

A New Approach to Near-Infrared Spectral Data Analysis Using Independent Component Analysis

J. Chen and X. Z. Wang*

Department of Chemical Engineering, The University of Leeds, Leeds LS2 9JT, UK

Received December 12, 2000

This paper presents a new approach to near-infrared spectral (NIR) data analysis that is based on independent component analysis (ICA). The main advantage of the new method is that it is able to separate the spectra of the constituent components from the spectra of their mixtures. The separation is a blind operation, since the constituent components of mixtures can be unknown. The ICA based method is therefore particularly useful in identifying the unknown components in a mixture as well as in estimating their concentrations. The approach is introduced by reference to case studies and compared to other techniques for NIR analysis including principal component regression (PCR), multiple linear regression (MLR), and partial least squares (PLS) as well as Fourier and wavelet transforms.

1. INTRODUCTION

Near-infrared (NIR) spectroscopy has received much attention during the past decade. The unrivalled combination of speed, accuracy, and simplicity makes NIR spectroscopy one of the most popular alternatives to wet chemistry procedures for determining concentrations of major classes of chemical compounds. It is now widely used in agricultural, food, and beverage industries.^{1,2} Because the NIR technique is noninvasive and safe, it is also increasingly used in the medical field.³ The widespread use of the NIR technology is also found in pharmaceutical, biomedical, textile, and petrochemical industries.^{4–7}

The NIR spectrum range is 780–2500 nm (12 820–4000 cm^{-1}). Most NIR spectra are the result of overtones or combinations of fundamental vibrations involving hydrogenic stretching modes in the mid-infrared region of the spectrum (4000–600 cm^{-1}).^{1,2,4} For a given molecule, many active overtone and combination bands are typically present in a narrow NIR region and force the peaks to overlap significantly, and other fundamental vibrations manifest in the NIR region only as high-order overtones that are generally too weak to provide useful analytical information. Because of the difficulties in dealing with the highly overlapping peaks and the weak absorption in the NIR region, it had remained unnoticed for almost the next two decades since it was first introduced in the early 1950s. In late 1980s NIR began to emerge as an important technique, mainly due to the development in advanced optical fiber materials and solid detectors as well as the progress in multivariate data analysis.^{4,8} The multivariate techniques used for NIR spectra analysis include multiple linear regression (MLR), principal component regression (PCR), and partial least squares (PLS) regression, as reviewed by Bouveresse and Massart,⁹ Hildrum et al.,⁸ and Williams and Norris.² Fourier,^{10,11} and more recently wavelet transforms^{12–16} have also been studied for NIR spectral data analysis.

NIR spectra of a mixture are often the linear combination of the spectra of its constituent species. It would be very useful if the component spectra can be separated from those of the mixture. This is a blind operation because the pure species in the mixture can be unknown. As will be analyzed in the next section, MLR, PCR, PLS, Fourier, and wavelet transforms are not designed for such a purpose. They cannot be used to identify unknown species in mixtures as well as their concentrations.

The independent component analysis (ICA) method introduced in this paper is able to deliver this function. ICA is a newly developed statistical approach to separating unobserved, independent source variables from the observed variables that are the combinations (or mixtures) of these source variables. Since it was developed in the 1990s,^{17–19} ICA has proved to be a successful technique in biomedical signal processing, e.g., electroencephalographic analysis,^{20,21} magnetic resonance imaging analysis,^{20,22} speech recognition,^{20,23} and machine monitoring.²⁴

The rest of the paper is organized as follows. In the next section, a brief introduction to NIR spectral data analysis will be given, which is followed by a presentation of the concept and algorithm of ICA in section 3. The procedure to NIR data analysis using ICA is introduced in section 4 by reference to two case studies. Conclusions and future work discussions are to be made in section 5.

2. NIR SPECTROSCOPY AND DATA ANALYSIS

2.1. NIR Spectroscopy. In transmission NIR spectroscopy of homogeneous and nonscattering samples, the spectral can be described by absorbance of light.^{1,2,4} The Beer–Lambert Law (or Beer's Law) states that for an absorbing compound dissolved in a nonabsorbing medium, the concentration c of the compound is proportional to the absorbance (A)

$$c \propto A = \log_{10}(I_0/I) \quad (1)$$

where I_0 is the light intensity incident on the medium and I

* Corresponding author phone: +44 113 233 2427; fax: +44 113 233 2405; e-mail: x.z.wang@leeds.ac.uk.

is the light intensity transmitted through the medium. For a medium containing several different absorbing compounds (except at very high concentrations) the overall absorbance A is simply the sum of the contributions of each compound.

In diffuse-reflectance NIR spectroscopy of opaque and nonabsorbing samples, the spectral can be described by reflectance R (or apparent absorbance $A = \log_{10}(1/R)$). Although the Beer–Lambert Law does not hold for this case, empirically the concentration c is still proportional to the apparent absorbance^{1,4}

$$c \propto A = \log_{10}(1/R) \quad (2)$$

According to the aforementioned linear relationship, the concentration c can be solved by modern multivariate regression analysis against the absorbance A or the apparent absorbance $\log_{10}(1/R)$ at different wavelength. The modern multivariate data analysis can also deal with overlapped NIR spectra effectively. Conventional NIR spectral analysis involves two stages of multivariate data analysis, i.e., the calibration and the prediction stages. Calibration establishes a model that correlates the spectral response of each sample at individual wavelengths to known chemical concentrations from laboratory analysis. The prediction makes use of the spectral response of new samples and the calibrated model to estimate unknown chemical concentrations in the samples. Three multivariate regression methods^{2,8,9,25–27} are commonly used for NIR spectral analysis, and they are described briefly next.

2.2. Multiple Linear Regression (MLR).^{2,25,27} According to the Beer's Law, and considering multicomponent concentration matrix \mathbf{C} and their corresponding spectral absorbance matrix \mathbf{A} at different wavelengths, the linear relationship between the concentration \mathbf{C} and the absorbance \mathbf{A} can be described as $\mathbf{C} = \mathbf{AB}$, where \mathbf{B} is the matrix of coefficients. So if \mathbf{C} and \mathbf{A} are known, \mathbf{B} can be solved by multiple linear regression $\mathbf{B} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{C}$.

The MLR can accurately build models for complex mixtures when only some of the constituent concentrations are known. The only requirement is to select the wavelengths that correspond to the absorbances of the desired constituents. However, due to the dimensionality of the matrix equations, the number of selected wavelengths cannot exceed the number of training samples. In theory, it should be possible to increase training samples to allow for additional wavelengths, but this brings a new challenge. The absorbances in a spectrum tend to all increase or decrease as the concentrations of the constituents in the mixture change. This effect, known as collinearity, may lead to an unstable mathematical solution with respect to each constituent. Another problem associated with adding more wavelengths to the model is an adverse effect known as overfitting. In theory, starting with very few wavelengths and adding more to the model (provided they are chosen to reflect the constituents of interest) will improve the accuracy of prediction. However, at some point, the prediction will start to get worse. When too much information in the spectrum is used in calibration, the model starts to include the spectral noise that is unique to the training set and the prediction for unknown samples deteriorates.

2.3. Principal Component Regression (PCR).^{2,8,26–31} The PCR method combines principal component analysis (PCA)

of the spectra with an MLR to devise a quantitative model for complex samples. PCA decomposes the spectra into several principal components, which are linear combinations of the original spectra and can represent the spectral variations that are common to all of the spectroscopic calibration data. Therefore, using the decomposed information to calculate a regression equation instead of the straight spectral responses produces a robust model for predicting concentrations of the desired constituents in very complex samples.

PCA decomposes the original spectral data matrix, \mathbf{A} , into a principal component matrix \mathbf{F} and the corresponding scaling score matrix \mathbf{S} , $\mathbf{A} = \mathbf{SF}$, where the rows of \mathbf{F} are the principal components or the eigenspectra of \mathbf{A} , and the scores in the \mathbf{S} matrix are unique to each calibration spectrum. The \mathbf{S} matrix is supposed to be linearly correlated to the concentration matrix \mathbf{C} , then it is the next step to regress \mathbf{C} against the \mathbf{S} matrix by the MLR model: $\mathbf{C} = \mathbf{SB}$. Therefore it can be solved that $\mathbf{B} = (\mathbf{S}^T\mathbf{S})^{-1}\mathbf{S}^T\mathbf{C}$. For unknown samples, \mathbf{A} , \mathbf{B} , and \mathbf{F} are known and \mathbf{S} is obtained from $\mathbf{S} = \mathbf{AF}^T$, and then the concentration matrix \mathbf{C} can be obtained.

