Somatic cell count determination in cow's milk by near-infrared spectroscopy: a new diagnostic tool

R. Tsenkova, S. Atanassova, S. Kawano and K. Toyoda

J ANIM SCI 2001, 79:2550-2557.

Somatic cell count determination in cow's milk by near-infrared spectroscopy: A new diagnostic tool[1]

R. Tsenkova*[2], S. Atanassova†, S. Kawano‡, and K. Toyoda*

*Department of Environmental Information and Bio-production Engineering, Kobe University, Kobe 657-8501, Japan; †Department of Mathematics and Physics, Thracian University, Stara Zagora 6000, Bulgaria; and †National Institute of Food Industry, Tsukuba 305, Japan

ABSTRACT: The potential of near-infrared spectroscopy (NIR) in the region from 1,100 to 2,500 nm to measure somatic cell count (SCC) content of cow's milk was investigated. A total of 196 milk samples from seven Holstein cows were collected for 28, consecutive days, starting from 7th d after calving, and analyzed for fat, protein, lactose, and SCC. Three of the cows were healthy, and the remainder had periods of mastitis during the experiment. Near-infrared transflectance milk spectra were obtained using an InfraAlyzer 500 spectrophotometer. The calibration for logSCC was performed using partial least square (PLS) regression and different spectral data pretreatment. The best accuracy of determination was found for an equation that was obtained using smoothed absorbance data and 10 PLS factors. The standard error of calibration was 0.361, the calibration coefficient of multiple correlation was 0.868, the standard error of prediction for independent validation set of samples was 0.382, the correlation coefficient was 0.854, and the coefficient of variation was 7.63%. The accuracy of logSCC determination by NIR spectroscopy would allow health screening of cows and differentiation between healthy and mastitic milk samples. It has been found that SCC determination by NIR milk spectra is based on the related changes in milk composition. The most significant factors that simultaneously influenced milk spectra with the elevation of SCC were alteration of milk proteins and changes in ionic concentration of milk.

Key Words: Infrared Spectroscopy, Mastitis, Milk, Somatic Cell Count

 J. Anim. Sci. 2001. 79:2550–2557

Introduction

Somatic cell count (**SCC**) is a recognized indicator of cow's health and milk quality. Milk SCC reflects the level of infection and resultant inflammation in the mammary gland of dairy cows, associated with mastitis. Milk somatic cells are primarily leukocytes, which depend on intensity of the cellular immune defense, and some cells are from the mammary ducts. Mastitis is one of the most common dairy cow diseases, and it can cause considerable economic losses to dairy farmers. The losses are caused by several factors, such as decreasing milk yield, marked compositional changes in milk that reduced milk quality, treatment and labor costs, and increasing risk of early culling of cows.

Somatic cell count in milk has been accepted as the world standard for mastitis diagnosis (International IDF Standard 148A, 1995). Milk from healthy udders usually has a SCC less than 200,000 cells/mL, whereas for cows with subclinical mastitis SCC is greater than 200,000 cells/mL. Milk from cows with clinical mastitis may have SCC of several million cells per milliliter (Smith, 1995). Somatic cell count is an indicator of milk quality, and shelf-life is reduced for high-SCC milk and the processing quality and yield of some milk products is reduced when SCC increases (Barbano et al., 1991).

Near-infrared spectroscopy (**NIRS**) has been applied to measure the content of various constituents in milk (Sato et al., 1987; Laporte and Paquin, 1999; Tsenkova et al., 1999). Near-infrared spectroscopy in the shortwave region has been successfully used for mastitis detection in quarter milk (Tsenkova et al., 1992). Near-infrared spectroscopy has gained popularity because it has several advantages such as rapidity, no need of sample preparation and reagents, nondestructive measurement, and possibilities for on-line measurements.

The purpose of this study was to investigate the ability of NIRS to quantitatively determine logSCC content in composite milk from individual cows in the spectral region from 1,100 to 2,500 nm.

[1]This work was supported by the Program for Promotion of Basic Research Activities for Innovative Bioscience (PROBRAIN). The authors are grateful to GL Science Inc., Japan, for providing the software Pirouette Version 2.6.

[2]Correspondence: 1-1 Rokkodai Nada (phone/fax: +81 78 803 5911; E-mail: rtsen@eng.ans.kobe-u.ac.jp).

Received November 2, 2000.
Accepted May 23, 2001.

