Available online at www.sciencedirect.com
ScienceDirect
ISPRS Journal of Photogrammetry & Remote Sensing 62 (2007) 2–12
PHOTOGRAMMETRY & REMOTE SENSING
www.elsevier.com/locate/isprsjprs

Prediction of sweetness and amino acid content in soybean crops from hyperspectral imagery

Sildomar Takahashi Monteiro [a,*], Yohei Minekawa [a], Yukio Kosugi [a], Tsuneya Akazawa [b], Kunio Oda [c]

[a] Department of Mechano-Micro Engineering, Tokyo Institute of Technology, 4259 Nagatsuta, Midori-ku, Yokohama, 226-8502, Japan
[b] University Farm, Faculty of Agriculture, Yamagata University, 5-3 Takasaka, Tsuruoka-shi, Yamagata, 997-0369, Japan
[c] Yamagata General Agricultural Research Center, 25, Yamanomae, Fujishima, Tsuruoka-shi, Yamagata, 999-7601, Japan

Received 22 May 2006; received in revised form 6 December 2006; accepted 6 December 2006
Available online 23 January 2007

Abstract

Hyperspectral image data provides a powerful tool for non-destructive crop analysis. This paper investigates a hyperspectral image data-processing method to predict the sweetness and amino acid content of soybean crops. Regression models based on artificial neural networks were developed in order to calculate the level of sucrose, glucose, fructose, and nitrogen concentrations, which can be related to the sweetness and amino acid content of vegetables. A performance analysis was conducted comparing regression models obtained using different preprocessing methods, namely, raw reflectance, second derivative, and principal components analysis. This method is demonstrated using high-resolution hyperspectral data of wavelengths ranging from the visible to the near infrared acquired from an experimental field of green vegetable soybeans. The best predictions were achieved using a nonlinear regression model of the second derivative transformed dataset. Glucose could be predicted with greater accuracy, followed by sucrose, fructose and nitrogen. The proposed method provides the possibility to provide relatively accurate maps predicting the chemical content of soybean crop fields.

Keywords: Agriculture; Hyperspectral image; Modeling; Neural networks; Spatial prediction

1. Introduction

Green vegetable soybeans (or *edamame*, in Japanese) are a popular crop throughout East Asia and Japan. They differ from ordinary green vegetable soybeans by their larger seed, and sweet and nutty flavor (Liu, 1997). Studies have demonstrated that the taste of vegetables can be attributed to, among other factors, their sweetness and amino acid contents (Toko, 1998). Sweetness can be estimated according to sucrose, glucose and fructose concentrations. Amino acids taste sweet or delicious (*umami*) to humans. The concentration of nitrogen provides an estimation of the amino acid content, as well as the plant stress status. The non-destructive and timely prediction of sucrose, glucose, fructose and nitrogen content would therefore provide farmers and producers with useful information, facilitating the selection of which crops to harvest and determining the appropriate harvest time.

Spectroscopy in the near infrared region has been used for performing analytical measurements in a

* Corresponding author. Tel.: +81 45 924 5484; fax: +81 45 924 5172.
E-mail address: monteiro@pms.titech.ac.jp (S.T. Monteiro).

0924-2716/$ - see front matter © 2006 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.isprsjprs.2006.12.002

variety of applications, providing both qualitative and quantitative results non-destructively for solids and liquids (Koehler et al., 2002). However, conventional methods usually rely on fiber-optic spectrometers that utilize only a few wavelengths in order to perform calculations of chemical contents in vegetables (e.g., Slaughter et al., 1996; Zude, 2003). Calibration methods are commonly based on multivariate mathematics that result in a calibration equation, which is a linear combination of spectral data (Osborne and Fearn, 1986). Furthermore, the spectral measurement methods that have been developed thus far need to be performed destructively on a fruit-by-fruit basis (Tsuta et al., 2002). The prediction of sweetness and nitrogen content in green vegetable soybean crops has not been addressed by these conventional methods.

Hyperspectral imagery has advantages over conventional spectroscopy such as providing simultaneous data acquisition over a large sample area (Richards and Jia, 1999). Recent advances in both sensor and lens technologies for hyperspectral imaging have improved their spatial and spectral resolution, while reducing the size and cost of the equipment. Hyperspectral sensors have been utilized primarily for land cover classification (Cihlar, 2000). Despite much research, applications for material estimation are still challenging, mainly due to the limitations of conventional airborne devices for hyperspectral data acquisition that make the analysis of the "pure spectra" of vegetables difficult (Schowengerdt, 1997).

This paper investigates the potential of high-resolution hyperspectral imagery, in the visible to near infrared region, to predict the sweetness and amino acid content of green vegetable soybeans. The objective is to forecast the concentration of sucrose, glucose, fructose and nitrogen in soybean crops by using hyperspectral data acquired locally without physical contact with the plants, thus not interfering with their growth process. An artificial neural network (ANN)-based method to process the hyperspectral data is proposed. We conducted a performance analysis of regression models using datasets preprocessed by different transformation methods, namely, raw reflectance, second derivative and principal component analysis.