In earlier literature, it was assumed that PCR does not require wavelength selection, and often the whole spectrum was used. This gives an averaging effect and makes the model less susceptible to spectral noise. However, more recently it has been demonstrated in several papers that the elimination of uninformative variables before PCA significantly improve the quality of the results.^{32,33} When collinearity between the original constituent concentrations occurs (although it should also be avoided when selecting them), principal components often allow better interpretation of the variations observed in the data set than those original spectra selected by the single MLR. It should be pointed out though that the principal components merely represent the largest common variations among all the spectra in the training set, and they do not mean the spectra of the pure constituents. As a modeling method, PCA is somewhat less accurate than MLR when both are used within the calibration range and when the model is indeed linear. However it is often more reliable than the MLR if extrapolation is needed.

2.4. Partial Least Squares (PLS) Regression.^{2,8,14,29,30,34,35} PLS is a quantitative spectral decomposition technique that resembles PCR. However it uses the concentration information during the decomposition process, and this causes the spectra containing higher constituent concentrations to be weighted more heavily than those with low concentrations.

The purpose of PLS is to build the linear model between the concentration matrix \mathbf{C} and the spectral matrix \mathbf{A} , $\mathbf{C} = \mathbf{AB}$. In comparison to the MLR method and PCR methods, PLS produces a weight matrix \mathbf{W} for \mathbf{A} such that $\mathbf{S} = \mathbf{AW}$, i.e., the columns of \mathbf{W} are weight vectors for the \mathbf{A} columns producing the corresponding score matrix \mathbf{S} . These weights are computed so that each of them maximizes the covariance between the concentration matrix \mathbf{C} and the spectra matrix \mathbf{A} . \mathbf{C} can then be decomposed as $\mathbf{C} = \mathbf{SF}$, where \mathbf{F} is the loading matrix for \mathbf{C} . Once \mathbf{F} is computed, there is $\mathbf{C} = \mathbf{AB}$, where $\mathbf{B} = \mathbf{WF}$, and the PLS model is established. There are various algorithms for computing the PLS model and readers can refer to Haaland and Thomas,³⁴ Williams and Norris,² and the webpage.³⁵

PLS adopts a single-step decomposition and regression technique and therefore is faster than PCA although it is more abstract and difficult to explain. Generally, for the purpose of prediction, PLS performs equally well as PCR.

2.5. Fourier and Wavelet Analysis. Fourier analysis^{10,11} transforms the spectral data in the wavelength domain into the frequency domain, where the original spectra are approximated to a desired degree of accuracy by sums of periodic sine and cosine functions of increasing frequency. The Fourier approximation of the spectra can be characterized by Fourier coefficients that are linear transformations of the original spectral data. Therefore a spectral model can be established between these coefficients and component concentrations by multiple regression analysis. The most important feature of the Fourier analysis is the reduction of the multicollinearity and the dimension of the original spectra. However, the Fourier coefficients bear no simple relationship to individual features of the spectrum so that it will not be clear what information is being used in calibration.¹

Wavelet analysis^{12–16} has recently been studied in NIR spectra data processing. It proves to be more powerful than Fourier transform in capturing the local features of a spectrum because it can decompose a signal into components that are well localized in both the time and frequency domains, while Fourier transform can only reflect upon the frequency information. Performing a wavelet analysis of a signal yields a vector of wavelet coefficients that are assigned to different frequency bands. Each band expands over the complete wavelength domain and reflects upon a certain frequency range of the signal. By selecting appropriate wavelet coefficients, similarly a spectral model can be established by regression of the coefficients against the concentrations.

Since NIR spectra are the combined result of the spectra of individual components, it would be useful if the component spectra can be separated. Especially in some circumstances where the spectra of the chosen samples are themselves mixtures of the unknown pure constituent spectra, the separation of these unknown spectra would become very important in identifying the unknown pure components. However the foregoing reviewed methods are mainly for calibration and prediction purposes, and none is designed for the purpose of blind source spectra separation. The ICA approach presented in this article is specifically designed for this purpose and is introduced in the following section. Although inverse calibration can be used even if the concentration of various constituents in the calibration sample is unknown, it cannot separate the spectra of the unknown pure species. Inverse calibration can be applied with PCA or PLS, its details can be found in Kalivas and Lang³² and Martens and Naes.³³

3. INDEPENDENT COMPONENT ANALYSIS

This section describes the mathematical background of ICA. Its application to NIR signal analysis will be described in section 4.

3.1. The Concept of Statistical Independence. Independent component analysis is designed for separating blind (unobserved) source signals or variables from the observations of their mixtures. The concept of statistical indepen-

dence needs to be clarified before the ICA algorithm is to be introduced.^{38–40} Denote x_1, x_2, \dots, x_m the random variables with the joint probability density function (*pdf*) of $p(x_1, \dots, x_m)$ and assume that these variables have zero-mean, and then they are said to be mutually statistically independent if the following condition holds

$$p(x_1, \dots, x_m) = p_1(x_1) \cdot p_2(x_2) \cdot \dots \cdot p_m(x_m) \quad (3)$$

where $p_i(x_i)$ ($i = 1, \dots, m$) denotes the marginal *pdf* of x_i , i.e., the *pdf* of x_i when it is considered alone (e.g. $p_1(x_1) = \int \Lambda \int p(x_1, \Lambda, x_m) dx_2 dx_3 \dots dx_m$).

It is worth pointing out that statistical independence is a different concept from uncorrelatedness. The latter implies that $E\{x_i x_j\} = E\{x_i\} \cdot E\{x_j\}$ for $i \neq j$, where $E\{\cdot\}$ stands for the mathematical expectation. The independence has a stricter condition than the uncorrelatedness. If the x_i and x_j are independent, for any functions f_i and f_j , they must have

$$\begin{aligned} E\{f_i(x_i)f_j(x_j)\} &= \int \int f_i(x_i)f_j(x_j)p(x_i,x_j)dx_idx_j \\ &= \int \int f_i(x_i)f_j(x_j)p_i(x_i)p_j(x_j)dx_idx_j \\ &= \int f_i(x_i)p_i(x_i)dx_i \int f_j(x_j)p_j(x_j)dx_j \\ &= E\{f_i(x_i)\}E\{f_j(x_j)\} \end{aligned} \quad (4)$$

and therefore if the variables are independent, they are uncorrelated but not vice versa.

3.2. Blind Source Signal Separation Using ICA. ICA is a useful extension of the principal component analysis. PCA finds principal components that are the linear combinations of the observed variables, while ICA finds independent components that constitute the observed variables. The PCs found in PCA are uncorrelated and orthogonal. However, as it has been stated that the uncorrelatedness is different from the concept of independence and in many applications including near-infrared spectroscopy analysis, independent components will be of more interest.

ICA was originally developed to deal with problems that are closely related to the real world cocktail-party problem.⁴⁰ Imagine two people are speaking simultaneously in a room, and there are two microphones placed in different locations. The microphones can provide two recorded time signals, $x_1(t)$ and $x_2(t)$. Each of these signals can be expressed as a weighted sum of the speech signals $s_1(t)$ and $s_2(t)$ emitted by the two speakers

$$\begin{aligned} x_1(t) &= m_{11}s_1(t) + m_{12}s_2(t) \\ x_2(t) &= m_{21}s_1(t) + m_{22}s_2(t) \end{aligned} \quad (5)$$

where m_{11}, m_{12}, m_{21} , and m_{22} are some parameters that depend on the distances of the microphones from the speakers. The cocktail-party problem is to separate or recover the two individual source signals $s_1(t)$ and $s_2(t)$ from the recorded mixed signals $x_1(t)$ and $x_2(t)$. If the preceding four parameters are known, the linear equations can be solved by traditional methods. However if they are not known, the problem is much more difficult and is therefore sometimes called blind source separation. Suppose $s_1(t)$ and $s_2(t)$ are statistically independent ICA is designed to estimate the four parameters

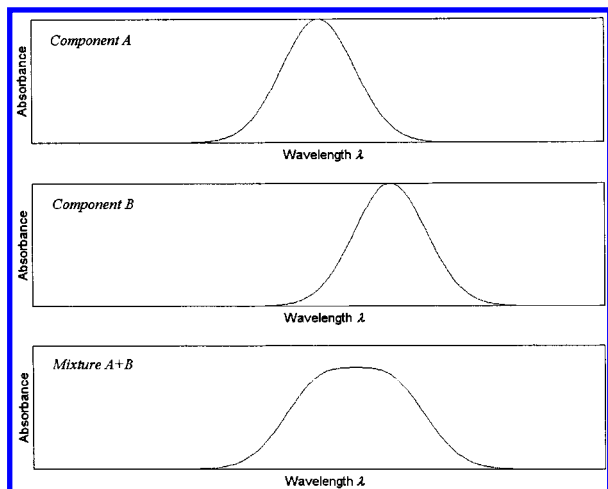


Figure 1. The spectra of *components A and B* and their mixture.

and to separate the two original source signals $s_1(t)$ and $s_2(t)$ from their mixtures $x_1(t)$ and $x_2(t)$.