Materials and Methods

Milk Samples and Spectra

A total of 196 composite milk samples from seven Holstein cows were analyzed. The samples were collected for 28 consecutive days, beginning 7 d after calving. Cows were fed a ration containing corn silage, Timothy hay, commercial concentrate mixture (corn, barley, alfalfa meal, beet pulp, and $CaCO_3$), and trace mineral and vitamin premix; soybean meal provided 48% of the total CP of the diet. The average BW of the cows was 552 kg. Animals were fed twice for daily libitum intake and always had access to drinking water.

Each milk sample was divided into two subsamples. One was subjected to spectral analysis and another was analyzed for SCC by a fluoro-opto-electronic method using FOSSOMATIC 400 (Foss-Electric A/C, Hillerod, Denmark). Somatic cell count standards were used to calibrate the Foss instrument throughout the study. The repeatability coefficients of variation of this method are 4 to 5% for the region between 400,000 and 500,000 cells/mL and 5 to 10% for the region between 100,000 and 200,000 cells/mL (International IDF Standard 148A, 1995). The $\log_{10}$SCC was calculated to normalize the SCC distribution, and further quantitative analysis was made on the transformed data. Samples were also analyzed for fat, total protein, and lactose content (AOAC, 1990) by Milko Scan (Foss-Electric A/S). Three of the examined cows were healthy and had SCC lower than 137,000 cells/mL. One cow was mastitic during the entire experimental period: the measured SCC varied from 204,000 to 11,876,000 cells/mL. Three cows had mastitic and healthy periods (SCC was between 80,000 and 4,737,000 cells/mL).

Near infrared transflectance (**T**) spectra were obtained by the InfraAlyzer 500 spectrophotometer (Bran+Luebbe, Nordestedt, Germany), in terms of optical density log(1/T) in a wavelength range from 1,100 to 2,500 nm. A flow cell with a pathlength of 0.2 mm, connected with an automated liquid sampling system and taking milk samples and cleaning solution alternatively, was used. Before the spectral analysis each sample was warmed to 40°C in a water bath with temperature control ± 0.1°C. During the analysis, the same temperature was controlled through the use of an integrated water-jacketed holder of the flow cell connected with the water bath.

NIR Data Analysis

A commercial software program (Pirouette Version 2.6, Infometrics, Woodinville, WA) was used to process the data and to develop regression equations for logSCC determination.

The data were randomly divided into a set of 128 calibration samples and a set of 64 validation samples. Both data sets covered similar ranges of each investigated parameter. Methods used for preliminary examination of the data included smoothing the spectral data, multiplicative scatter correction, standard normal variance correction, baseline correction, and first or second derivative transformation of log (1/T) data. The smoothing and derivative transformations were based on the Savitzki-Golay second-order polynomial filter (Savitzky and Golay, 1964).

Calibration for SCC was performed by partial least square (**PLS**) regression using the calibration set of samples. The PLS used both the spectra and the respective reference data for the examined samples to determine latent variables (PLS factors). Criteria for calculating PLS factors are to describe the possible maximum amount of variance of NIR data and to be maximally correlated to the dependent variables. The PLS performs both the calculation of PLS factors and regression of the reference data simultaneously. After all PLS factors are calculated it is possible to combine the loading of these factors into a calibration (regression) equation. In this equation there is one coefficient for each wavelength. In developing the calibration equations, 15 PLS factors were set up as a maximum. The optimum number of PLS factors used in the models was determined by a cross-validation method. In cross-validation, five samples were temporarily removed from the calibration set to be used for validation. With the remaining samples, a PLS model was developed and applied to predict the respective milk component for each sample in the group of five. Results were compared with the respective reference values. This procedure was repeated several times until prediction for all samples was obtained. Performance statistics were accumulated for each group of removed samples. The validation errors were combined into a standard error of cross-validation. The optimum number of PLS factors in each equation was defined to be the one that corresponded to the lowest standard error of cross-validation. The performance of each regression was tested with an independent validation set of samples.

Calibration and validation statistics for each regression included the standard error of calibration, coefficient of multiple correlation, standard error of prediction, correlation coefficient between measured and NIRS-predicted values; bias and variation coefficients for calibration and validation set of samples were calculated. The statistical parameters were used for evaluating the accuracy of the NIRS determination.

Correlation coefficients between PLS factors and regression coefficients, logSCC, fat, protein, and lactose content of milk samples, respectively, were calculated by Microsoft Excel 97. The *F*-ratio of variance was calculated and used to determine significance of correlation coefficients (Steel and Torrie, 1980).

