis after passing the measuring cell. In a final step, spectra of complete milkings per cow were additionally averaged to predict constituents in composite milk samples.

Milk sampling

Bypassed milk was taken for reference analysis at each time 2 kg of milk had passed the milk meter. An aliquot of 45 ml from all milk samples was conserved with bronopol after sampling and sent to laboratory for reference analysis of fat, protein and lactose (Milkoscan FT+; Foss-Electric A/S, Hillerød, Denmark). The rest of the bypassed milk was collected and summed up with all subsamples of one complete cow milking to gather a milk sample of the total cow milking. This sample was also sent for reference analysis to the laboratory.

Dairy farms and measurement frequency

In-line measurements were done on three different farms located in the federal state of Schleswig-Holstein, Germany: one commercial dairy farm (ME); one research dairy farm (Max-Rubner-Institute Kiel, in Schaedtbek; SB); and one education and research dairy farm (Chamber of Agriculture Schleswig-Holstein, in Futterkamp; FK). Farm ME has been visited six times for measurements and milk sampling, farms SB and FK three times each. An overview on measurement sequence, total number of milk samples and feeding ratios is provided in Fig. 1. Each visit to farm ME comprised three consecutive milkings (two evening and one morning milking). The data sets of two consecutive visits were summarised in one data set (ME_{a1+a2}; ME_{b1+b2}; ME_{c1+c2}), referred to as visit a, b, and c, respectively. Milk samples were taken during 154 cow milkings of 95 different cows (total 1313 samples). The cows were in their first to eleventh lactation and 22–490 d in milk ($\bar{x} = 2\cdot6$ lactations and 217 DIM). Each visit (a, b, and c) to farms SB and FK comprised four consecutive milkings (two evening and two morning milkings). Milk samples at farm SB were taken during 116 cow milkings of 51 different cows (total 937 samples). The
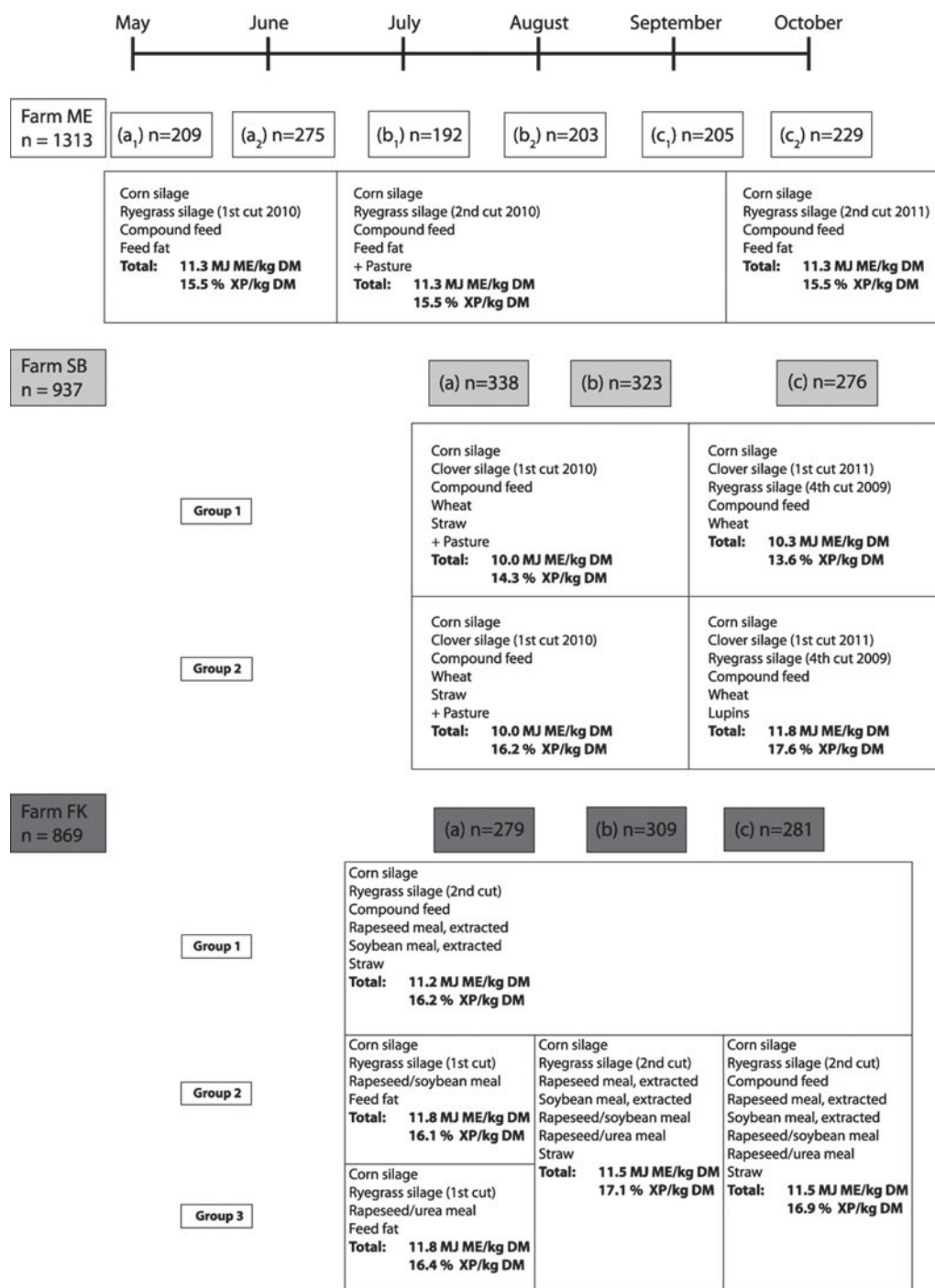
Fig. 1. Chronological sequence of milk sampling dates (a, b, c) at the three dairy farms (ME, SB, FK), total number of milk samples (*n*) per farm and diets fed during the experiment on the farms. (ME: metabolisable energy; XP: crude protein; DM: dry matter)