2. Materials and methods

The problem of predicting a dependent variable (e.g., sucrose, glucose, fructose and nitrogen concentrations) by combining a number of independent variables (reflectance spectral bands) can be defined in statistics as a regression problem (Fukunaga, 1990). In spectroscopy, this problem can also be referred to as a calibration problem. Spectrum calibration involves searching for predictive relationships between spectral data and reference values, using either laboratory standards, field standards, or modeling (Siesler et al., 2002). ANNs have been widely applied as regression models (Bishop, 1995), which are one of the main tools used in spectrum calibration.

Fig. 1 shows the data analysis scheme proposed in this study. The hyperspectral image data was initially preprocessed and prepared to serve as input data for the regression analysis. Only vegetated areas were used in the calculations and the remainder of the study area was masked out. ANN-based regression models were obtained using datasets preprocessed by different methods of hyperspectral data transformation and reduction, providing a basis for the comparative performance analysis. The transformation algorithms that were assessed are derivative analysis and principal components analysis (PCA).

Along with the hyperspectral data, samples of green vegetable soybeans from within the study area were collected in order to provide an independent measurement of their chemical contents to serve as "ground

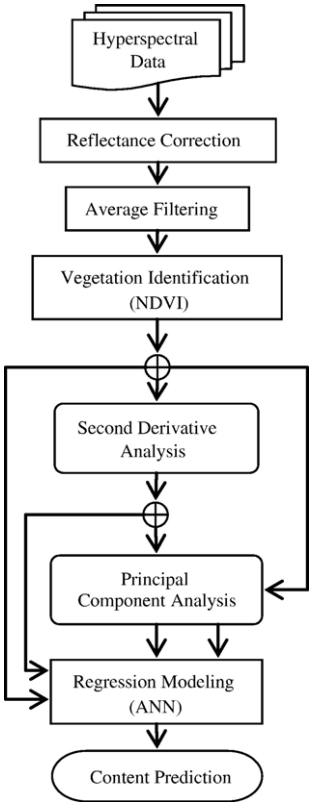

Fig. 1. Diagram of the ANN-based hyperspectral data processing.

reference" data. The concentrations of sucrose, glucose, fructose and nitrogen were measured in the laboratory using liquid chromatography analysis of freeze-dried soybean samples (Ardrey, 2003, chap. 2). The values obtained by these procedures were used as target data for the supervised training of the ANN regression models.

2.1. Hyperspectral image data

The sample data were acquired using a hyperspectral line sensor ImSpector. The hyperspectral line sensor was coupled with a CCD camera and a computer controller that were mounted on the tip of a crane. By means of rotating the crane arm, the camera could scan over the crop field to produce hyperspectral images. Minekawa et al. (2004) demonstrated that this setup allows data acquisition at a spatial resolution suitable for the analysis of agricultural data from localized crop fields. The crane-based system presents several advantages compared to satellite or airplane-based systems, e.g., higher data accuracy and spatial resolution, and reduced atmospheric effects. The hyperspectral data comprised the visible to the near-infrared range of the spectrum from 400 nm to 1000 nm with a 5-nm resolution, thus producing 121 bands of spectral information. In the spatial plane, the hyperspectral camera acquired 484 pixels in each line, which, given the distance to the target, generated a high-resolution image of the crop field. The swath width was about 10 m and the pixel size was approximately 20 mm.

Assume that the hyperspectral data matrix I is composed of N spectral images $I(\lambda)$, ($\lambda = 1,...,N$). Each image pixel (i, j) of a given spectral band ë is considered an observation.

The hyperspectral imaging device captures the radiance from the sample (Spectral Imaging Ltd., 2003). Conversion from raw radiance to reflectance was achieved by acquiring data from a white board used as reference in the field being measured.

In order to reduce or attenuate the intrinsic noise of the hyperspectral dataset, a special three-dimensional filter was designed to simultaneously provide a mean value of the combined spectral and spatial dimensions (Du et al., 2005). The 3D mean filter increases signal-to-noise level by taking advantage of the high spectral resolution and band correlation provided by the sensor, in contrast with traditional 2D mean filters that use only spatial information, although the spatial resolution is somewhat reduced after the smoothing process in both cases. The filter was applied as a 3D cubic window in which each pixel was assigned the mean value of the reflectance of its surrounding pixels, including neighboring spectral bands. This filter can be formulated as

$$y[n_1, n_2, n_3] = \frac{1}{D^3} \sum_{i=0}^{D-1} \sum_{j=0}^{D-1} \sum_{\lambda=0}^{D-1} I[n_1 - i, n_2 - j, n_3 - \lambda], \quad (1)$$

where $[n_1, n_2, n_3]$ refer to the current window position at the hyperspectral dataset, corresponding to [*row*, *column*, *band*]; D is the window size; and I is the pixel's reflectance value.

After the filtering process, image regions containing vegetation were identified using a normalized difference vegetation index (NDVI) (Rouse et al., 1973). The portions of the hyperspectral images corresponding to the different varieties of soybeans (i.e. labeled regions) within the crop-field were then separated manually. The typical spectral signature of the soybean crop is displayed in Fig. 2. Finally, the labeled regions' reflectance data were normalized to the interval [−1, 1] then used as input vectors, constituting, along with the corresponding ground reference data used as target vectors, the training data for the ANNs.