Clearly NIR spectral analysis represents a very similar problem to the cocktail party. One of the keys to quantitative spectral analysis is the assumption that the concentrations of the components of interest are somehow related to the data from the measurement technique used for analyzing the samples. This relation is well described by the Beer's Law. The law defines a simple linear relationship between the spectrum and the composition of a sample and claims that the absorbances of multiple constituents at the same wavelength are additive. For example, in Figure 1, there are two spectra for *component A* and *component B*, denoted by $s_A(\lambda)$ and $s_B(\lambda)$, the spectra of the mixture of the components *A* and *B* is a linear combination of the spectra $s_A(\lambda)$ and $s_B(\lambda)$. Thus if there are two observed spectra, $x_1(\lambda)$ and $x_2(\lambda)$, of two different mixtures of the components *A* and *B*, there should be the following two equations

$$\begin{aligned} x_1(\lambda) &= c_{11}s_A(\lambda) + c_{12}s_B(\lambda) \\ x_2(\lambda) &= c_{21}s_A(\lambda) + c_{22}s_B(\lambda) \end{aligned} \quad (6)$$

where c_{11} , c_{12} , c_{21} , and c_{22} are the four parameters that are linearly related to the concentrations of the *components A* and *B* in the mixtures. Then similar to the cocktail-party problem, suppose $s_A(\lambda)$ and $s_B(\lambda)$ are statistically independent, ICA is to estimate the four parameters and to separate the spectra of the components *A* and *B*, $s_A(\lambda)$ and $s_B(\lambda)$, from their mixture spectra $x_1(\lambda)$ and $x_2(\lambda)$.

3.3. An ICA Model. Assume that there are l linear mixtures x_1, x_2, \dots, x_l of m independent (source) components s_1, s_2, \dots, s_m , the equations in the last section can be generalized as

$$x_i = m_{i1}s_1 + m_{i2}s_2 + \dots + m_{im}s_m, i = 1, \dots, l \quad (7)$$

Here each mixture x_i and component s_j ($j = 1, \dots, m$) is generalized as a random variable instead of a proper time signal or spectra in the above examples. It is necessary to have at least as many mixtures as independent variables, i.e., $l \geq m$, to perform ICA analysis.

Denote the random column vector $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_l]^T$, $\mathbf{s} = [s_1 \ s_2 \ \dots \ s_m]^T$, and \mathbf{M} the matrix with elements m_{ij} ($i=1, \dots, l; j=1, \dots, m$), the above equations can be written as

$$\mathbf{x} = \mathbf{M}\mathbf{s} \quad (8)$$

which is called an ICA model. For simplicity, in developing the algorithms of ICA, it is assumed that \mathbf{x} is a prewhitened vector so that its components are uncorrelated and their variances equal to unity. The goal of ICA consists of the estimation of the mixing matrix \mathbf{M} and/or the independent source vector \mathbf{s} from the observed mixed vector \mathbf{x} . This goal is equivalent to find a separating matrix \mathbf{W} that satisfies

$$\hat{\mathbf{s}} = \mathbf{W}\mathbf{x} \quad (9)$$

where $\hat{\mathbf{s}}$ is the estimation of \mathbf{s} .

The separating matrix \mathbf{W} can be trained as the weight matrix of a two-layer feed-forward neural network in which \mathbf{x} is input and $\hat{\mathbf{s}}$ is output. The constraint of the neural network is that the elements of $\hat{\mathbf{s}}$ are statistically independent, which can be reflected by nonGaussianity.^{18,41} Thus a measure of nonGaussianity can be used as the objective (contrast) function when training the neural network. There are two common measures of nonGaussianity: one is kurtosis, and the other is negentropy. A brief introduction to these two measures is given below, and more details can be found in Hyvärinen^{38,39} and Hyvärinen and Oja.⁴⁰

The kurtosis of a random variable y is conventionally defined by

$$\text{kurt}(y) = E\{y^4\} - 3(E\{y^2\})^2 \quad (10)$$

For simplicity, assume that the variable y has been preprocessed to have zero-mean and unit variance, and therefore the kurtosis is simplified to $\text{kurt}(y) = E\{y^4\} - 3$. This shows that kurtosis is simply a normalized version of the fourth moment $E\{y^4\}$. For a Gaussian y , the fourth moment equals $3(E\{y^2\})^2$. Thus, kurtosis is zero for a Gaussian random variable. For most (but not quite all) nonGaussian random variables, kurtosis is nonzero. Kurtosis can be either positive or negative. Random variables that have a negative kurtosis are called subGaussian whose *pdf* is flatter than that of Gaussian, and those with positive kurtosis are called super-Gaussian whose *pdf* is shaper than that of Gaussian. The kurtosis measure is a special case of negentropy that is discussed as follows.

Negentropy is based on the information-theoretic quantity of (differential) entropy. The entropy of a random variable can be interpreted as the degree of information that the observation of the variable gives. The more random, i.e., unpredictable and unstructured, the variable is, the larger its entropy. The differential entropy H of a random variable y with density $p(y)$ is defined as $H(y) = -\int p(y) \log p(y) dy$. A fundamental result of information theory is that a Gaussian variable has the largest entropy among all random variables of equal variance. This means that entropy could be used as a measure of nonGaussianity. To obtain a measure of nonGaussianity that is zero for a Gaussian variable and always nonnegative, one often uses a slightly modified version of the definition of differential entropy, negentropy, which is defined as $J(y) = H(y_{\text{gauss}}) - H(y)$, where y_{gauss} is a Gaussian random variable of the same covariance matrix as y . Due to the above-mentioned properties, negentropy is always nonnegative and is zero if and only if y has a Gaussian distribution.

The negentropy J can have the following approximation

$$J(y) \approx k[E\{G(y)\} - E\{G(v)\}]^2 \quad (11)$$

where G is a practically any nonquadratic function, k is an irrelevant constant, and v is a Gaussian variable of zero mean and unit variance. Therefore, to find one independent component, one can maximize the function J_G

$$J_G(\mathbf{w}) = [E\{G(\mathbf{w}^T \mathbf{x})\} - E\{G(v)\}]^2 \quad (12)$$

where \mathbf{w} , an l -dimensional (weight) vector in the weight matrix \mathbf{W} , is constrained so that $E\{(\mathbf{w}^T \mathbf{x})^2\} = 1$. Several independent components can then be estimated one-by-one that will be introduced later.

According to Hyvärinen,^{38,39} the function G has the following choices:

$$G_1(u) = \frac{1}{a_1} \log \cosh(a_1 u) \quad (13)$$

$$G_2(u) = -\frac{1}{a_2} \exp(-a_2 u^2/2) \quad (14)$$

$$G_3(u) = \frac{1}{4} u^4 \quad (15)$$

where $1 \leq a_1 \leq 2$, $a_2 \approx 1$. Among these three choices, G_1 is a good general-purpose function. When the independent components are highly superGaussian, or when robustness is very important, G_2 may be better. G_3 is actually based on the kurtosis and is justified on statistical grounds only for estimating subGaussian independent components when there is no outliers.

Hyvärinen and Oja⁴⁰ have developed a fast fixed-point algorithm (FastICA) that can be described as follows.

(1) Choose an initial random weight vector $\mathbf{w}(0)$ and let $k = 1$.

(2) Let $\mathbf{w}(k) = E\{v g(\mathbf{w}(k-1)^T \mathbf{x})\} - E\{g'(\mathbf{w}(k-1)^T \mathbf{x})\} \mathbf{w}(k-1)$, where g is the first-derivative of the function G . The expectations can be estimated using a large sample of \mathbf{x} vectors by computing over successive points.

