Near-infrared Spectroscopy for
On-line Real-time Monitoring of Milk Quality:
Spectrum Analysis by Principal Component Analysis

Masataka Kawasaki, Graduate Student
Agricultural and Food Process Engineering Lab. Hokkaido University, Sapporo 060-8589, Japan, kawasaki@bpe.agr.hokudai.ac.jp

Shuso Kawamura, Associate Professor
Agricultural and Food Process Engineering Lab. Hokkaido University, Sapporo 060-8589, Japan, shuso@bpe.agr.hokudai.ac.jp

Hiroki Nakatsuji, Associate Professor
Animal Production Systems Lab. Hokkaido University, Sapporo 060-8589, Japan nakahiro@anim.agr.hokudai.ac.jp

Motoyasu Natsuga, Associate Professor
Agricultural Ecology and Engineering, Yamagata University, Tsuruoka 997-8555, Japan, toko@tds1.tr.yamagata-u.ac.jp

Written for presentation at the
2006 ASABE Annual International Meeting
Sponsored by ASABE
Portland Convention Center
Portland, Oregon
9 - 12 July 2006

Abstract. We have constructed a near-infrared (NIR) spectroscopic sensing system on an experimental basis. The NIR system can be used for on-line real-time monitoring of milk quality items such as fat, protein, lactose, somatic cell count, and milk urea nitrogen during milking with sufficient...
precision and accuracy. However, when the calibration models developed from a data set were validated using a different data set, the performance of the calibration models was poor except for fat content. It seemed that various factors such as cow individuality, lactation stage and calving time caused the poor performance. However, it was not known which factor affected milk spectra and which wavelength range of the spectra was affected by the factors. We therefore analyzed milk spectra by principal component analysis in order to determine the reasons for poor performance. It was found that milk spectra were greatly affected by fat content and calving time. Calving time had a particularly great effect on the spectra in the wavelength range of 860 to 880 nm. In summary, using principal component analysis, we found the factor most affecting milk spectra and the wavelength of the spectra affected by the factor.

**Keywords.** fat content, calving time, loading, score plot
Introduction

Near-infrared spectroscopy (NIRS) is a new nondestructive method for obtaining qualitative information on foods and agricultural commodities. NIRS has already been put to practical use in automatic rice-quality inspection systems in Japan (Kawamura et al., 2002; Kawamura et al., 2003a). NIRS has also been used to assess milk quality (Sato et al., 1987; Tsenkova et al., 1999; Tsenkova 2001a; Tsenkova et al., 2001b; Tsenkova et al., 2001c; Natsuga et al., 2002), but it has been difficult to apply NIRS to real-time on-line monitoring of milk quality of individual cows during milking.

We have constructed an on-line near-infrared (NIR) spectroscopic sensing system on an experimental basis to assess milk quality, and we (Kawamura et al., 2003b) have reported that the NIR sensing system can be used for real-time assessment of milk quality during milking with sufficient precision and accuracy. We (Kawasaki et al., 2005) have also reported that the performance of the calibration models was poor except for fat when the calibration models developed from a data set were validated using a different data set. It seemed that various factors such as cow individuality, lactation stage and calving time caused the poor performance. However, it was not known which factor affected milk spectra and which wavelength range of the spectra was affected. The aim of this study was to determine the reasons for poor performance by analyzing milk spectra using principal component analysis (PCA).

PCA is an effective method of multivariate analysis for revealing variations between spectra. The objective of PCA is to describe the data by assessing a small number of variables which linear combination of the original variables. These synthetic variables are called principal components (PCs). Initially calculated PCs include more information on spectral variances than do later PCs. The main information that the spectral data hold is in the first PC (PC1) and the second PC (PC2). The characteristic of a PC is indicated by the loading. The PC and the loading are linked by

$$PC = \sum_{k=1}^{n} l_k a_k,$$

where $l_k$ is the loading of the $k$ nm wavelength, $n$ is the number of wavelength points, and $a_k$ is the absorbance of the $k$ nm wavelength. Analyses of PCs provide the main information on spectral data. Analyses of loadings provide the useful wavelength range of spectra that contribute to the information of PCs (Robert et al., 1996; Tanaka et al., 1995).