Results and Discussion

Raw (log1/T) spectra of unhomogenized milk samples are shown in Figure 1. Due to strong absorbance by O-H groups in water, two bands around 1,445 and 1,930

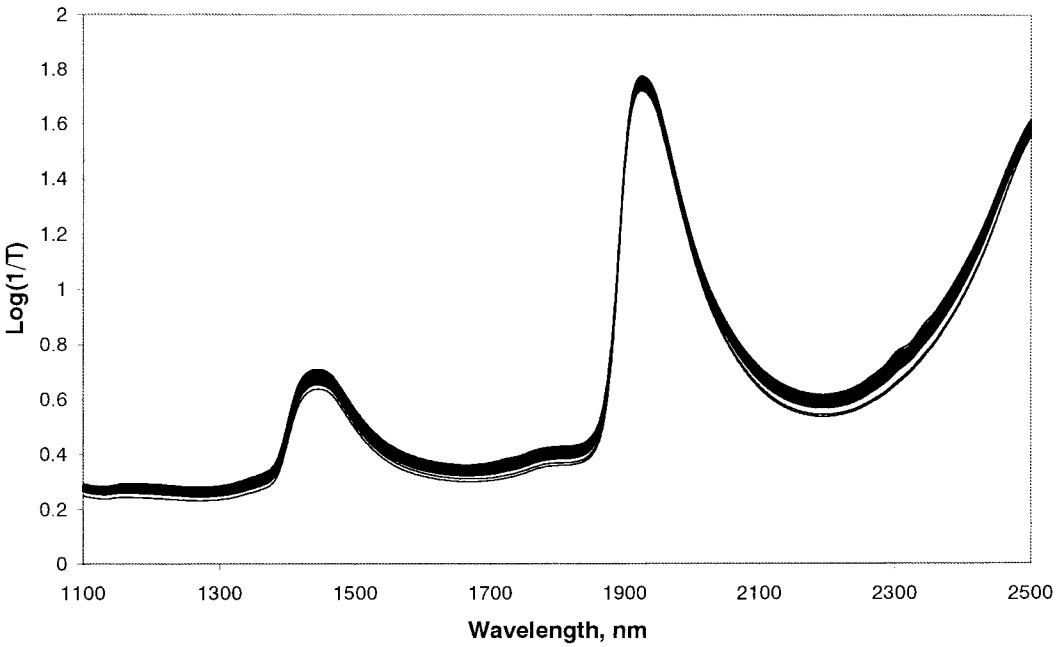

Figure 1. Near-infrared spectra of milk samples.

nm dominated the spectra. The characteristic absorption bands of fat and other milk components such as protein and lactose were very weak in comparison with the water bands and were difficult to visualize. This fact necessitated application of multivariate analysis to extract the spectral information connected with milk composition.

Results for the range, mean values, and standard deviation of logSCC, fat, protein, and lactose content of milk samples used in the calibration and the validation sets measured by reference methods are presented in Table 1. Calibration and validation statistics for regression equations are summarized in Table 2.

The best accuracy (the smallest standard error of prediction and the highest validation correlation coefficient) for determination of logSCC for the validation set of samples was found when smoothed log (1/T) data and 10 PLS factors were used. Figure 2 illustrates the relation between logSCC data and the respective values predicted by NIR regression based on smoothed log(1/T) data. The equations derived using multiple scatter correction or first derivative spectral data transformation also had similar accuracy. The obtained standard error of prediction of 0.382 and variation coefficient of validation of 7.63% would allow screening of milk samples and differentiation of healthy and mastitic milk samples. The obtained standard errors of prediction in this investigation were better than the value of 0.60 reported by Whyte et al. (2000) for the spectral region from 400 to 1,100 nm. The equations based on

Table 1. Range, mean, and standard deviation (SD) of fat, protein, lactose, and log somatic cell count content of the examined milk samples measured by reference methods

Parameter and set of samples	n	Mean	Min.	Max.	SD
Fat, %					
Calibration set	128	3.724	1.36	6.02	0.850
Validation set	64	3.762	1.28	6.25	0.864
Protein, %					
Calibration set	128	3.106	2.50	3.92	0.281
Validation set	64	3.106	2.49	3.86	0.284
Lactose, %					
Calibration set	128	4.458	4.04	4.86	0.162
Validation set	64	4.443	3.51	4.79	0.190
LogSCC					
Calibration set	128	4.964	3.845	7.075	0.693
Validation set	64	5.004	3.778	7.275	0.775

Table 2. NIRS calibration and validation statistics for log somatic cell count determination