(3) Let $\mathbf{w}(k) = \mathbf{w}(k)/\|\mathbf{w}(k)\|$.

(4) If $|\mathbf{w}(k)^T \mathbf{w}(k-1)|$ is not close enough to 1, let $k = k + 1$ and go back to step 2. Otherwise, output the vector $\mathbf{w}(k)$.

The above describes the procedure to estimate just one of the independent components. To estimate several independent components, the procedure should be run as many times as required. To prevent different vectors from converging to the same maxima, the outputs $\mathbf{w}_1^T \mathbf{x}$, ..., $\mathbf{w}_m^T \mathbf{x}$ after the every run of the procedure must be decorrelated. After one has estimated p vectors, \mathbf{w}_1 , ..., \mathbf{w}_p , this just simply needs two more steps for \mathbf{w}_{p+1}

$$\text{let } \mathbf{w}_{p+1} = \mathbf{w}_{p+1} - \sum_{j=1}^p \mathbf{w}_{p+1}^T \mathbf{w}_j \mathbf{w}_j \quad (16)$$

and

$$\text{let } \mathbf{w}_{p+1} = \mathbf{w}_{p+1} / \sqrt{\mathbf{w}_{p+1}^T \mathbf{w}_{p+1}} \quad (17)$$

4. NIR SPECTRAL DATA ANALYSIS USING ICA

In this section, the ICA approach for NIR spectral data analysis will be illustrated by reference to two case studies.

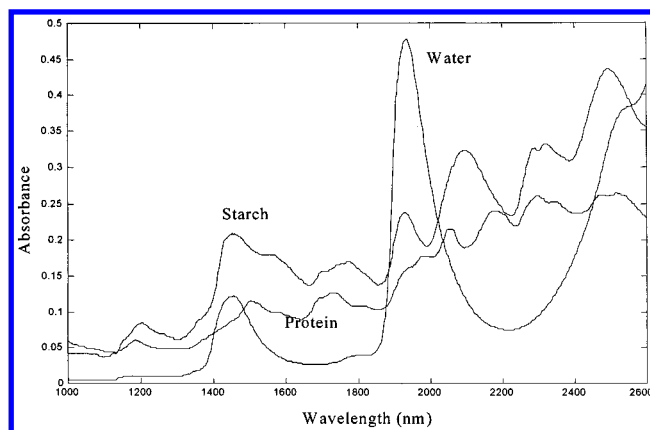


Figure 2. The spectra of water, starch, and protein.

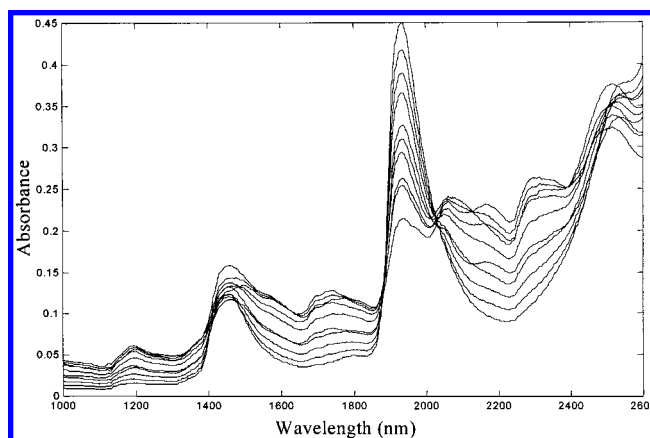


Figure 3. The spectra of 10 samples consisting of water, starch, and protein.

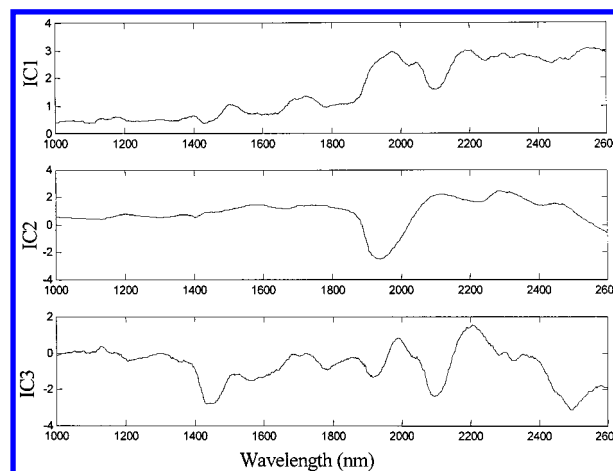


Figure 4. The ICs obtained by directly applying ICA to Figure 3.

In the first case study, all the spectra of mixtures and three pure components, water, starch, and protein, are known. The purpose is to test the method by comparing the separated ICs with the spectra of the pure components.

The second case study is concerned with analysis of the NIR spectra of 200 mixture samples with known concentrations. However, the spectra of the individual components, moisture, fat, and protein, are not available. In ICA analysis, we have studied the separation of the mixture spectra into three, four, and five ICs and found that there is strong correlation between two ICs and the concentrations of moisture and fat. It indicates that the two ICs represent the

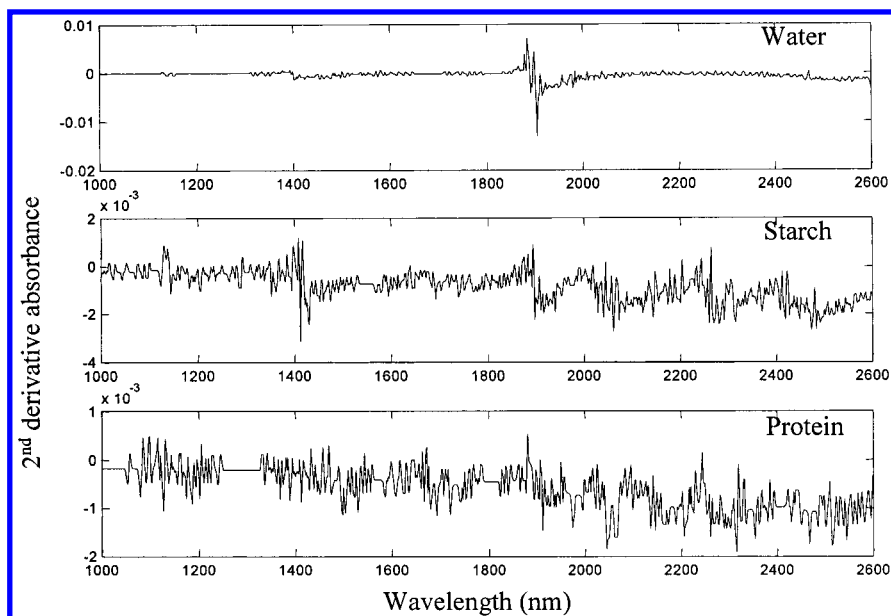


Figure 5. The second-derivative spectra of water, starch, and protein.

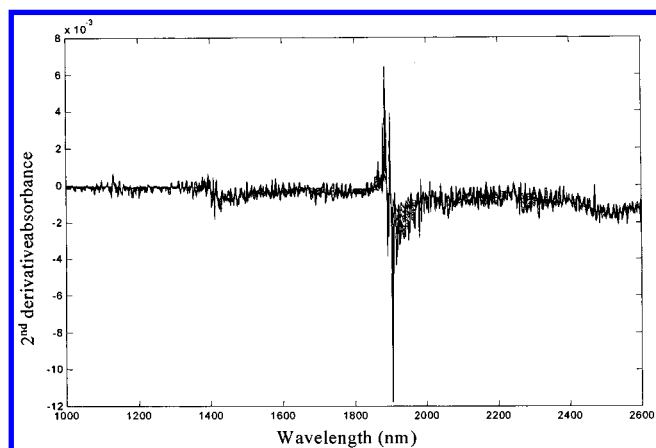


Figure 6. The second-derivative spectra of the 10 samples.

spectra of moisture and fat. The procedure to use ICA as a calibration method is also described.

4.1. Case Study 1. The first case study is about spectral analysis of the mixtures of water, starch, and protein. Though in practical application, the constituent components of mixtures may not be known and the ICA approach is supposed to identify these individual components through blind separation of their spectra, in the case study here, the mixtures of known components are used to illustrate and test the proposed method.

