Comparison of two-dimensional correlation analysis and chemometrics in near infrared spectroscopy: protein and fat concentration-dependent spectral changes of milk

B. Czarnik-Matusewicz,[a,c] K. Murayama,[b] R. Tsenkova[b] and Y. Ozaki[a*]

[a]*Department of Chemistry, School of Science, Kwansei-Gakuin University, Uegahara, Nishinomiya 662-8501, Japan.*

[b]*Department of Environment Information and Bioproduction Engineering, Faculty of Agriculture, Kobe University, Rokkodai, Nada-ku, Kobe 657-8501, Japan.*

Introduction

The analysis of near infrared (NIR) spectra has recently made conspicuous advances because of three major factors.[1] One is the development of new multivariate analysis, such as soft independent modelling of class analogy (SIMCA) and genetic algorithms.[7–9] The progress in pre-treatment methods for multivariate analysis is also notable. Another is the reconsideration of conventional spectral analysis methods.[10–14] Of particular note is the use of difference spectra to detect subtle spectral differences in near infrared (NIR) spectra. Recent rapid improvement in spectrometers enables one to obtain NIR spectra with a high signal-to-noise ratio and high accuracy. Thus, one can calculate reliable difference spectra easily. Yet another factor is the introduction of two-dimensional (2D) spectroscopy to the NIR region.[15–24] We have demonstrated that generalised 2D correlation spectroscopy is powerful in exploring complicated NIR spectra.[4,5] It enhances apparent spectral resolution by spreading spectral peaks over the second dimension, enabling one to deconvolute overlapped bands.[15] One can also monitor the specific order of the spectral intensity changes occurring during perturbation by use of the asynchronous spectrum. It is also possible to investigate various inter- and intramolecular interactions through selective correlations of peaks.

The purpose of this article is to demonstrate the potential of generalised 2D correlation spectroscopy in analysing NIR spectra of milk, which is a complicated biological fluid, consisting mainly of proteins, lipids and carbohydrates. The NIR spectra of milk show rather poor signal-to-noise ratio and have changing baselines from one spectrum to another. Thus, to expand the use of 2D correlation spectroscopy to raw biological materials, we have investigated pre-treatment procedures. NIR spectra of 165 milk samples, with different protein and fat concentrations, were subjected to multiplicative scatter correction (MSC) and smoothing and the synchronous and asynchronous spectra were calculated for protein or fat concentration-dependent spectral changes of milk. The most important conclusion of the 2D correlation spectroscopy study is that the contribution of a particular component to the NIR spectra of milk can be extracted by calculating a power spectrum along the diagonal line of a synchronous spectrum generated from the component-dependent spectral variations of milk.

[c]Present address: Faculty of Chemistry, University of Wrocław, F. Joliot-Curie 14, 50-383 Wrocław, Poland.

Near Infrared Spectroscopy: Proceedings of the 9th International Conference

Another purpose of the present study is to compare 2D correlation analysis with principal component analysis (PCA). Both 2D correlation analysis and PCA are powerful tools for resolution enhancement, emphasising spectral features not readily observable in conventional spectra. Therefore, it is of particular interest to compare these two methods for improving spectral analysis of complicated biological materials.

Experimental

Milk samples from cows were employed for the present study. These cows were under routine feeding management at the National Institute of Animal Industry, Tsukuba, Japan. Protein and fat contents of the milk samples were determined by a Milkoscan 134 A/B (N Foss Electric, Denmark). The milk samples were homogenised and incubated at 40°C in a water bath prior to the NIR measurements. NIR spectra were measured with a step size of 2 nm at 40°C by an NIRSystems 6500 spectrometer. A liquid sample cell (1.0 mm path length) was used while carrying out this task.

Synchronous and asynchronous 2D NIR spectra were calculated using the algorithm recently developed by Noda.[25] Eight sets of 2D NIR correlation spectra have been completed, based upon eight series of dynamic spectra, constructed from NIR spectra of the milk samples, arranged in increasing order of fat or protein concentration. The data pretreatment (MSC and smoothing) were performed by the GRAMS/386 program package (Galactic Industries Co., Salem, NH, USA). For constructing the 2D correlation spectra, software written by Y. Wang (Kwansei-Gakuin University, Nishinomiya, Japan) with the Array Basic programming language (The Galactic Industries Corp.) was used.