Spectra transformation	Calibration set				Validation set			
	PLS factors	SEC	R	VCC,%	SEP	r	Bias	VCV, %
Smooth	10	0.361	0.868	7.27	0.382	0.854	–0.007	7.63
MSC	12	0.325	0.896	6.55	0.400	0.841	–0.005	7.99
SNV	11	0.342	0.883	6.89	0.423	0.820	–0.018	8.45
BC	9	0.362	0.865	7.29	0.445	0.800	0.016	8.89
1D	12	0.334	0.890	6.73	0.407	0.834	–0.044	8.13
2D	11	0.356	0.872	7.17	0.472	0.771	0.002	9.43

[a]MSC = multiple scatter correction; SNV = standard normal variate; BC = baseline correction; 1D = first derivative transformation; 2D = second derivative transformation; PLS = partial least square; SEC = standard error of calibration; R = coefficient of multiple correlation; SEP = standard error of prediction; r = validation correlation coefficient; VCC,% = variation coefficient for calibration set ([SEC/mean value]) × 100); VCV,% = variation coefficient for validation set ([SEP/mean value]) × 100).

spectral data transformed as second derivative or by using baseline correction showed a slightly lower accuracy. This result showed that baseline variations, which are corrected by second derivative transformation or baseline correction, contained information that is significant for SCC determination. Further investigation is needed to examine whether the accuracy could be improved by extracting the baseline and scatter information.

The PLS modeling not only aided development of quantitative models for logSCC determination but was also used as a tool for discerning the location of spectral information related to SCC. Regression equations for logSCC, based on smoothed log (1/T) spectral data and loading of factors included, were studied (Figure 3 through 5). Correlation coefficients between loading weight of each PLS factor and regression coefficients, and between scores of each PLS factor and logSCC, fat, protein, and lactose content of milk samples, respectively, were calculated. Results are presented in Table 3.

Only PLS factor 1 had some relationship to logSCC, but the contribution of that factor to the regression was very small. This result was probably due to some increase of the scattering caused by the presence of SCC. There was no correlation between the remainder of the factors and log SCC. These results showed that NIR determination of logSCC was based on relative changes in milk composition affecting milk spectral changes.

The highest significant correlation between PLS factors and fat content was found for factors 1 and 2,

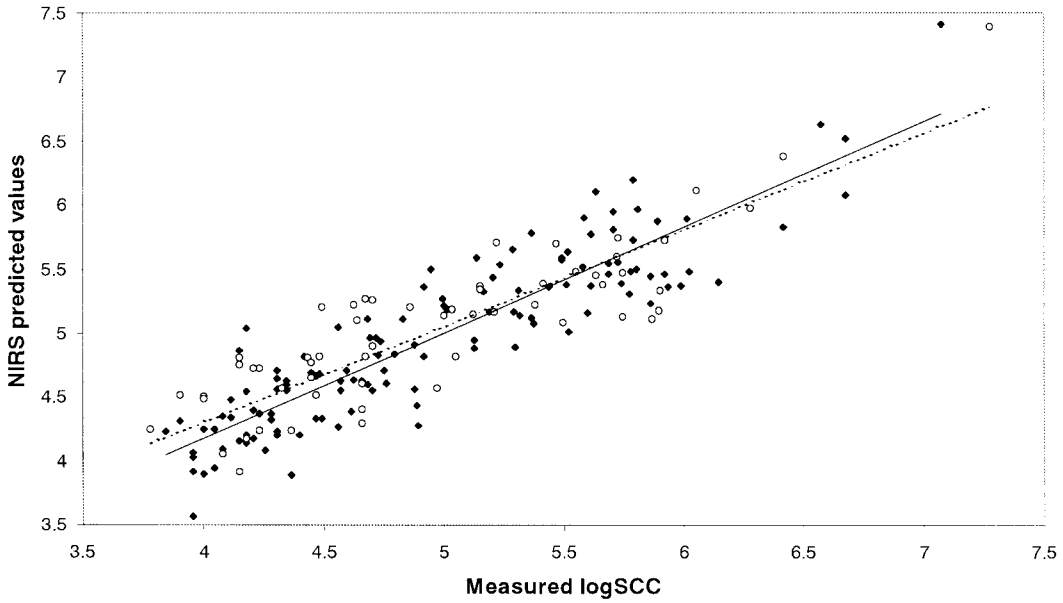

Figure 2. Relationship between actual and near-infrared-predicted values of log somatic cell counts (logSCC) ◆ = samples from calibration set, ○ = samples from validation set, —— = regression line for calibration set, and ------- = regression line for validation set.