cows were in their first to ninth lactation and 39–467 d in milk ($\bar{x}$ = 2·4 lactations and 189 DIM).

Milk samples at farm FK were taken during 84 cow milkings of 72 different cows (total 869 samples). The cows were in their first to fifth lactation and 7–312 d in milk ($\bar{x}$ = 2·3 lactations and 165 DIM).

The amount of analysed cow milkings depended on the number of visits, the total number of cows and the type of milking parlour. The number of cows that appeared at more than one milking in the analysis was incidental due to randomised appearance of cows at the milking place at which the measuring setup was installed. All farms had the

Table 1. Overview of the different calibration sets and corresponding validation set

Name	Cal set	n cal set	Val set	n val set
ME-*RDM*	$ME_{2/3}$	875	$ME_{1/3}$	438
ME-*INT*	ME_{a+b}	879	ME_c	434
ME-*EXT*	$SB_{total}+FK_{total}$	1806	ME_c	434
ME-*EXT+1/3*	$SB_{total}+FK_{total}+ME_a$	2290	ME_c	434
ME-*EXT+2/3*	$SB_{total}+FK_{total}+ME_{a+b}$	2685	ME_c	434
SB-*RDM*	$SB_{2/3}$	624	$SB_{1/3}$	313
SB-*INT*	SB_{a+b}	661	SB_c	276
SB-*EXT*	$ME_{total}+FK_{total}$	2182	SB_c	276
SB-*EXT+1/3*	$ME_{total}+FK_{total}+SB_a$	2520	SB_c	276
SB-*EXT+2/3*	$ME_{total}+FK_{total}+SB_{a+b}$	2843	SB_c	276
FK-*RDM*	$FK_{2/3}$	579	$FK_{1/3}$	290
FK-*INT*	FK_{a+b}	588	FK_c	281
FK-*EXT*	$ME_{total}+SB_{total}$	2250	FK_c	281
FK-*EXT+1/3*	$ME_{total}+SB_{total}+FK_a$	2529	FK_c	281
FK-*EXT+2/3*	$ME_{total}+SB_{total}+FK_{a+b}$	2838	FK_c	281

Cal set: calibration set; val set: validation set

same cow breed (German Holstein) with a high lactation performance (9·500–10·600 l/305d).

Calibration and validation sets

A total of 3119 NIR spectra and milk samples were acquired (spectra taken during parts of the milking process ($n = 2765$) and averaged spectra ($n = 354$)). In a first step, spectra were merged into one dataset per visit per farm in farms SB and FK (SB_a–SB_c; FK_a–FK_c), and per two consecutive visits per farm, respectively, in farm ME (ME_a–ME_c) (Fig. 1.). Additional farm datasets contained all spectra and results of reference milk samples of the respective farm (ME_{total}, SB_{total}, FK_{total}). In a second step, datasets were combined in different calibration and validation sets (CV-set) (Table 1). The nomenclature of the CV-sets is based on the farm in the validation set on the one hand and on the status of the calibration set (random internal (RDM), internal (INT), external (EXT), combined (EXT+1/3 & EXT+2/3)) on the other hand.

Fully random internal calibration models (RDM): NIR spectra of the complete dataset per farm were subdivided into two random sample sets: a calibration set containing two-thirds of all samples and a test set containing the remaining samples.

Internal calibration models (INT): NIR spectra of the visits a and b were summarised in one calibration set per farm. Validation was done on visit c to the same farm (ME_c; SB_c; FK_c, Table 1).

External calibration models (EXT): The whole spectral information of two farms was summarised in one calibration set. Validation was done on the milk samples of visit c to the respective remaining farm, i.e. on the equal validation set like the validation in INT (Table 1). The aim of the external calibration sets was to evaluate their robustness to predict milk constituents in milk samples of an unknown third farm.