Results and discussion

2D correlation analysis

Figure 1 shows NIR spectra in the 1100–2500 nm region of the 165 milk samples. The spectra closely resemble an NIR spectrum of water.[21] A broad band near 1450 nm is assigned to the combination of O–H symmetric and antisymmetric stretching modes of water, while an intense feature near 1930 nm is due to the combination of O–H bending and symmetric stretching modes of water. It is also noted that the signal-to-noise ratio of the spectra is not high and the baseline is changing from one spectrum to another. It is almost impossible to detect bands arising directly from fats, proteins and other metabolites of milk in the NIR spectra.

Milk contains light scattering particles in the form of fat globules and protein micelles. These particles may introduce erroneous information into NIR spectral measurements. We used MSC as a preprocessing tool to overcome the significant light scattering problems. For the MSC treatment, two data segments were constructed, one from 1100 to 1900 nm and another from 2000 to 2400 nm. The MSC procedure eliminated the undesired scattering of baseline level sufficiently but the signal-to-noise ratio was still too low to perform the 2D correlation analysis. Thus, we employed smoothing to solve the problem of the low signal-to-noise ratio.

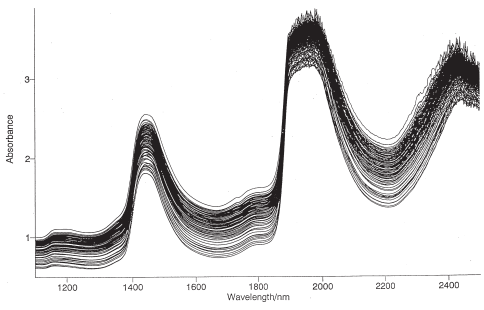

Figure 1. NIR spectra in the 1100–2500 nm region of the milk samples (Reproduced from Reference 21 with permission. Copyright© Society for Applied Spectroscopy).

Figures 2(a) and (b) show synchronous 2D correlation spectra in the 2000–2400 nm region constructed from the fat or protein concentra-

tion-dependent NIR spectral changes after pre-treatments, respectively. Note that the two synchronous spectra are markedly different from each other. The synchronous spectrum, constructed from the fat concentration-dependent NIR spectral variations, shows autopeaks at 2311 and 2346 nm, while the corresponding spectrum from the protein concentration-dependent spectral changes develops autopeaks at 2047, 2075, 2100, 2351 and 2375 nm. The appearances of the autopeaks mean that the intensities of these bands vary most significantly with the increase in concentrations of fats or proteins in milk.

Figures 3(a) and (b) depict power spectra along the diagonal line on the synchronous spectra shown in Figures 2(a) and (b), respectively.[21] It is noted that most of the bands which appeared in the power spectrum for the fat concentration-dependent spectral changes [Figure 3(a)] are assigned to fats while those in the power spectrum for the protein concentration-dependent spectral variations [Figure

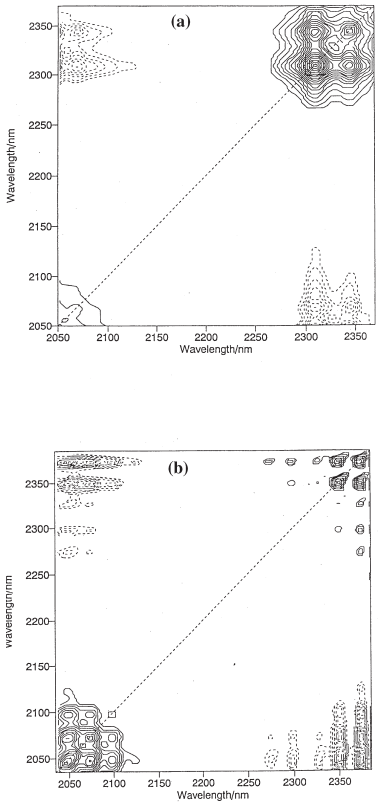

Figure 2. Synchronous 2D NIR correlation spectra in the 2000–2400 nm region constructed from fat (a) and protein (b) concentration-dependent spectral changes of milk after MSC and smoothing. (Reproduced from Reference 21 with permission. Copyright © Society for Applied Spectroscopy).