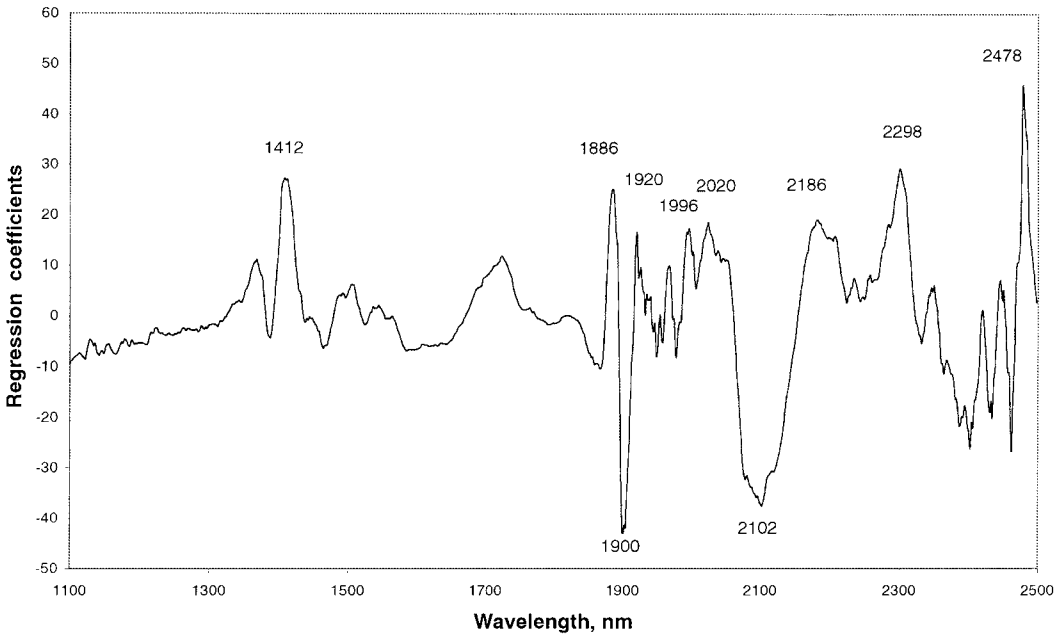

Figure 3. Regression coefficients for each wavelength in the partial least square model determining log somatic cell counts and based on smoothed log (1/T) milk spectra.

but the contribution of these factors to the log SCC regression was very small. Therefore, it could be concluded that fat content had little influence on logSCC determination.

The percentage of variations explained by each PLS factor decreased with its number. The third factor had the highest correlation with the protein content compared to the rest of the milk components, but protein content had the highest correlation with factor 9. Factor 4 had a higher contribution than factor 3 to the logSCC regression vector and the highest correlation with lactose compared to the rest of the factors. Mastitis decreases lactose content, which could explain the relationship between lactose content and determination of logSCC (Harmon, 1994). This finding emphasizes the possibilities for detecting changes with lactose when analyzing milk spectra and prove its strong relationship to SCC.

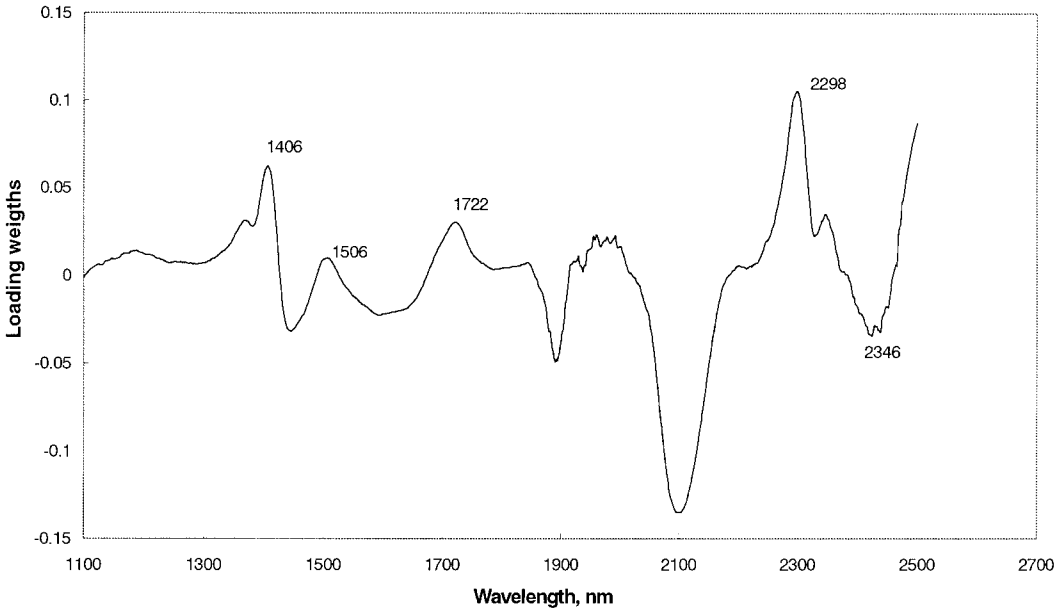

Figure 4. Loading factor 4 in the partial least square model, highly correlated with the regression coefficients and the lactose content of milk.