Assume that the NIR spectrum of a mixture is the weighted sum of the spectra of the three constituent components, i.e., water, starch, and protein, and the weights depend on the concentration of each component. The spectral contribution of each constituent in a sample is defined by its near-infrared spectrum multiplied by its concentration in the sample. The spectra of water, starch, and protein are shown in Figure 2.² Figure 3 shows the spectra of 10 different samples to be analyzed, consisting of different concentrations of water, starch, and protein.

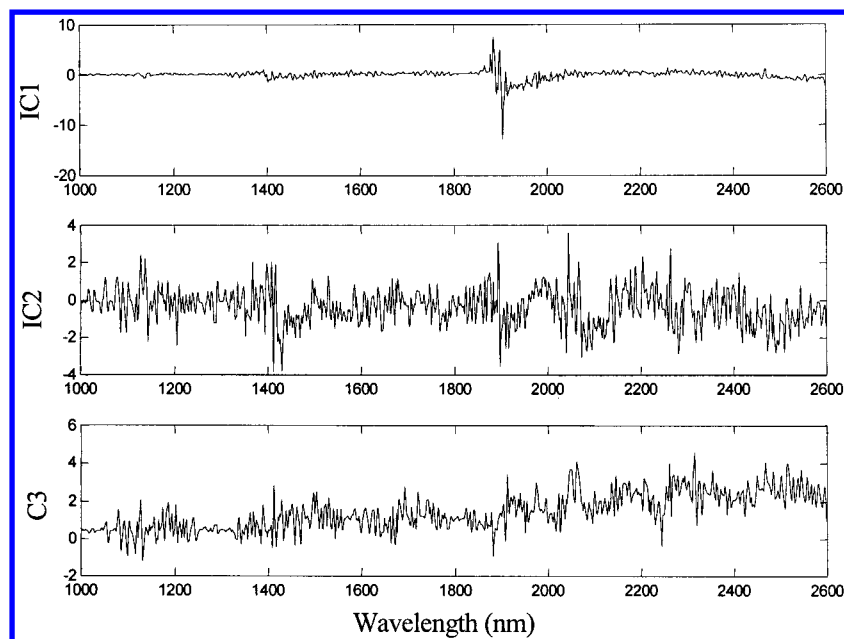


Figure 7. The separated ICs from Figure 6.

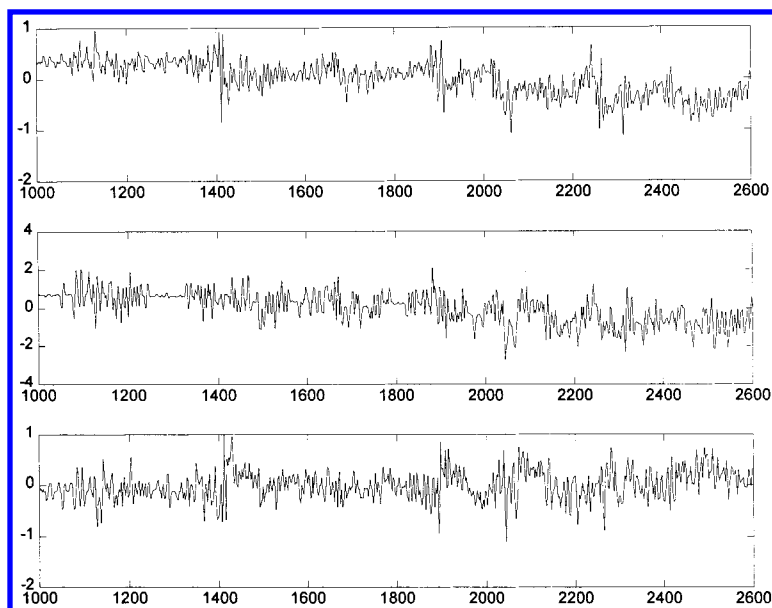


Figure 8. The difference between the spectra of water, starch, and protein (Figure 5) and that of the three ICS (Figure 7).

Applying ICA directly to the data represented by Figure 3 and choosing three ICs give the spectra of the three ICs as shown in Figure 4. Clearly they do not match the spectra of water, starch, and protein shown in Figure 2.

Further examination of the spectra in Figures 2 and 3 reveals that they are all tilted and have baselines (in other words, there are autocorrelations). This could be the reason the separated ICs (spectra) in the preceding paragraph are not as those expected. As a matter of fact, the baseline in NIR spectra is normally nonreproducible and carries no information. Thus the NIR spectra almost always need to be baseline corrected. An effective and conventional approach to carrying out baseline correction is to make use of second derivative spectra.^{42,27,1,2} This approach can not only strip baselines but also help resolve overlapping spectral bands.

The second-derivative spectra of the three constituents and 10 mixture samples are obtained and shown in Figures 5 and 6, and ICA is performed based on these second-derivative spectra. Figure 7 shows the separated ICs. Comparing Figures 7 and 5, it can be clearly seen that the spectra of the three separated ICs correspond to that of water, starch, and protein. The difference between the spectra of water, starch, and protein (Figure 5) and that of the three ICs (Figure 7) is plotted in Figure 8. The spectrum of IC3 resembles the sign-reversed spectrum of protein. This happens in ICA calculation but does not affect the separation result.

4.2. Case Study 2. The data of case study 2, obtained from the literature⁴³ was recorded on a Tecator Infratec Food and Feed Analyzer working in the wavelength range 850–1050 nm by the near-infrared transmission (NIT) principle. Each sample contains finely chopped pure meat with different moisture, fat, and protein contents. For each meat sample, the data consists of a 100-channel spectrum of absorbances. The absorbance is $-\log_{10}$ of the transmittance measured by the spectrometer. The original task was to predict the fat content of a meat sample on the basis of its near-infrared absorbance spectrum. The three contents, moisture (water), fat, and protein, determined by analytic chemistry are

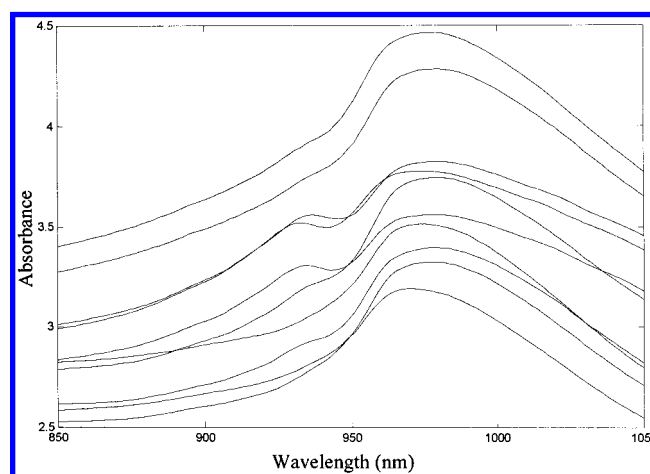


Figure 9. The spectra of 10 samples in case study 2.

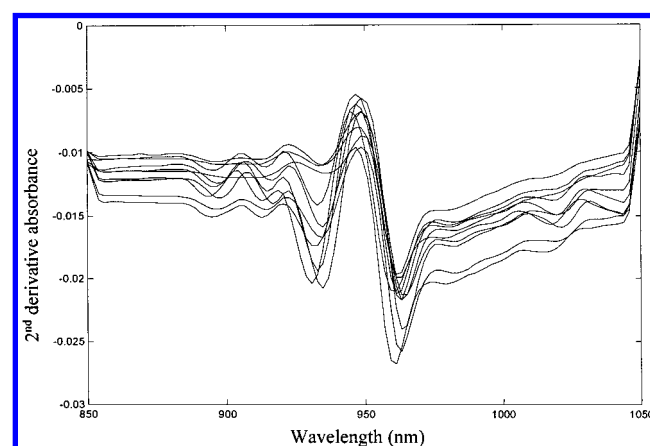


Figure 10. The second-derivative spectra of Figure 9.

available for verification. There are totally 240 samples, however here samples 1–100 will be used for training and samples 101–215 will be used for testing. The spectra of samples 1–10 are given in Figure 9 for illustration. Similar to the first case study, ICA will be based on the second-derivative spectra of the samples. Figure 10 shows the second-derivative spectra of Figure 9. The purpose of this

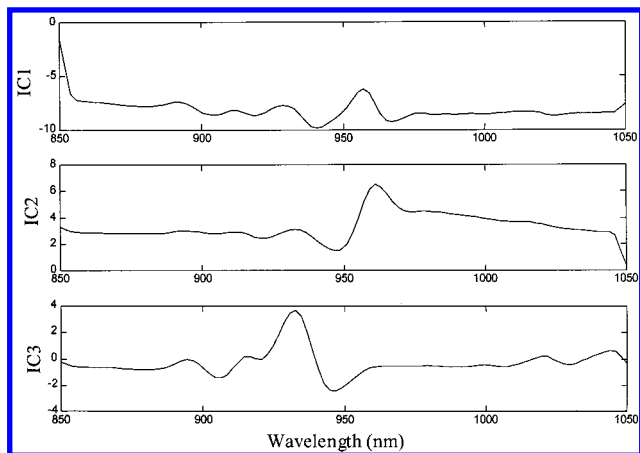


Figure 11. The three separated ICs.

case study is to demonstrate that through analysis of the separated independent components, it is possible to identify the substances in the mixture as well as their concentrations.