In a further step, the external calibration sets were complemented with spectral information of either visit a (EXT+1/3) or visits a and b (EXT+2/3) of the remaining validated farm. The validation set remained equal to the validation set in EXT and INT (Table 1).

Preprocessing of NIR data

Chemometric tools (SL Calibration Wizard v.1.1.0; SensoLogic GmbH, Norderstedt, Germany) were used for preprocessing of spectra, calibration of different milk constituents, and validation of results. Spectra normalisation was done to NIR reflectance spectra in order to reduce scattering effects. The applied normalisation of spectra scales the measured reflection to an ordinate between 0 and 1, where the minimum and maximum reflection value of each spectrum corresponds to none (i.e. 0) and total (i.e. 1) reflection, respectively, in the normalised spectra. The statistical PLS-1 method (partial least squares) was used for calibration of each constituent. Cross-validation (20 cross-validation segments) was used in the calibration process to estimate the number of principal components for the calibration. Validation was based on the previously created test set. Validation criteria for each PLS model were coefficient of determination (R^2), Root Mean Square Error of Prediction (RMSEP) and Ratio of Prediction to Deviation (RPD). In general the RPD is defined as ratio of SD of the validation set to Standard Error of Prediction (SEP) when no or only small bias is existing (Williams, 2001). Since the RMSEP is independent of appearances of high bias values, RPD is calculated as ratio of SD of the validation set to RMSEP in this study.

Statistical significance of differences between the varying CV-sets was tested by comparing absolute values of the residuals of NIR predicted and laboratory analysed milk contents (IBM SPSS Statistics, Version 19).

Results

An overview of the range, mean and SD of the fat, protein and lactose contents in the reference milk samples of each visit per farm is shown in Table 2. Since reference milk samples were taken every 2 kg of total milking, large ranges of the single milk contents were covered.

The data of all three milk contents in the calibration and validation sets of all created datasets were normally distributed.

Statistical calibration performance

The statistical performances of fat, protein and lactose prediction for the fully random internal CV-sets (ME-RDM; SB-RDM; FK-RDM) are shown in Table 3. Excellent calibration results were achieved with respect to RPD values in all three farms for predicting fat content in milk. Very good RPD values up to 6·36 were achieved when protein content in milk was predicted. RPD of lactose prediction showed

Table 2. Overview on the contents of milk constituents in the reference milk samples per visit per farm

			Fat(%)	Protein(%)	Lactose(%)
ME	**a**	Range	0·7–10·3	2·7–4·2	3·5–5·3
$n = 484$		Mean	4·1	3·3	4·8
		SD	1·55	0·30	0·25
	b	Range	0·7–14·8	2·4–3·9	3·8–5·3
$n = 395$		Mean	4·1	3·2	4·8
		SD	1·85	0·28	0·19
	c	Range	0·4–11·1	2·3–4·2	4·2–5·2
$n = 434$		Mean	4·1	3·2	4·8
		SD	1·72	0·35	0·17
	total	Range	0·4–14·8	2·3–4·2	3·5–5·3
$n = 1313$		Mean	4·1	3·3	4·8
		SD	1·70	0·32	0·21
SB	**a**	Range	0·8–7·8	2·7–4·0	3·5–5·1
$n = 338$		Mean	3·8	3·2	4·7
		SD	1·49	0·23	0·22
	b	Range	0·7–10·3	2·6–3·8	3·7–5·3
$n = 323$		Mean	3·9	3·2	4·7
		SD	1·70	0·30	0·28
	c	Range	0·8–11·0	2·4–4·0	3·9–5·3
$n = 276$		Mean	4·2	3·2	4·7
		SD	1·81	0·30	0·25
	total	Range	0·7–11·0	2·4–4·0	3·5–5·3
$n = 937$		Mean	3·9	3·2	4·7
		SD	1·67	0·28	0·25
FK	**a**	Range	0·7–10·6	2·6–3·9	3·9–5·1
$n = 279$		Mean	3·5	3·1	4·7
		SD	1·76	0·30	0·26
	b	Range	0·8–12·3	2·4– 3·7	4·0–5·2
$n = 309$		Mean	3·8	3·1	4·8
		SD	2·00	0·28	0·19
	c	Range	0·7–8·6	2·6–4·0	4·1–5·2
$n = 281$		Mean	3·5	3·1	4·7
		SD	1·58	0·32	0·19
	total	Range	0·7–12·3	2·4–4·0	3·9–5·2
$n = 869$		Mean	3·6	3·1	4·7
		SD	1·80	0·30	0·22

useful results for the farms ME and SB and a good accuracy in farm FK. Comparable RMSEP were achieved in all three farms for the prediction of fat and protein content. RMSEP for the prediction of lactose content was considerably lower at FK compared with the other farms. The bias was rather small for all milk constituents with regard to the range of data.