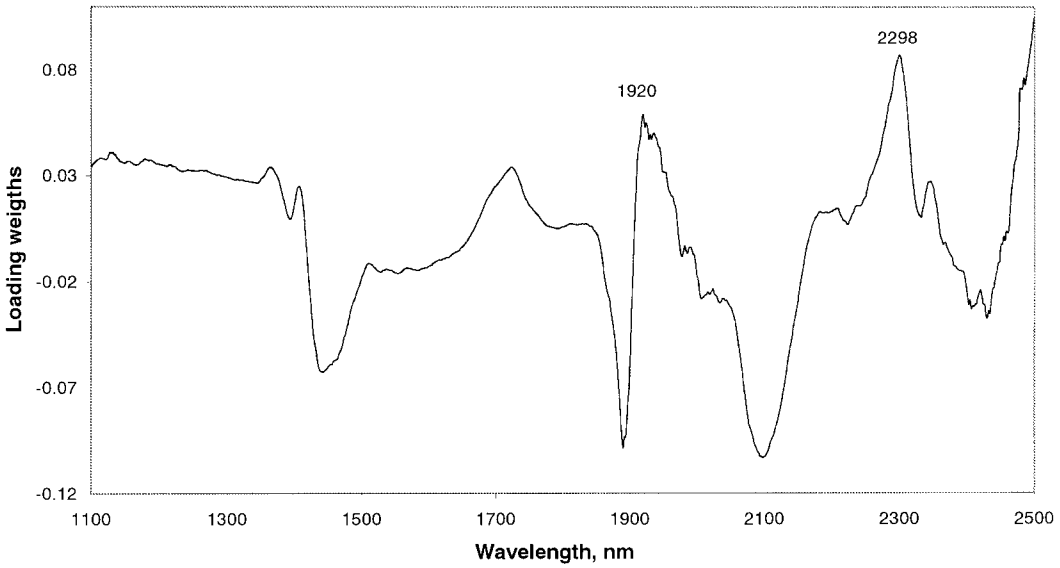

Figure 5. Loading factor 5 in the partial least square model, highly correlated with the regression coefficients and the protein content of milk.

Factors 5, 8, 6, and 10, which had the highest correlation with regression coefficients, had the highest correlation with protein content. This is not surprising, because mastitis caused alteration of protein fractions of milk. Mastitic milk has more proteolytic activity than normal milk, due to an increase of proteinase plasmin, which hydrolyzes the casein (Verdi et al., 1987; Urech et al., 1999). Harmon (1994) and Urech et al. (1999) reported decreased α_s-casein and β-casein content and elevated whey proteins and γ-casein in the total protein of mastitic milk. However, mastitis increases capillary permeability, which facilitates passage of proteins from blood to milk. These proteins are mainly serum albumin and immunoglobulins that are implicated in udder defense mechanisms (Poutrel et al., 1983).

High positive coefficients in the regression plot (Figure 3) were found at 1,412, 1,886, 1,920, 1,996, 2,020, 2,186, 2,298 and 2,478 nm. The wavelength 1,412 nm could be interpreted as the O-H water absorption band (Maeda et al., 1995). Probably the importance of this wavelength was related to an influence of the altered ionic concentration in milk on the water band caused by mastitis. Mastitis increases sodium and chloride content of milk (Harmon, 1994). The changes in ionic concentration could provoke a shift in the position of the water absorbance bands (Binette and Buijs, 1996; Molt et al., 1998). The other significant peaks at 1,886 nm and 1,920 nm are in the region of O-H stretching mode, first overtone and C=O stretch vibration, second overtone (Shenk et al., 1992). Two factors could explain the importance of those wavelengths. One of them is the influence of changes in quantity of water-soluble proteins and ionic concentration on the water absorption bands. The C=O absorption is likely due to absorption

Table 3. Correlation between loadings of partial least square (PLS) factors and regression vector and between scores of PLS factors and milk components