In ICA analysis, the n -channel spectrum matrix $\mathbf{A}_{l \times n}$ of l mixture samples is decomposed as

$$\mathbf{A} = \mathbf{M}\mathbf{I} \quad (18)$$

In eq 18, $\mathbf{I}_{m \times n}$ is the matrix containing information about the independent components. In an ideal situation, if the m ICs exactly match the pure substances constituting the mixtures, then $\mathbf{I}_{m \times n}$ is the matrix of the spectral data of the pure substances. $\mathbf{M}_{l \times m}$ is a matrix related to the concentrations of the substances in mixtures. To make an analogy to PCR, \mathbf{M} in ICA is equivalent to the score matrix \mathbf{S} , while \mathbf{I} corresponds to the principal component matrix \mathbf{F} .

In practice, not all the m ICs can match the pure substances exactly, and therefore it cannot be taken for granted that the elements in the matrix \mathbf{M} are concentrations. However like PCR, regression analysis can be performed between the concentration matrix $\mathbf{C}_{l \times t}$ and $\mathbf{M}_{l \times m}$

$$\mathbf{C} = \mathbf{M}\mathbf{B} \quad (19)$$

In eq 19, $\mathbf{B}_{m \times t}$ is a regression coefficient matrix between \mathbf{C} and \mathbf{M} , and t is the number of constituents. In this case, it is assumed that the number of independent components is not known. In practice, different numbers of ICs can be analyzed in order to determine the optimum number.

First, the ICA analysis was carried out under the assumption that there are three independent components. The separation result is shown in Figure 11. In the matrix \mathbf{I} , there are three rows representing vectors for the three ICs, and each first column of \mathbf{M} can be regarded as the weight of IC1, representing the contribution of IC1 to the NIR spectrum of a mixture sample. Similarly, the second and third columns of \mathbf{M} indicate the weight contributions of IC2 and IC3. Since each column of \mathbf{M} is considered as the weight of each IC and according to eq 19, each row of the matrix \mathbf{B} represents the contribution of that IC to a respective constituent concentration. Therefore, the first row of \mathbf{B} corresponds to the contribution of IC1, the second row of IC2, and the third row of IC3.

In addition, in the matrix \mathbf{C} , the first column stands for the concentration of moisture. The second and the third columns indicate the concentrations of fat and protein, respectively. Thus the first column of \mathbf{B} represents the

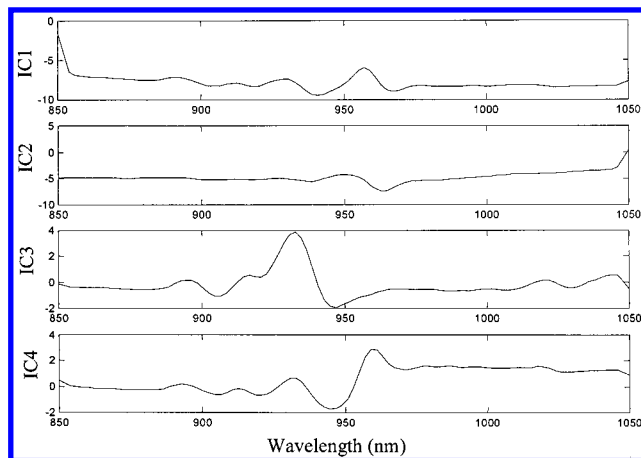


Figure 12. The four separated ICs.

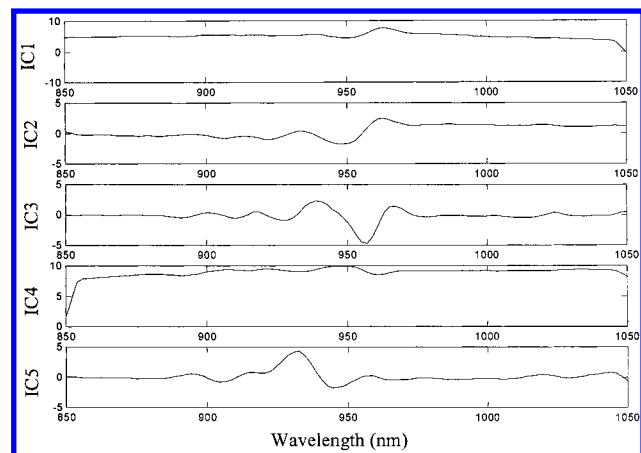


Figure 13. The five separated ICs.

contributions of the different ICs to the concentration of moisture. The second column of \mathbf{B} accounts for the IC contributions to the concentration of fat and the third column of \mathbf{B} to the concentration of protein.

For the three ICs in Figure 11, the regression coefficient matrix \mathbf{B} between the mixing matrix \mathbf{M} of these components and the concentrations of moisture, fat, and protein is

$$\begin{bmatrix} 12328 & 2231 & 7796 \\ -22187 & -3328 & -5537 \\ 17515 & -14160 & 5275 \end{bmatrix}$$

In this matrix, each row represents the contribution of IC1, IC2, and IC3, respectively, and each column indicates the contribution to the concentration of moisture, fat, and protein. Thus it can be seen that the moisture is mainly determined by IC2, and the fat is mainly determined by IC3. Therefore these two ICs could correspond to the second-derivative spectra of moisture and fat, respectively.

Figure 12 shows those separated by ICA under the assumption of four independent components. The regression coefficient matrix \mathbf{B} between the mixing matrix \mathbf{M} of these four components and the concentrations of moisture, fat, and protein is

$$\begin{bmatrix} 10562 & 1714 & 7119 \\ 31580 & 5194 & 10378 \\ 15998 & -14342 & 4322 \\ -1122 & -2370 & 1685 \end{bmatrix}$$

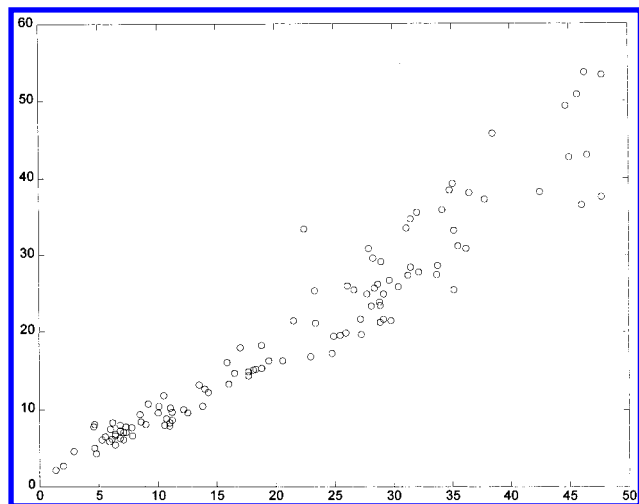


Figure 14. Comparison of the prediction and analytical values for fat (three ICs were used).

From this matrix, it can also be seen that the moisture is mainly determined by the second plot (IC2), and the fat is mainly determined by the third plot (IC3). Thus these two plots could be corresponding to the second-derivative spectra of moisture and fat, respectively. This conclusion is in line with that which has been explained in the last paragraph because these two plots show little difference from those in Figure 11.

If it is assumed that there are five independent components, the regression coefficient matrix **B** between the mixing matrix **M** of the separated five independent components (Figure 13) and the concentrations of moisture, fat, and protein is

$$\begin{bmatrix} -32913 & -4665 & -10855 \\ 1633 & -2597 & 2486 \\ 7524 & -3574 & 1506 \\ -15450 & -1029 & -9337 \\ 13793 & -14120 & 3824 \end{bmatrix}$$

Similarly from this matrix, it can be seen that the moisture is mainly determined by the first plot (IC1), and the fat is mainly determined by the last plot (IC5). These two plots could be corresponding to the second-derivative spectra of moisture and fat, respectively. This conclusion is in line with those which have been explained before because these two plots show little difference from those in Figures 11 and 12.