Near-infrared spectroscopic sensing systems

An on-line near-infrared (NIR) spectroscopic sensing system for assessing milk quality of individual cows during milking was constructed on an experimental basis. The system consists of an NIR instrument, a milk flow meter, a milk sampler and a laptop computer (Table 1). The system was attached between a teatcup cluster and a milk bucket of a milking machine. Raw milk from the teatcup cluster continuously flowed into the milk chamber of the spectrum sensor and flowed out through an outlet pipe for surplus milk to the milk flow meter. The optical axes of a halogen lamp and optical fiber were set to the same levels (Figure 1). The volume of milk sample in the chamber was about 230 mL. The spectrum sensor acquired diffusion transmittance spectra through the milk. The spectra were recorded in the range of 600 to 1050 nm at 1-nm intervals every 10 seconds during milking. Six continuous spectra were averaged to obtain a spectrum for one minute. (Figure 2)
Table 1. Specifications of the near-infrared spectroscopic instrument constructed on an experimental basis.

<table>
<thead>
<tr>
<th>Devices</th>
<th>Specifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spectrum sensor</td>
<td>Diffusion transmittance spectrum sensor</td>
</tr>
<tr>
<td>Light source</td>
<td>Halogen lamp</td>
</tr>
<tr>
<td>Optical fiber</td>
<td>Silica glass fiber, 0.6 mm</td>
</tr>
<tr>
<td>Milk chamber surface</td>
<td>Glass</td>
</tr>
<tr>
<td>Volume of milk sample</td>
<td>Approx 230 mL</td>
</tr>
<tr>
<td>Distance between</td>
<td>93 mm</td>
</tr>
<tr>
<td>optical axis and milk level</td>
<td></td>
</tr>
<tr>
<td>Spectrometer</td>
<td>Diffraction grating spectrometer</td>
</tr>
<tr>
<td>Optical density</td>
<td>Transmittance</td>
</tr>
<tr>
<td>Wavelength range</td>
<td>600 - 1050 nm, 1-nm intervals</td>
</tr>
<tr>
<td>Wavelength resolution</td>
<td>Approx 5 nm</td>
</tr>
<tr>
<td>Photocell</td>
<td>Linear array CCD, 2048 pixels</td>
</tr>
<tr>
<td>Thermocontroller</td>
<td>Peltier cooling system</td>
</tr>
<tr>
<td>Data processing computer</td>
<td>DELL, Windows XP, Celeron 1.06GHz, RAM 384KB</td>
</tr>
<tr>
<td>A/D converter</td>
<td>12 bit</td>
</tr>
<tr>
<td>Spectrum data acquisition</td>
<td>Every 10 seconds</td>
</tr>
</tbody>
</table>

Figure 1. Plane view of the near-infrared spectrum sensor constructed on an experimental basis.

Figure 2. Flow chart of an on-line near-infrared spectroscopic sensing system for assessing milk quality.
**Cows and milk samples**

Two Holstein cows belonging to Hokkaido University were used in this experiment (Table 2). These cows had calved during the experimental period. Milk samples were collected from the milk sampler every minute during milking. The milk constituents were determined using a MilkoScan 4000 (Foss Electric, Hillerod, Denmark).

**Spectral data analyses**

Transmittance spectra were transferred to absorbance spectra. Pretreatments of the spectra such as second derivative (Savitzky-Golay derivatives, number of smoothing points: 51) and multiplicative scatter correction (MSC) were performed. The Unscrambler ver.6.11 (Camo AS, Trondheim, Norway) was used for PCA with cross-validation.

Two PCs (PC1 and PC2) were calculated for raw milk spectra from milking of each cow. Two PCs were used to develop a score plot (PC1 on the x-axis and PC2 on the y-axis). The score plot provides graphical comparisons of the samples by taking into account the main information extracted from the spectra.

<table>
<thead>
<tr>
<th>Cow number</th>
<th>Calving time</th>
<th>Date of calving</th>
<th>Experimental period</th>
<th>Number of reference samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1116</td>
<td>2</td>
<td>16 Oct. 01</td>
<td>23 Oct. 01 - 12 Jun. 02</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>19 Nov. 02</td>
<td>19 Nov. 02 - 25 Sep. 03</td>
<td>105</td>
</tr>
<tr>
<td>1110</td>
<td>2</td>
<td>20 Sep. 00</td>
<td>22 May. 01 - 18 Jul. 01</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>25 Nov. 01</td>
<td>6 Nov. 01 - 22 May. 02</td>
<td>54</td>
</tr>
</tbody>
</table>

**Results and discussion**

**Near-infrared spectra**

Figure 3 shows an example of two sets of absorbance spectra (raw NIR and their pretreated spectra) of raw milk from a cow. In the pretreated spectra, the large negative peak in the spectra at the wavelength of 970 nm indicates second-overtone absorption by O-H strings. Two negative peaks in the spectra at the wavelengths of 740 nm and 840 nm indicate overtone absorptions by C-H strings and C-C strings.