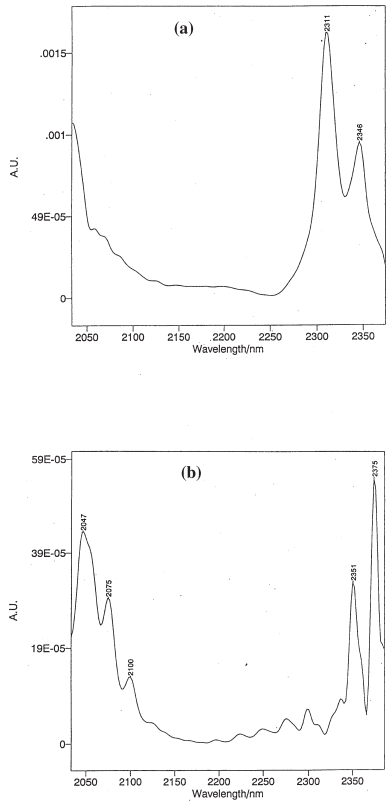

Figure 3. (a) A power spectrum along the diagonal line on the synchronous spectrum shown in Figure 2(a); (b) a power spectrum along the diagonal line on the synchronous spectrum shown in Figure 2(b). (Reproduced from Reference 21 with permission. Copyright © Society for Applied Spectroscopy).

3(b)] are ascribed to proteins. In fact, the power spectra presented in Figures 3(a) and (b) are close to the NIR diffuse reflectance (DR) spectra of fat and protein in solid states, respectively. In this way, generalised 2D correlation spectroscopy can separate out the bands due to fats and proteins from the complicated NIR spectra of milk.

Figure 4(a) shows the asynchronous 2D NIR correlation spectrum in the 2000–2400 nm region, constructed from fat concentration-dependent spectral changes of milk after the MSC and smoothing pretreatment.[21] The asynchronous map shows a number of peaks. The analysis of intensity changes in this map can be simplified by extracting slice spectra. Figure 4(b) shows a slice spectrum at 2311 nm, a

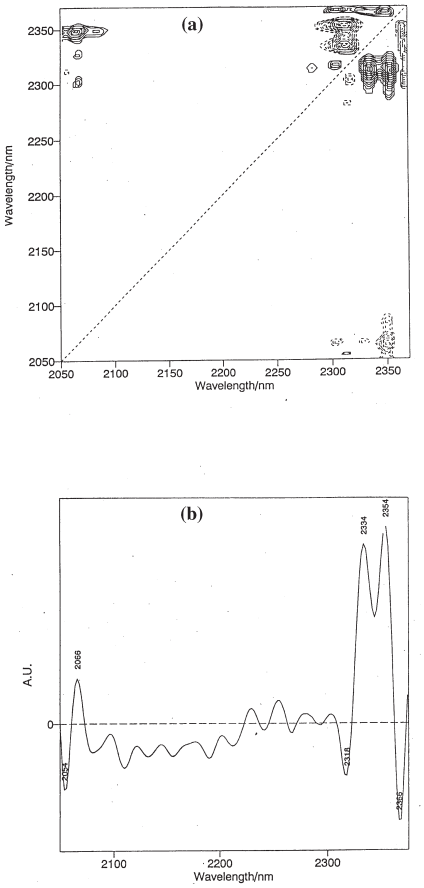

Figure 4. (a) An asynchronous 2D NIR correlation spectrum in the 2000–2400 nm region constructed from fat concentration-dependent spectral changes of milk after the pretreatments; (b) a slice spectrum along 2311 nm line in the asynchronous spectrum shown in Figure 4 (a). (Reproduced from Reference 21 with permission. Copyright © Society for Applied Spectroscopy).

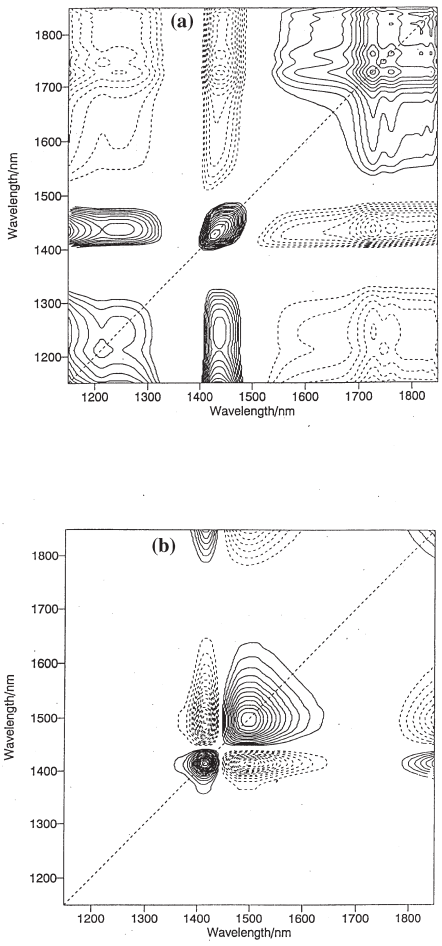

Figure 5. Synchronous 2D NIR correlation spectra in the 1100–1900 nm region constructed from fat (a) and protein (b) concentration-dependent spectral changes of milk after the pretreatments. (Reproduced from Reference 21 with permission. Copyright © Society for Applied Spectroscopy).