Figures 14–16 show the predictions of the content of fat using three, four, and five independent components separated in Figures 11–13. The correlation coefficients (*R*) between the predictions and the analytical values of fat are 0.95, 0.96, and 0.97, respectively. The results are made with the test data sets (samples 101–215). Comparing Figures 14–16, it can be seen that five ICs give slightly better results than three and four ICs, but the difference is almost negligible.

5. CONCLUSIONS AND FUTURE WORK

The new method introduced has proved to be an effective approach to NIR spectral data analysis. It is distinctive to other methods such as MIR, PCR, and PLS as well as Fourier and wavelet transforms, because it is aimed at separating the spectra of the independent constituent components of the

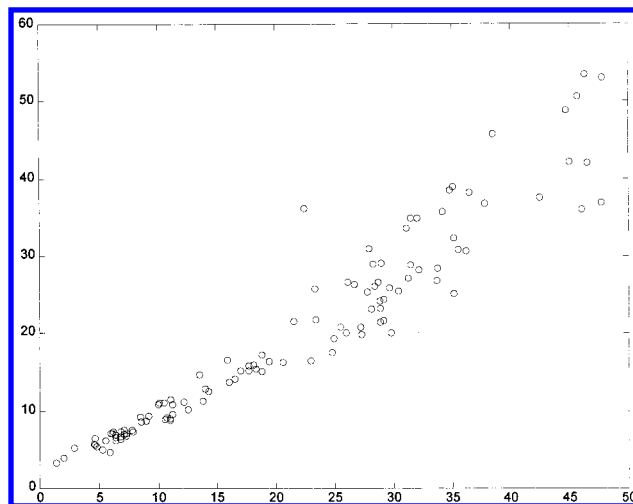


Figure 15. Comparison of the prediction and analytical values for fat (four ICs were used).

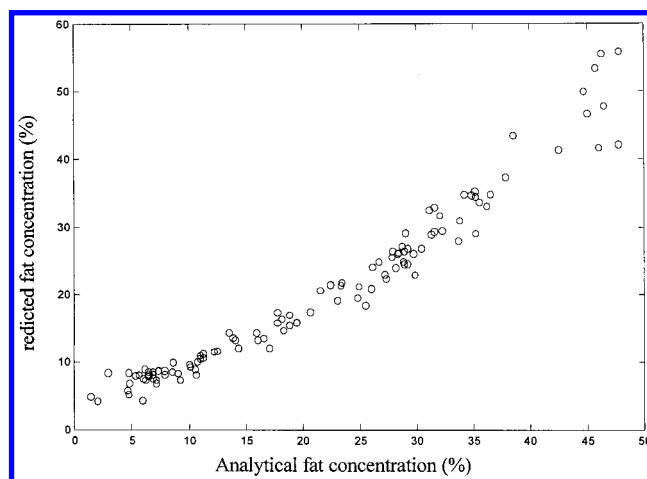


Figure 16. Comparison of the prediction and analytical values for fat (five ICs were used).

mixture as well as determining their concentrations. This is clearly useful when all or some of the components of a mixture are unknown. By comparing the spectra of separated ICs with the spectra library of pure components, it is possible to identify unknown components in the mixture.

Although NIR analysis used to be a laboratory technique, it has recently been found to be potentially useful for online product quality monitoring and control.^{7,44–52} In online applications, speed is clearly a dominant factor. Using the current ICA algorithm which is implemented in Matlab on a Pentium 200, each case study takes about 10 s which clearly can satisfy online requirement. However, the present ICA algorithm is still regarded as computationally intensive and is expected to take much longer for more complex mixtures; therefore, further work needs to be done to improve the ICA method in this aspect.

REFERENCES AND NOTES

- (1) Osborne, B. G.; Fearn, T.; Hindle, P. H. *Practical NIR Spectroscopy with Applications in Food and Beverage Analysis*; Longman Scientific & Technical: 1993.
- (2) *Near-infrared Technology: in the Agricultural and Food Industries*; Williams, P., Norris, K., Eds.; American Association of Cereal Chemists: St. Paul, Minneapolis, 1987.