**Principal component analysis**

Figures 4, 5, 6 and 7 show score plots for raw milk spectra of each cow. The total variances of the two PCs of cow 1116 and cow 1110 were 83.7% and 84.4%, respectively. When the score plots were classified into three levels of fat content shows in Table 3 (low fat content: x, medium fat content: ■, high fat content: ◊), the score plots were separated into three groups along the x-axis (PC1) (Figures 4, 5). PC1 has the main information of the spectra, and scores were separated into groups according to fat content. Therefore, milk spectra were greatly affected by fat content. We (Kawasaki et al., 2005) have reported that the performance of the calibration model for determining fat content is good even if the model is affected by changes in various factors such as cow individuality, lactation stage and calving time. The reason for the
robustness of the calibration model for determining fat content was that milk spectra had much
information on fat content.

When the scores were classified on the basis of difference in calving time (second calving
time: ●, third calving time: ○), the score plots were separated into two groups along the y-axis
(PC2) (Figures 6, 7). PC2 has the main information of the spectral data next to PC1, and
scores were separated into groups according to calving time. Therefore, milk spectra were
greatly affected by calving time. We (Kawasaki et al., 2005) have reported that the difference in
calving time causes the poor performance of the calibration model. The reason for the poor
performance was that milk spectra were affected by the difference in calving time.

Figures 8 and 9 show the loadings of PC1 and PC2 and the average spectrum of each calving
time of each cow's datasets. In the case of PC1, peaks of the loading appeared at 740 and 840
nm. The wavelengths were the same as those of the negative peak of the spectra. Therefore,
the peaks (740 and 840 nm) were associated with fat (triacylglycerol).

In the case of PC2, the prominent peak of the loading appeared in the wavelength range
between 860 and 880 nm. Compared with the spectra of each calving time, the spectra differed
widely in the same wavelength range. Therefore, the spectra of the wavelength range were
affected by the difference of calving time. The spectra of the wavelength of 874 nm indicate
overtone absorptions by C-C strings. When the calving time is changed, mammary gland cells
are renewed. The size of fat globules changes with calving, and nutritional conditions also
change with calving. The composition of unsaturated fatty acid is changed by change in
nutritional conditions. Therefore, there is a possibility that the spectra wavelength was affected
by changes in the composition of fatty acid and size of fat globules. However, we could not
clarify which milk constituent with C-C strings was affected by the difference in calving time.

Figure 3. (a) Raw NIR spectra and (b) their pretreated spectra.

Table 3. Range of fat contents on score plots (Figures 4 and 5).

<table>
<thead>
<tr>
<th>Fat range (%)</th>
<th>cow 1110</th>
<th>cow 1116</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low fat content</td>
<td>0.85 - 4.28</td>
<td>0.83 - 3.75</td>
</tr>
<tr>
<td>Medium fat content</td>
<td>4.33 - 5.40</td>
<td>3.79 - 5.31</td>
</tr>
<tr>
<td>High fat content</td>
<td>5.41 - 8.33</td>
<td>5.35 - 8.20</td>
</tr>
</tbody>
</table>
Figure 4. Score plot of PC1 and PC2 for raw milk spectra of cow 1116. (Low fat content: ✗, Medium fat content: ■, High fat content: ◊)

Figure 5. Score plot of PC1 and PC2 for raw milk spectra of cow 1110. (Low fat content: ✗, Medium fat content: ■, High fat content: ◊)
Figure 6. Score plot of PC1 and PC2 for raw milk spectra of cow 1116. (second calving time: ●, third calving time: ○)

Figure 7. Score plot of PC1 and PC2 for raw milk spectra of cow 1110. (second calving time: ●, third calving time: ○)
Figure 8. Loadings of PC1 and PC2 and the average spectrum of each calving time of cow 1116.

Figure 9. Loadings of PC1 and PC2 and the average spectrum of each calving time of cow 1110.
Conclusion

We analyzed milk spectra by principal component analysis in order to determine the reasons for poor performance of calibration models. It was found that milk spectra were greatly affected by fat content and calving time. Calving time had a particularly great effect on the spectra in the wavelength range between 860 and 880 nm. Using principal component analysis, we found the factors most affecting milk spectra and the wavelength of the milk spectra affected by the factors.

References