	Correlation coefficient				
PLS factor	Regression vector	Fat,%	Protein, %	Lactose, %	Log somatic cell count
1	0.045	–0.850**	0.152	–0.082	0.166
2	0.091	0.715**	–0.097	–0.077	$-3.7.10^{-9}$
3	0.234*	0.248*	0.323*	0.292*	$-2.9.10^{-7}$
4	0.319**	–0.340*	–0.314*	–0.345*	$-9.6.10^{-7}$
5	0.502**	0.179	0.324*	0.117	$-2.9.10^{-6}$
6	0.386**	0.147	–0.304*	0.008	$3.4.10^{-6}$
7	0.271*	–0.146	–0.031	–0.070	$-3.4.10^{-6}$
8	0.402**	–0.088	0.235*	0.015	$2.1.10^{-6}$
9	0.307*	0.090	–0.403*	0.250*	$2.3.10^{-7}$
10	0.352**	–0.230	0.334*	–0.085	$-1.8.10^{-5}$

*$P < 0.05$.
**$P < 0.01$.

of urea, the main nonprotein nitrogen compound in milk. Ng-Kwai-Hang et al. (1985) reported a significant contribution of SCC to the variation of nonprotein nitrogen in individual milk samples from 3,600 cows. The absorption at 1,996, 2,020, 2,186, 2,298 and 2,478 nm might be explained by protein and urea absorption. Diaz-Carrillo et al. (1993) reported peaks in NIR spectra of albumin at 2,170 and 2,300 nm and in spectra of casein at 1,968, 2,166, and 2,282 nm, respectively. The obtained peak in the regression vector at 2,298 nm was nearer the reported albumin peak at 2,300 nm than to the casein peak at 2,282 nm, which might show the greatest influence of water-soluble proteins such as albumin for SCC determination. The absorption at 2,186 nm in milk has been associated with Amide A+Amide III absorption by Charnik-Matusewicz et al. (1999), using two-dimensional correlation spectroscopy. Murayama et al. (2000) associated absorption at 2,290 nm in albumin with the contribution of C-H and amide III vibrations. Peaks at 2,020 nm could be associated with absorption due to the C=O stretch, second overtone, which corresponds to urea.

High negative coefficients in the regression vector plot were found at 1,900 and 2,102 nm (Figure 3). Absorption at these wavelengths might be explained by absorption of O-H stretch/C-O stretch combination vibration at 1,900 nm and O-H bend/C-O stretch combination vibration at 2,100 nm, associated with carbohydrates (Shenk et al. 1992). In the case of milk, the main carbohydrate component is lactose. Lactose content is negatively correlated with SCC (Harmon, 1994), which could explain the high negative coefficients obtained in the regression vector at 1,900 and 2,102 nm.

Further convincing evidence that determination of logSSC was based mainly on changes in lactose content, ionic concentration, and alteration of protein fractions of milk was given by the loading plot of significant PLS factors. Strong contributions of 1,406, 1,506, 2,102, 2,298 and 2,346 nm in the regression vector were also found to be very prominent in the loading of factor 4 (Figure 4), related to lactose. The loading of factor 5 (Figure 5) revealed 1,900 and 1,920 nm to be related to protein. Some wavelengths, such as 1,506 and 2,298 nm, were present in both plots. Further study is required to fully understand the assignment of all prominent wavelengths.

Implications

The accuracy of somatic cell count determination by near-infrared spectroscopy would allow health screening of cows and differentiation between healthy and mastitic milk samples. It has been found that somatic cell count determination by near-infrared milk spectra was based on the related changes in milk composition. The most significant factors that influenced near-infrared spectra were found to be the alteration of milk proteins and changes in ionic concentration of mastitic milk. Further investigations are needed to prove whether this approach is valuable for mastitis diagnosis in other ruminants and whether the short-wavelength near-infrared region, which is suitable for on-line measurements, would give similar results.

Literature Cited

AOAC. 1990. Official Methods of Analysis. 15th ed. Association of Official Analytical Chemists, Arlington, VA.

Barbano, D. M., R. R. Rasmussen, and J. M. Lynch. 1991. Influence of milk somatic cell count and milk age on cheese yield. J. Dairy Sci. 74:369–388.

Binette, J. F., and H. Buijs. 1996. Fourier transform near infrared process monitoring of multiple inorganic ions in aqueous solution. In: A. M. C. Davies and P. Williams (ed.) Near Infrared Spectroscopy: The Future Waves. pp 286–289. NIR Publications, Charlton, U.K.