In Table 4 the model performances of farm internal CV-sets are shown (ME-INT; SB-INT; FK-INT). Compared with the results in Table 3, considerably higher bias can be observed for fat, protein and lactose prediction. Especially absolute values of protein and lactose bias were rather high with up to 0·0955 and − 0·0701%, respectively. Excellent calibration results were still achieved regarding RPD values of fat prediction for all three farms. Slightly higher RMSEP than in RDM can be observed for predicting fat content (Tables 3 & 4).

The prediction of protein content in the internal CV-set INT achieved a much lower accuracy than in RDM, with

RPD values around 2·0. An insufficient accuracy was achieved for all three farms when lactose content in milk was predicted. NIR prediction results in INT for protein and lactose showed a significantly higher deviation from the laboratory values compared with validation of randomised selected samples in RDM (Kruskal–Wallis-Test ($P \leqslant 0·001$)).

Model results of CV-sets with external validation spectra (ME-EXT; SB-EXT; FK-EXT) are shown in Table 4. According to RPD classification by Williams (2001), excellent prediction accuracy was achieved with regard to the fat content. Suitable RPD values for analytical purposes in most NIR applications in agricultural products with low SD were achieved for protein calibration. RPD values for lactose prediction in raw milk samples were rather small and classified as not recommended for analytical purposes.

Even though results in EXT did not achieve the same accuracy as in RDM, the prediction of protein content in raw milk in EXT was realised with a significantly higher accuracy than in INT (Kruskal–Wallis-Test ($P \leqslant 0·001$); Table 4). In addition, bias was smaller in EXT compared with the high bias levels in INT.

In the extended external calibration sets the previous external data sets were amended with farm information of that farm from which spectra were taken for validation. The statistical performances of calibration and validation for the extended external calibration sets EXT + 1/3 and EXT + 2/3, with different amounts of additional farm information, are shown in Table 5. The prediction of the fat content in raw milk was done with an accuracy classified as excellent regarding the high RPD values in all six calibration sets. Suitable up to good calibrations were developed for predicting protein content in milk. The RPD values of lactose prediction were rather small in all calibration sets.

Comparing EXT + 1/3 and EXT + 2/3 (Table 5) with results from EXT (Table 4), slightly improved calibration results were observed the more farm information was added. RMSEP of fat prediction was comparable with EXT. RMSEP was improved significantly with regard to lactose content in milk (Kruskal–Wallis-Test ($P \leqslant 0·05$)). Moderate bias values existed for CV-sets FK-EXT + 1/3 and FK-EXT + 2/3. The bias for prediction of fat and lactose in milk was rather high in FK-EXT + 1/3.

Discussion

The variation in the content of the milk ingredients, shown in Table 2, was in accordance with typical ranges of subsamples from total milkings (Nielsen et al. 2005). The changes of milk contents during milking time of cows have a positive effect on the performance of the calibration since they have a strong influence on spectra variability. Including subsamples of cow milkings in the calibration sets is particularly important for predicting the whole range of milk contents during the course of milking with sufficient accuracy in a practical in-line application.

Table 3. Statistical performances of random calibration and validation for the calibration sets ME-RDM; SB-RDM; FK-RDM with each having randomised spectra selection from the datasets ME, SB and FK respectively

ME-RDM	Calibration ($n = 875$)			Validation ($n = 438$)			
	SEC	RMSECV	R^2 cal	RMSEP	R^2 val	Bias	RPD
Fat [%]	0·19	0·19	0·99	0·17	0·99	0·0069	10·09
Protein [%]	0·06	0·07	0·97	0·06	0·96	0·0064	5·27
Lactose [%]	0·08	0·09	0·87	0·10	0·78	0·0084	2·12
SB-RDM	Calibration ($n = 624$)			Validation ($n = 313$)			
	SEC	RMSECV	R^2 cal	RMSEP	R^2 val	Bias	RPD
Fat [%]	0·16	0·16	0·99	0·16	0·99	− 0·0064	10·77
Protein [%]	0·04	0·06	0·98	0·06	0·95	− 0·002	4·61
Lactose [%]	0·07	0·11	0·93	0·09	0·89	0·005	2·99
FK-RDM	Calibration ($n = 579$)			Validation ($n = 290$)			
	SEC	RMSECV	R^2 cal	RMSEP	R^2 val	Bias	RPD
Fat [%]	0·15	0·16	0·99	0·16	0·99	− 0·0013	11·05
Protein [%]	0·04	0·05	0·99	0·05	0·98	0·0018	6·36
Lactose [%]	0·05	0·06	0·95	0·05	0·95	− 0·0013	4·35

SEC: standard error of calibration; RMSECV: root mean square error of cross validation; RMSEP: root mean square error of prediction; RPD: ratio of sD of the validation set to RMSEP