characteristic wavelength for fat species.[21] The slice spectrum develops peaks at 2054, 2066, 2318, 2334, 2354 and 2366 nm. It is notable that the peaks at 2054, 2066, 2318, 2334 and 2366 nm are missing in the corresponding synchronous spectrum shown in Figure 2(a). In general, asynchronous correlation spectra have a more powerful deconvolution ability for highly overlapped bands. Due to this fact the asynchronous spectrum correlated with changes in one component (fat or protein) detects peaks corresponding to the characteristic bands of fat or protein. The assignments of bands appearing in the 2D correlation spectra are reported elsewhere in detail.[21]

Figures 5(a) and (b) show synchronous 2D correlation maps in the 1100–1900 nm region, generated from the fat and protein concentration-dependent spectral variations, respectively.[21] The synchronous correlation spectra, constructed from the fat and protein concentration-dependent spectral changes of milk, show marked differences in the 1400–1500 nm region where bands due to the combination modes of the various species of water are expected to appear. The most important difference is that the bands in the 1400–1500 nm region are separated into two in the synchronous spectrum for the protein concentration-dependent spectral variations [Figure 5(b)]. The first essential factor, which may cause the separation of the two water bands, is that proteins are hydrophilic while fats are hydrophobic. Therefore, even if the variation of protein content (1.18%) is much smaller than that of fat content (4.59%), the separation of the bands, due to water, is larger in the synchronous spectrum for the protein concentration-dependent spectral variations [Figure 5(b)] than that for the fat concentration-dependent spectral changes [Figure 5(a)]. It is very likely that the two autopeaks at 1419 and 1485 nm in the former [Figure 5(b)] arise from bulk and hydrated water, respectively.[21]

Comparison between 2D correlation analysis and PCA

Figure 6 shows the first loadings plots in the 2000–2400 nm region for the PCA model, based on the NIR spectra of milk samples.[26] It is noted that the first loadings plot is similar to the power spectrum shown in Figure 3(a). In both cases, bands due to milk fats are enhanced remarkably. Therefore, the synchronous spectrum and the first loadings plot of PCA have a similar ability for selecting out bands due to a particular component. The similarity between the power spectrum and the first loadings is also observed in the 1400–1700 nm region.[26]

It seems quite reasonable that the synchronous 2D correlation spectrum and the first loadings plots for the PCA model are very close to each other because both represent overall spectral variations. Autopeaks in the synchronous spectrum represent the overall extent of perturbation-induced intensity changes observed during the measurement. In PCA, variables with a high degree of systematic variation, typically, have large absolute loadings.

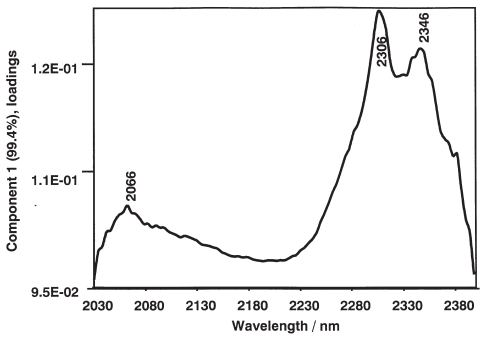

Figure 6. The first loadings plots for the PCA model based upon the NIR spectra of milk.

Recently, we have carried out a systematic investigation into comparisons between generalised 2D correlation spectroscopy and PCA by using NIR spectra of human serum albumin (HSA) aqueous solutions.[27] We have found that there is correlation between the asynchronous spectrum and the second loadings plots.[27] Some slice spectra in the asynchronous spectra very closely resemble the second loadings plot. The second loadings plot is orthogonal to the first loadings plot and the asynchronous spectrum is orthogonal to the synchronous spectrum. Therefore, it is also reasonable that the asynchronous spectrum corresponds well to the second loadings plots.

A more detailed comparison between 2D correlation spectroscopy and PCA for milk is now under investigation and will be reported elsewhere.[26]

Acknowledgements

The authors thank Dr S. Tanabe, Dr T. Hayashi, Dr F. Terada and Dr M. Amari of the National Institute of Animal Industry, Japan, Dr S. Kawano of the National Food Research Institutes, Japan, for the preparation of milk samples, the determination of protein and fat contents in the milk samples and the NIR measurements. This work was supported by the Programme for Promotion of Basic Research Activities for Innovative Biosciences (PROBRAIN).