Charnik-Matusewicz, B., K. Murayama, R. Tsenkova, and Y. Ozaki. 1999. Analysis of near-infrared spectra of complicated biological fluids by two-dimensional correlation spectroscopy: Protein and fat concentration-depending spectral changes of milk. Appl. Spectrosc. 53:1582–1594.

Díaz-Carrillo, E., A. Munoz-Serrano, A. Alonso-Moraga, and J. M. Serradilla-Manrique. 1993. Near infrared calibrations for goat's milk components: protein, total casein, β and κ-casein, fat and lactose. J. Near Infrared Spectrosc. 1:141–146.

Harmon, R. J. 1994. Physiology of mastitis and factors affecting somatic cell counts. J. Dairy Sci. 77:2101–2112.

International IDF Standard 148A. 1995. Milk Enumeration of Somatic Cells. International Dairy Federation, Brussels, Belgium.

Laporte, M., and P. Pacuin. 1999. Near-infrared analysis of fat, protein, and casein in cow's milk. J. Agric. Food Chem. 47:2600–2650.

Maeda, H., Y. Ozaki, M. Tanaka. N. Hayashi, and T. Kojima. 1995. Near infrared spectroscopy and chemometrics studies of temperature-dependent spectral variations of water: relationship between spectral changes and hydrogen bonds. J. Near Infrared Spectrosc. 3:191–201.

Molt, K., S. Berentsen, V. J. Frost, and A. Niemoller. 1998. NIR spectrometry—an alternative for the analysis of aqueous systems? Infrared Spectrosc. 10:16–21.

Murayama, K., B. Charnik-Matusewicz, Y. Wu, R. Tsenkova, and Y. Ozaki. 2000. Comparison between conventional spectral analysis methods, chemometrics, and two-dimensional correlation spectroscopy in the analysis of near-infrared spectra of protein. Appl. Spectrosc. 54:978–985.

Ng-Kwai-Hang, K. F., J. F. Hayes, J. E. Moxley, and H. G. Monardes. 1985. Percentage of protein and nonprotein nitrogen with varying fat and somatic cells in bovine milk. J. Dairy Sci. 68:1257–1262.

Poutrel, B., J. P. Caffin, and P. Rainard. 1983. Physiological and pathological factors influencing bovine serum albumin content of milk. J. Dairy Sci. 66:535–541.

Savitzky, A., and M. J. E. Golay. 1964. Smoothing and differentiation of data by simplified least squares procedures. Anal. Chem. 36:1627–1639.

Sato, T., M. Yoshino, S. Farukawa, Y. Someya, N. Yano, J. Uozumi, and M. Iwamoto. 1987. Analysis of milk constituents by the near infrared spectrophotometric method. Jpn. J. Zootech. Sci. 58:698–706.

Shenk, J. S., J. Workman, and M. O. Westerhaus. 1992. Application of NIR spectroscopy to agricultural products. In: D. Burns and E. Ciurczak (ed.) Handbook of Near-Infrared Analysis. Marcel Dekker, New York.

Smith, K. L. 1995. Standards for somatic cells in milk: Physiological and regulatory. IDF Mastitis Newslett. 144(21):7–9.

Steel, R. G. D., and J. H. Torrie. 1980. The Principles and Procedures of Statistics. 2nd ed. McGraw-Hill, New York.

Tsenkova, R., S. Atanassova, K. Toyoda, Y. Ozaki, K. Itoh, and T. Fearn. 1999. Near infrared spectroscopy for dairy measurement: Measurement of unhomogenized milk composition. J. Dairy Sci. 82:2344–2352.

Tsenkova, R. N., K. I. Yordanov, and Y. Shinde. 1992. Near-infrared spectroscopy for evaluating milk quality. In: Prospects for Automatic Milking. EAAP Publ. No. 65. pp 185–192. Pudoc, Wageningen, The Netherlands.

Urech, E., Z. Puhan, and M. Schallibaum. 1999. Changes in protein fraction as affected by subclinical mastitis. J. Dairy Sci. 82:2402–2411.

Verdi, R. J., D. M. Barbano, M. E. Dellavalle, and G. F. Senyk. 1987. Variability in true protein, casein, nonprotein nitrogen, and proteolysis in high and low somatic cell milks. J. Dairy Sci. 70:230–242.

Whyte, D., R. Claycomb, and R. Kunnemeyer. 2000. Measurement of somatic cell count, fat and protein in milk using visible to near infra-red spectroscopy. In: Proc. ASAE Annu. Int. Mtg., Milwaukee, WI.