Comparing the prediction results of RDM data sets with calibration results from other randomised data sets in literature (Tsenkova et al. 1999, 2001, 2006; Schmilovitch et al. 2000; Chen et al. 2002; Kawasaki et al. 2008; Saranwong & Kawano, 2008) considerably improved or comparable prediction accuracy was achieved for all three farms with regard to the protein and lactose content. An important problem of fully randomised data sets is that the consideration of variability in future milk samples and of samples from different origin is missing in general. Naturally, accuracy of validation results from such fully randomised models is over-optimistic with regard to future samples (Peirs et al. 2003), a fact that is ignored in the RDM data set as well as in most literature calibration models of raw milk contents. The milk samples in a calibration set should span the natural range of variability of influences on the milk spectra in order to be able to predict milk constituents in future unknown milk samples with a sufficient accuracy (Naes et al. 2004). The distributions of milk content variation in randomised calibration and validation sets are nearly identical, which leads to advanced prediction accuracy compared with most independent validation sets. Nevertheless, additional instrumental and environmental variations must be included in calibration models to achieve accurate prediction results with high robustness in commercial farm applications, especially when the spectroscopic signal from the compound of interest is rather small (Thomas & Ge, 2000; Williams & Norris, 2001). Since spectroscopic signals in raw milk spectra in the selected wavelength range are dominated by absorption bands of O–H and C–H bonds this is of particular importance for the milk ingredients protein and lactose.

Thomas & Ge (2000) suggest two different ways of including sufficient sample information variability in the calibration set to gain robust calibration models. The variant of *controlled calibration* with selection or preparation of concentration series including the designated concentrations of the ingredient is useful for manufactured products. Due to high variability of agricultural products this method seems insufficient for creating robust calibration models for the prediction of future samples (Dhanoa et al. 1999; Peirs et al. 2003; Sileoni et al. 2011). The approach of *passive* or *natural calibration*, with a global calibration includes most part of biological variability over a sufficient period of time with samples from different origin. Despite the effort of time consuming development of natural calibrations (Thomas & Ge, 2000) and continuous efforts of updating or replacements with the introduction of new varieties to the model (Guthrie et al. 1998) this approach is preferable for reaching an adequate prediction accuracy in agricultural products (Peirs et al. 2003; Naes et al. 2004). Since variability in NIR spectra of raw milk is in addition to the milk constituent also influenced by external disturbances such as season, farm or cow (see above) it requires a large data set of calibration spectra to cover this variation and to achieve a sufficient robustness.

The approach of an internal calibration set (ME-INT; SB-INT; FK-INT) or the procedure of an external calibration set with calibration data of different origin (ME-EXT; SB-EXT; FK-EXT) is a more adequate solution for independent validation. As expected, the results were much poorer for these internal CV-sets compared with fully randomised CV-sets. Both calibration sets include the intra-farm variance of different cows with different lactation state and number

Table 4. Statistical performances of internal (ME-INT; SB-INT; FK-INT) and external calibration and validation (ME-EXT; SB-EXT; FK-EXT) with independent validation set consisting the last visit of the farms ME, SB and FK respectively (MEc, SBc, FKc)

ME-INT	Calibration (*n* = 879)			Validation (*n* = 434)			
	SEC	RMSECV	R^2 cal	RMSEP	R^2 val	Bias	RPD
Fat [%]	0·16	0·17	0·99	0·20	0·99	0·0197	8·45
Protein [%]	0·06	0·07	0·96	0·16	0·88	0·0955	2·22
Lactose [%]	0·15	0·16	0·55	0·18	0·18	0·0615	0·96

SB-INT	Calibration (*n* = 661)			Validation (*n* = 276)			
	SEC	RMSECV	R^2 cal	RMSEP	R^2 val	Bias	RPD
Fat [%]	0·13	0·14	0·99	0·19	0·99	− 0·0116	9·44
Protein [%]	0·07	0·07	0·94	0·16	0·77	0·0556	1·94
Lactose [%]	0·09	0·10	0·89	0·16	0·67	− 0·0498	1·56

FK-INT	Calibration (*n* = 588)			Validation (*n* = 281)			
	SEC	RMSECV	R^2 cal	RMSEP	R^2 val	Bias	RPD
Fat [%]	0·14	0·16	0·99	0·17	0·99	0·038	9·15
Protein [%]	0·05	0·06	0·97	0·18	0·74	0·0733	1·77
Lactose [%]	0·12	0·13	0·73	0·15	0·55	− 0·0701	1·28

ME-EXT	Calibration (*n* = 434) (*n* = 1806)			Validation (*n* = 434)			
	SEC	RMSECV	R^2 cal	RMSEP	R^2 val	Bias	RPD
Fat [%]	0·15	0·15	0·99	0·19	0·99	− 0·0073	8·94
Protein [%]	0·09	0·09	0·90	0·12	0·90	0·0425	2·95
Lactose [%]	0·10	0·12	0·81	0·17	0·29	− 0·0554	1·02