References

1. I. Murray and P.C. Williams, in *Near-Infrared Technology in the Agricultural and Food Industries,* Ed by P. Williams and K. Norris. American Association of Cereal Chemists, Inc., Minnesota, USA, p. 23 (1987).
2. D.A. Burns and E.W. Cirurczak, *Handbook of Near Infrared Analysis.* Marcel Dekker, New York, USA (1992).
3. B.G. Osborne, T. Fearn and P.H. Hindle, *Practical NIR Spectroscopy with Applications in Food and Beverage Analysis.* Longman Scientific and Technical, Harlow, UK (1993).
4. Y. Ozaki and I. Noda*, J. Near Infrared Spectrosc*. **4,** 85 (1996).
5. Y. Ozaki, Y. Liu and I. Noda, *Appl. Spectrosc*. **51,** 466 (1997).
6. Y. Ozaki, *NIR news* **9(2),** 1 (1998).
7. H. Martens and T. Næs, *Multivariate Calibration*. John Wiley & Sons, Chichester, UK (1989).
8. R. Malinowski, *Factor Analysis in Chemistry,* 2nd Edition. John Wiley & Sons, New York, USA (1991).
9. D.L. Massart, B.G.M. Vandenginste, L.M.C. Buydens, S. De Jong, P.J. Lewi and J. Smeyers-Verbeke, *Handbook of Chemometrics and Qualimetrics*. Elsevier, Amsterdam, The Netherlands (1997).
10 M. Iwahashi, Y. Hayashi, N. Hachiya, H. Matsuzawa and H. Kobayashi, *J. Chem. Soc*. *Faraday Trans.* **89,** 707 (1993).
11. M. Iwahashi, N. Hachiya, Y. Hayashi, H. Matsuzawa, M. Suzuki, Y. Fujimoto and Y. Ozaki, *J. Phys. Chem*. **97,** 3129 (1993).
12. M.A. Czarnecki, Y. Liu, Y. Ozaki, M. Suzsuki and M. Iwahashi, *Appl. Spectrosc*. **47,** 2162 (1993).
13. Y. Liu, H. Maeda, Y. Ozaki, M.A. Czarnecki, M. Suzuki and M. Iwahashi, *Appl. Spectrosc*. **49,** 1661 (1995).
14. J. Wang, M.G. Soma, M.K. Ahmed and H.H. Mantsch*, J. Phys, Chem*. **98,** 4748 (1994).
15. I. Noda, *Appl. Spectrosc*. **47,** 1329 (1993).
16. F.E. Barton, II, D.S. Himmelsbach, J.H. Duckworth and M.J. Smith, *Appl. Spectrosc*. **46,** 420 (1992).
17. I. Noda, Y. Liu, Y. Ozaki and M.A. Czarnecki, *J. Phys. Chem*. **99,** 3068 (1995).
18. F.E. Barton, II, D.S. Himmelsbach and D.D. Archibald*, J. Near Infrared Spectrosc*. **4,** 139 (1996).
19. Y. Wang, K. Murayama, Y. Myojo, R. Tsenkova, N. Hayashi and Y. Ozaki, *J. Phys. Chem. B* **102,** 6655 (1998).
20. Y. Wang, R. Tsenkova, M. Amari, F. Terada, T. Hayashi, A. Abe and Y. Ozaki, *Analusis* **26,** 64 (1998).
21. B. Czarnik-Matusewicz, K. Murayama, R. Tsenkova and Y. Ozaki, *Appl. Spectrosc*. **53,** 1582 (1999).
22. N.L. Sefara, N.P. Magtoto and H.H. Richardson, *Appl. Spectrosc*. **51,** 536 (1997).

23. Y. Liu, Y. Ozaki and I. Noda, *J. Phys. Chem*. **100,** 7326 (1996).
24. C.P. Schultz, H. Fabian and H.H. Mantsch, *Biospectrosc.* **4,** 19 (1998).
25. I. Noda, in *Proceedings of International Symposium on Two-Dimensional Correlation Spectroscopy.* American Institute of Physics, in press.
26. B. Czarnik-Matusewicz, K. Murayama, O.M. Kvalheim, R. Tsenkova and Y. Ozaki, to be submitted.
27. K. Murayama, B. Czarnik-Matusewicz, Y. Wu, R. Tsenkova and Y. Ozaki, *Appl. Spectrosc*., in press