- (3) Tenhunen, J.; Kopola, H.; Myllylä, R. Noninvasive Glucose Measurement Based on Selective Near Infrared Absorption; Requirements on Instrumentation and Spectral Range. *Measurement* **1998**, *24*, 173–177.
- (4) Chang, S. Y.; Wang, N. S. Monitoring Polymerization Reactions by Near Infrared Spectroscopy. In *Multi-Dimensional NMR, FTIR/Raman, and Fluorescence Spectroscopy of Polymers*; Urban, M. W., Ed.; ACS Symp. Ser., 1995; Vol. 598, pp 147–165.
- (5) Hammond, R. P. Practical and Successful On-line Near Infrared Applications. *Process Contr. Qual.* **1997**, *9*, 117–121.
- (6) *Near-infrared Applications in Biotechnology*; Raghavachari, R.; Ed.; Marcel Dekker: New York, 2000.
- (7) Yeung, K. S. Y.; Hoare, M.; Thornhill, N. F.; Williams, T.; Vaghjani, J. D. Near-infrared Spectroscopy for Bioprocess Monitoring and Control. *Biotechnol. Bioeng.* **1999**, *63*, 684–693.
- (8) Hildrum, K. I.; Isaksson, T.; Naes, T.; Tandberg, A. *Near Infrared Spectroscopy; Bridging the Gap Between Aata Analysis and NIR Applications*; Ellis Horwood: Chichester, 1992.
- (9) Bouveresse, E.; Massart, D. L. Standardization of Near-infrared Spectrometric Instruments: A Review. *Vibrational Spectrosc.* **1996**, *11*, 3–15.
- (10) McClure, W. F.; Hamid, A.; Giesbrecht, F. G.; Weeks, W. W. Fourier Analysis Enhances NIR Diffuse Reflectance Spectroscopy. *Appl. Spectrosc.* **1984**, *38*, 322–329.
- (11) Pasti, L.; Jouan-Rimbaud, D.; Massart, D. L.; deNoord, O. E. Application of Fourier Transform to Multivariate Calibration of Near-infrared Aata. *Anal. Chim. Acta* **1998**, *364*, 253–263.
- (12) Jetter, K.; Depczynski, U.; Molt, K.; Niemöller, A. Principles and Applications of Wavelet Transformation to Chemometrics. *Anal. Chim. Acta* **2000**, *20*, 169–180.
- (13) Jouan-Rimbaud, D.; Walczak, B.; Massart, D. L.; Poppi, R. J.; deNoord, O. E. Application of Wavelet Transform to Extract the Relevant Component from Spectral Data for Multivariate Calibration. *Anal. Chem.* **1997**, *69*, 4317–4323.
- (14) Trygg, J.; Wold, S. PLS Regression on Wavelet Compressed NIR Spectra. *Chemometr. Intell. Lab. Sys.* **1998**, *42*, 209–220.
- (15) Walczak, B.; Bouveresse, E.; Massart, D. L. Standardization of Near-infrared Spectra in the Wavelet Domain. *Chemometr. Intell. Lab. Sys.* **1997**, *36*, 41–51.
- (16) Wold, S.; Sjöström, M. Chemometrics, Present and Future Success. *Chemometr. Intell. Lab. Sys.* **1998**, *44*, 3–14.
- (17) Comon, P. Independent Component Analysis – a New Concept? *Signal Processing* **1994**, *36*, 287–314.
- (18) Delfosse, N.; Loubaton, P. Adaptive Blind Separation of Independent Sources: a Deflation Approach. *Signal Processing* **1995**, *45*, 59–83.
- (19) Jutten, C.; Herault, H. Blind Separation of Sources, Part I: an Adaptive Algorithm Based on Neuromimetic Architecture. *Signal Processing* **1991**, *23*, 1–10.
- (20) Lee, T. W. *Independent Component Analysis: Theory and Applications*; Kluwer Academic Publishers: 1998.
- (21) Vigário, R. N. Extraction of Ocular Artifacts from EEG using Independent Component Analysis. *Electroencephalogr. Clin. Neurophysiol.* **1997**, *103*, 395–404.
- (22) Biswal, B. B.; Ulmer, J. L. Blind Source Separation of Multiple Signal Sources of fMRI Data Sets Using Independent Component Analysis. *J. Comput. Assist. Tomogr.* **1999**, *23*, 265–271.
- (23) Park, H. M.; Jung, H. Y.; Lee, T. W.; Lee, S. Y. Subband-based Blind Signal Separation for Noisy Speech Recognition. *Electronics Lett.* **1999**, *35*, 2011–2012.
- (24) Ypma, A.; Pajunen P. Rotating Machine Vibration Analysis With Second-order Independent Component Analysis. *Proc. Workshop on Independent Component Analysis and Signal Separation (ICA'99)*; Aussois, France, 1999; pp 37–42.
- (25) Brown, C. W.; Lynch, P. F.; Obremski, R. J.; Lavery, D. S. Matrix Representations and Criteria for Selecting Analytical Wavelengths for Multicomponent Spectroscopic Analysis. *Anal. Chem.* **1982**, *54*, 1472–1479.
- (26) Maesschalck, R. D.; Estienne, F.; Verdú-Andrés, J.; Candolfi, A.; Centner, V.; Despagne, F.; Jouan-Rimbaud, D.; Walczak, B.; Massart, D. L.; deJong, S.; deNoord, O. E.; Puel, C.; Vandeginste, B. M. G. The Development of Calibration Models for Spectroscopic Data Using Principal Component Regression. *Internet J. Chem.* **1999**, *2*, 19; <http://www.ijc.com/articles/1999v2/19/>.
- (27) Webpage 1, <http://www.galactic.com/algorithms/default.asp>.
- (28) Fredericks, P. M.; Lee, J. B.; Osborn, P. R.; Swinkels, D. A. J. Materials Characterization Using Factor Analysis of FT-IR Spectra. Part 1: Results. *Appl. Spectrosc.* **1985**, *39*, 303–310.
- (29) Jouan-Rimbaud, D.; Walczak, B.; Massart, D. L.; Last, I. R.; Prebble, K. A. Comparison of Multivariate Methods Based on Latent Vectors and Methods Based on Wavelength Selection for the Analysis of Near-infrared Spectroscopic Data. *Anal. Chim. Acta* **1995**, *304*, 285–295.
- (30) Marbach, R.; Heise, H. M. Calibration Modeling by Partial Least-squares and Principal Component Regression and its Optimization Using an Improved Leverage Correction for Prediction Testing. *Chemometr. Intell. Lab. Sys.* **1990**, *9*, 45–63.
- (31) Naes, T.; Martens, H. Principal Component Regression in NIR Analysis: Viewpoints, Background Details and Selection of Components. *J. Chemometr.* **1988**, *2*, 155–167.
- (32) Kalivas, J. H.; Lang, P. M. *Mathematical Analysis of Spectral Orthogonality*; Marcel Dekker: New York, 1994.
- (33) Martens, H.; Naes, T. *Multivariate Calibration*; Wiley: New York, 1989.
- (34) Haaland, D. M.; Thomas, E. V. Partial Least-squares Methods for Spectral Analyses. 1. Relation to Other Quantitative Calibration Methods and the Extraction of Qualitative Information. *Anal. Chem.* **1988**, *60*, 1193–1202.
- (35) Webpage 2, <http://www.galactic.com/algorithms/pls.htm>.
- (36) Centner, V.; Massart, D. L.; deNoord, O. E.; deJong, S.; Vandeginste, B. M.; Sterna, C. Elimination of Uninformative Variables for Multivariate Calibration. *Anal. Chem.* **1996**, *68*, 3851–3858.
- (37) Jouanrimbaud, D.; Massart, D. L.; Leardi, R.; deNoord, O. E. Genetic Algorithms as a Tool for Wavelength Selection in Multivariate Calibration. *Anal. Chem.* **1995**, *67*, 4295–4301.
- (38) Hyvärinen, A. Fast and Robust Fixed-point Algorithms for Independent Component Analysis. *IEEE Trans. Neural Networks* **1999a**, *10*, 626–634.
- (39) Hyvärinen, A. Survey on Independent Component Analysis. *Neural Computing Surveys* **1999b**, *2*, 94–128.
- (40) Hyvärinen, A.; Oja, E. Independent Component Analysis: Algorithms and Applications. *Neural Networks* **2000**, *13*, 411–430.
- (41) Hyvärinen, A.; Oja, E. A Fast Fixed-point Algorithm for Independent Component Analysis. *Neural Computation* **1997**, *9*, 1483–1492.
- (42) Adar, F.; Geiger, R.; Noonan, J. Ramen Spectroscopy for Process/quality Control. *Appl. Spectrosc. Rev.* **1997**, *32*, 45–101.
- (43) Webpage 3, <http://lib.stat.cmu.edu/datasets/tecator>.
- (44) Brink, M. V. D.; Herk, A. M. V.; German, A. L. On-line Monitoring and Control of the Solution Polymerization of *n*-butyl Acrylate in Dioxane by Raman Spectroscopy. *Process Contr. Qual.* **1999**, *11*, 265–275.
- (45) DeThomas, F. A.; Hall, J. W.; Monfre, S. L. Real-time Monitoring of Polyurethane Production Using Near-infrared Spectroscopy. *Talanta* **1994**, *41*, 425–431.
- (46) Everall, N.; King, B. Raman Spectroscopy for Polymer Characterization in an Industrial Environment. *Macromol. Symp.* **1999**, *141*, 103–116.
- (47) González, F.; Pous, R. Quality Control in Manufacturing Process by Near Infrared Spectroscopy. *J. Pharmaceutical Biomed. Analysis* **1995**, *13*, 419–423.
- (48) Hailey, P. A.; Doherty, P.; Tapsell, P.; Oliver, T.; Aldridge, P. K. Automated System for the On-line Monitoring of Powder Blending Processes Using Near-infrared Spectroscopy Part I. System Development and Control. *J. Pharmaceutical Biomed. Analysis* **1996**, *14*, 551–559.
- (49) Iwanski, P.; Giardino, R.; Janis, B. Dewaxing Process Control Using Near-infrared Spectroscopy. *Ind. Eng. Chem. Res.* **1999**, *38*, 571–574.
- (50) Rohe, T.; Becker, W.; Kölle, S.; Eisenreich, N.; Eyerer, P. Near Infrared (NIR) Spectroscopy for In-line Monitoring of Polymer Extrusion Processes. *Talanta* **1999**, *50*, 283–290.
- (51) Wu, C.; Danielson, J. D. S.; Callis, J. B.; Eaton, M.; Ricker, N. L. Remote In-line Monitoring of Emulsion Polymerization of Styrene by Short-wavelength Near-infrared Spectroscopy, Part 1: Performance During Normal Runs. *Process Contr. Qual.* **1996**, *8*, 1–23.
- (52) a_1, a_2 , constant coefficients; $a_{11}, a_{12}, a_{21}, a_{22}$, weighted coefficients of the two speech signals; A , absorbance; \mathbf{A} , absorbance matrix; \mathbf{B} , regression coefficient matrix; c , concentration; $c_{11}, c_{12}, c_{21}, c_{22}$, weighted coefficients of the spectra of Component A and Component B; \mathbf{C} , concentration matrix; E , mathematical expectation; \mathbf{F} , principal component matrix; g , first derivative of function G ; G , nonquadratic function; H , differential entropy; i , subscript; I , light intensity transmitted through the media; separated independent component matrix; I_0 , light intensity incident on the media; j , subscript; J , negentropy; J_G , contrast function; k , constant coefficient; iteration step l , the number of mixed observed variables; m , the number of independent components; \mathbf{M} , mixing matrix; n , the number of spectral channels; p, p_i , marginal pdf; R , reflectance; correlation coefficient; $s_1(t), s_2(t)$, two original speech signals; $s_A(\lambda), s_B(\lambda)$, NIR spectra of Component A and Component B; \mathbf{s} , vector of m independent variables; \mathbf{S} , principal component score matrix; t , time; the number of constituents; u , independent variable in function G ; \mathbf{w} , l -dimensional vector in the weight matrix \mathbf{W} ; \mathbf{W} , weighting matrix in PLS; separating matrix in ICA; \mathbf{x} , vector of l observed mixed variables; x_1, \dots, x_m , random variables; $x_1(t), x_2(t)$, two mixed speech signals; y , random variable; λ , wavelength; ν , Gaussian variable of zero mean and unit variance.