SB-EXT	Calibration (*n* = 2182)			Validation (*n* = 276)			
	SEC	RMSECV	R^2 cal	RMSEP	R^2 val	Bias	RPD
Fat [%]	0·17	0·17	0·99	0·19	0·99	− 0·0105	9·39
Protein [%]	0·08	0·09	0·94	0·12	0·85	0·0323	2·48
Lactose [%]	0·15	0·15	0·53	0·18	0·55	0·0494	1·41

FK-EXT	Calibration (*n* = 2250)			Validation (*n* = 281)			
	SEC	RMSECV	R^2 cal	RMSEP	R^2 val	Bias	RPD
Fat [%]	0·17	0·17	0·99	0·16	0·99	0·0565	9·64
Protein [%]	0·10	0·10	0·90	0·10	0·92	0·0296	3·14
Lactose [%]	0·12	0·13	0·72	0·15	0·53	–0·0649	1·27

SEC: standard error of calibration; RMSECV: root mean square error of cross validation; RMSEP: root mean square error of prediction; RPD: ratio of sᴅ of the validation set to RMSEP

and health status. Nevertheless, lack of robustness of the calibration model can be observed, which made it difficult to predict milk contents in future raw milk samples with satisfying accuracy. According to Tsenkova et al. (2000, 2009), Cattaneo et al. (2009) and Melfsen et al. (2012c), spectral information of individual cows is dominant in raw milk spectra, a factor that was underrepresented in the calibration sets in INT and EXT. Only, a few of the cows were represented both in the calibration and in the validation sets in INT. Since the spectra variability from these cows was already included in calibration creation, better prediction results for the milk contents can be expected. As expected, the RMSEP for the milk samples from these cows were significantly smaller (Mann–Whitney U-Test (*P* < 0·001)) than the RMSEP for milk samples from cows that were new

in the validation set (results not plotted). This was apparent for the ingredients protein and lactose of all three farms.

The objective of the analysis was to investigate whether a large external data set for calibration, with or without additional information from the validated farms and cows, avoids these effects and hence improves the robustness of the calibration models.

Even though no information of the farm, the cows or the feeding were existent in the form of raw milk spectra in the model development of external calibrations, these sets provided equal or even better prediction accuracy results compared with internal CV-sets. Since sample numbers were considerably higher in the external CV-set, improved prediction accuracy might have been caused by that fact. However, additional model developments with randomly

Table 5. Statistical performances of extended external calibration and validation for the calibration sets ME-EXT + 1/3; SB-EXT + 1/3; FK-EXT + 1/3 and ME-EXT + 2/3; SB-EXT + 2/3; FK-EXT + 2/3 with calibration sets containing total datasets of two farms amended with visits from the remained farm and validation sets containing the last third of the visits of the remained farm (MEc, SBc, FKc)

ME-EXT + 1/3	Calibration ($n = 2290$)			Validation ($n = 434$)			
	SEC	RMSECV	R^2 cal	RMSEP	R^2 val	Bias	RPD
Fat [%]	0·16	0·16	0·99	0·19	0·99	− 0·0193	8·89
Protein [%]	0·09	0·09	0·92	0·11	0·91	0·0131	3·17
Lactose [%]	0·12	0·13	0·76	0·15	0·38	− 0·0387	1·12
ME-EXT + 2/3	Calibration ($n = 2685$)			Validation ($n = 434$)			
	SEC	RMSECV	R^2 cal	RMSEP	R^2 val	Bias	RPD
Fat [%]	0·16	0·16	0·99	0·19	0·99	0·0054	8·96
Protein [%]	0·09	0·09	0·91	0·11	0·91	0·0143	3·27
Lactose [%]	0·12	0·13	0·74	0·14	0·38	− 0·02	1·18
SB-EXT + 1/3	Calibration ($n = 2520$)			Validation ($n = 276$)			
	SEC	RMSECV	R^2 cal	RMSEP	R^2 val	Bias	RPD
Fat [%]	0·16	0·17	0·99	0·19	0·99	− 0·0124	9·40
Protein [%]	0·09	0·09	0·92	0·12	0·84	0·0222	2·46
Lactose [%]	0·10	0·11	0·78	0·16	0·62	0·0365	1·58
SB-EXT + 2/3	Calibration ($n = 2843$)			Validation ($n = 276$)			
	SEC	RMSECV	R^2 cal	RMSEP	R^2 val	Bias	RPD
Fat [%]	0·16	0·16	0·99	0·19	0·99	0·0019	9·50
Protein [%]	0·10	0·10	0·89	0·12	0·85	0·023	2·52
Lactose [%]	0·11	0·12	0·75	0·14	0·67	− 0·0093	1·74
FK-EXT + 1/3	Calibration ($n = 2529$)			Validation ($n = 281$)			
	SEC	RMSECV	R^2 cal	RMSEP	R^2 val	Bias	RPD
Fat [%]	0·17	0·17	0·99	0·16	0·99	0·0551	9·64
Protein [%]	0·09	0·10	0·91	0·10	0·92	0·0312	3·18
Lactose [%]	0·13	0·14	0·71	0·14	0·56	− 0·0673	1·32
FK-EXT + 2/3	Calibration ($n = 2838$)			Validation ($n = 281$)			
	SEC	RMSECV	R^2 cal	RMSEP	R^2 val	Bias	RPD
Fat [%]	0·17	0·17	0·99	0·17	0·99	0·0526	9·51
Protein [%]	0·10	0·10	0·90	0·11	0·90	0·0163	2·99
Lactose [%]	0·15	0·16	0·58	0·13	0·53	− 0·0326	1·42

SEC: standard error of calibration; RMSECV: root mean square error of cross validation; RMSEP: root mean square error of prediction; RPD: ratio of SD of the validation set to RMSEP

reduced sample numbers at a level equal to the internal CV-sets still showed improved prediction results (results not plotted). Most likely, the occurrence of additional variability when more than one farm was used to build up the calibration sets represented feeding and seasonal changes better than INT and hence was responsible for a more robust calibration. The calibration sets included milk samples from a larger quantity of cows, a higher variability of lactation characteristics, multiple different diet layouts, as well as a higher variability of instrumental and environmental conditions.

Further improvements were observed when external calibration sets were provided with additional spectral information from that farm at which the validation was done. In this case, clear improvements for RMSEP of prediction of lactose content in milk and smaller bias values for protein and lactose prediction were observed. An increase in the amount of farm information in the external calibration sets led to an increased accuracy for predicting lactose content in milk and a further reduction of bias values for all milk constituents. Unexpectedly large bias values were still observed, however, when fat content in test set FK_c was predicted with the calibration sets in FK-EXT + 1/3 and FK-EXT + 2/3.

Conclusion

In this study the robustness and prediction accuracy of raw milk constituents with NIRS by means of different multivariate models comprising internal and external effects were

evaluated and compared. The analysed results in this study underlined that excellent to good accuracy was achieved for predicting fat, protein and lactose content in milk in fully randomised calibration and validation sets. In most cases, the variability of future milk samples is missing in randomised test sets, which causes an insufficient robustness towards independent spectra of future measurements. In consequence, validation on temporally independent spectra achieved much poorer prediction results, especially for the prediction of protein and lactose content. Prediction was improved when calibration was done on external spectra of other farms, which was probably due to the superior amount of sample variation in terms of feeding diets and cows in external calibration sets. Further improvements were achieved when additional information of the farm at which the validation was done was added to the calibration set. In summary, randomised created calibration and validation sets reflect the accuracy of raw milk content analysis only in a minor way. A robust calibration that covers up most of the variability of raw milk samples is more suitable and beneficial than calibration sets of single farms. For the creation of robust calibrations, it is recommended to include as much cow, feeding and seasonal variation as possible.

References

Aernouts B, Polshin E, Lammertyn J & Saeys W 2011 Visible and near-infrared spectroscopic analysis of raw milk for cow health monitoring: reflectance or transmittance? *Journal of Dairy Science* **94** 5315–5329

Brandt M, Haeussermann A & Hartung E 2010 Invited review: technical solutions for analysis of milk constituents and abnormal milk. *Journal of Dairy Science* **93** 427–436

Cattaneo TMP, Cabassi G, Profaizer M & Giangiacomo R 2009 Contribution of light scattering to near infrared absorption in milk. *Journal of Near Infrared Spectroscopy* **17** 337–343

Chen JY, Iyo C, Terada F & Kawano S 2002 Effect of multiplicative scatter correction on wavelength selection for near infrared calibration to determine fat content in raw milk. *Journal of Near Infrared Spectroscopy* **10** 301–307

Delwiche SR & Norris KH 1993 Classification of hard red wheat by near-infrared diffuse reflectance spectroscopy. *Cereal Chemistry* **70** 29–35

Dhanoa MS, Lister SJ, France J & Barnes RJ 1999 Use of mean square prediction error analysis and reproducibility measures to study near infrared calibration equation performance. *Journal of Near Infrared Spectroscopy* **7** 133–143

Friggens NC, Ridder C & Lovendahl P 2007 On the use of milk composition measures to predict the energy balance of dairy cows. *Journal of Dairy Science* **90** 5453–5467

Guthrie J, Wedding B & Walsh K 1998 Robustness of NIR calibrations for soluble solids in intact melon and pineapple. *Journal of Near Infrared Spectroscopy* **6** 259–265

Kawasaki M, Kawamura S, Nakatsuji H & Natsuga M 2005 Online real-time monitoring of milk quality during milking by near-infrared spectroscopy. *ASAE Meeting Presentation*, Paper Number: 053045

Kawasaki M, Kawamura S, Tsukahara M, Morita S, Komiya M & Natsuga M 2008 Near-infrared spectroscopic sensing system for on-line milk quality assessment in a milking robot. *Computer and Electronics in Agriculture* **63** 22–27

McGlone VA & Kawano S 1998 Firmness, dry-matter and soluble-solids assessment of postharvest kiwifruit by NIR spectroscopy. *Postharvest Biology and Technology* **13** 131–141

Melfsen A, Hartung E & Haeussermann A 2012a Accuracy of milk composition analysis with near infrared spectroscopy in diffuse reflection mode. *Biosystems Engineering* **112** 210–210

Melfsen A, Hartung E & Haeussermann A 2012b Accuracy of in-line milk composition analysis with diffuse reflectance near infrared spectroscopy. *Journal of Dairy Science* **95** 6465–6476

Melfsen A, Hartung E & Haeussermann A 2012c Potential of cow individual scatter correction for an improved accuracy of near infrared milk composition analysis. *Journal of Near Infrared Spectroscopy* **20** 477–482

Naes T, Isaksson T, Fearn T & Davies T 2004 *A User-Friendly Guide To Multivariate Calibration And Classification*. Chichester, UK: NIR Publications

Neitz MH & Robertson NH 1991 *Composition Of Milk And Factors That Influence It*. Department of Agriculture and Water Supply. Pretoria, South Africa

Nielsen NI, Larsen T, Bjerring M & Ingvartsen KL 2005 Quarter health, milking interval, and sampling time during milking affect the concentration of milk constituents. *Journal of Dairy Science* **88** 3186–3200

Peirs A, Lammertyn J, Ooms K & Nicola BM 2001 Prediction of the optimal picking date of different apple cultivars by means of VIS/NIR-spectroscopy. *Postharvest Biology and Technology* **21** 189–199

Peirs A, Tirry J, Verlinden B, Darius P & Nicola BM 2003 Effect of biological variability on the robustness of NIR models for soluble solids content of apples. *Postharvest Biology and Technology* **28** 269–280

Saranwong S & Kawano S 2008 System design for non-destructive near infrared analyses of chemical components and total aerobic bacteria count of raw milk. *Journal of Near Infrared Spectroscopy* **16** 389–398

Schmilovitch Z, Shmulevich I, Notea A & Maltz E 2000 Near infrared spectrometry of milk in its heterogeneous state. *Computers and Electronics in Agriculture* **29** 195–207

Sileoni V, van den Berg F, Marconi O, Perretti G & Fantozzi P 2011 Internal and external validation strategies for the evaluation of long-term effects in NIR calibration models. *Journal of Agricultural and Food Chemistry* **59** 1541–1547

Thomas EV & Ge N 2000 Development of robust multivariate calibration models. *Technometrics* **42** 168–177

Tsenkova R, Atanassova S, Toyoda K, Ozaki Y, Itoh K & Fearn T 1999 Near-infrared spectroscopy for dairy management: measurement of unhomogenized milk composition. *Journal of Dairy Science* **82** 2344–2351

Tsenkova R, Atanassova S, Itoh K, Ozaki Y & Toyoda K 2000 Near infrared spectroscopy for biomonitoring: Cow milk composition measurement in a spectral region from 1100 to 2400 nanometers. *Journal of Animal Science* **78** 515–522

Tsenkova R, Atanassova S, Ozaki Y, Toyoda K & Itoh K 2001 Near-infrared spectroscopy for biomonitoring: influence of somatic cell count on cow's milk composition analysis. *International Dairy Journal* **11** 779–783

Tsenkova R, Atanassova S, Morita H, Ikuta K, Toyoda K, Iordanova IK & Hakogi E 2006 Near infrared spectra of cows' milk for milk quality evaluation: disease diagnosis and pathogen identification. *Journal of Near Infrared Spectroscopy* **14** 363–370

Tsenkova R, Meilina H, Kuroki S & Burns DH 2009 Near infrared spectroscopy using short wavelengths and leave-one-cow-out cross-validation

for quantification of somatic cells in milk. *Journal of Near Infrared Spectroscopy* **17** 345–351

Wang Y, Veltkamp DJ & Kowalski BR 1991 Multivariate instrument standardization. *Analytical Chemistry* **63** 2750–2756

Williams PC 2001 Implementation of near-infrared technology. In *Near-Infrared Technology in the Agricultural and Food Industries*, pp. 145–170. (Eds P Williams & KH Norris) St. Paul Minn: American Association of Cereal Chemists

Williams PC & Norris K 2001 Variables affecting near-infrared spectroscopic analysis. In *Near-Infrared Technology In The Agricultural And Food Industries*, pp. 171–185. (Eds P Williams & KH Norris) St. Paul Minn: American Association of Cereal Chemists