2.2. Derivative analysis

The derivatives provide a measure of the slope of the spectral curve at every point allowing for the resolution of overlapping peaks and correction of baseline effects (Hruschka, 2001). Derivative analysis of reflectance spectra has been used in hyperspectral remote sensing and in analytical chemistry to increase the estimation accuracy of target information (Petisco et al., 2005). Derivatives are relatively less susceptible to variations in illumination intensity in remote sensing applications. Nevertheless, a secondary effect of the derivative process is that, after each successive derivative is performed, the signal-to-noise ratio decreases and sometimes the noise present in the data might be intensified. Thus, there is the necessity of inserting a carefully designed pre-processing phase to reduce this noise.

Myneni et al. (1995) analyzed lower-order spectral derivatives as a generalization of vegetation indexes and demonstrated that, in the case of optically dense vegetation, they are indicative of the abundance and activity of the absorbers in the plant's leaves. Additionally, the second derivative of the near-infrared region has been reported to have a high correlation with the sugar content in melons (Tsuta et al., 2002). In the preliminary investigation on the first and second derivatives used as input vectors for the ANNs (Monteiro et al., 2006), the second derivative produced slightly more accurate predictions with a correlation increase of roughly 3%. Since higher-order derivatives would be affected by noise that

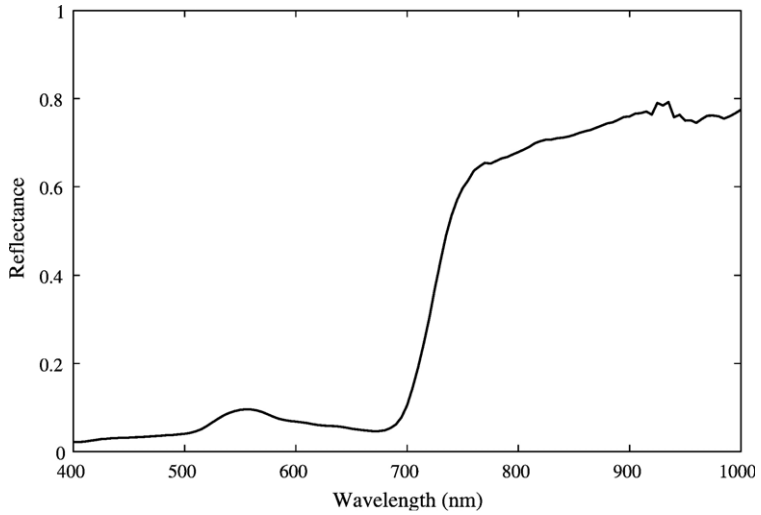

Fig. 2. Spectral profile of the soybean field, mean values calculated from a 10×10-pixel window for each wavelength band.

could jeopardize the ANN training, it was decided to focus the analysis on the second derivative.

The derivative of hyperspectral data can be calculated by finite approximation using suitable difference schemes in accordance with a definite band resolution (Tsai and Philpot, 1998). Given a finite separation between adjacent bands $\Delta\lambda$, the second derivative at wavelength λ_v can be estimated as

$$\left.\frac{\mathrm{d}^2 I}{\mathrm{d}\lambda^2}\right|_v = \left.\frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\frac{\mathrm{d}I}{\mathrm{d}\lambda}\right)\right|_v \approx \frac{I(\lambda_u) - 2I(\lambda_v) + I(\lambda_w)}{(\Delta\lambda)^2}, \qquad (2)$$

where the constraints for the calculation are $\Delta\lambda = \lambda_w - \lambda_v = \lambda_v - \lambda_u$ and $\lambda_w > \lambda_v > \lambda_u$. The derivation interval $\Delta\lambda$ was determined experimentally to be about 20 nm.

2.3. Principal components analysis

Principal components analysis (PCA) is a widely used technique to reduce the dimension of hyperspectral datasets. PCA is used to identify and extract interesting features from the input dataset by retaining only those features that account for a greater part of the dataset variation (Jolliffe, 1988).

PCA was implemented by performing singular value decomposition on the co-variance matrix of the data. The principal components were ordered according to the magnitude of their variance. A threshold parameter was utilized to discard those components that provide only small contributions to the total variance in the dataset. Particularly on hyperspectral datasets of agricultural crop fields, the size of the dataset can be drastically reduced, thus demonstrating the highly correlated nature of this kind of dataset. The objective was to evaluate the ability of the PCA-reduced dataset in providing a prediction of chemical contents when applied as input for the regression model.

2.4. Artificial neural networks

One of the most common kinds of ANN is the multilayer perceptron (MLP) network. The basic MLP network architecture is composed of the input layer, one hidden layer, and the output layer, sequentially interconnected in a feed-forward way (Haykin, 1999). The output of the MLP can be expressed mathematically as

$$\boldsymbol{y} = f(\boldsymbol{x}) = B\varphi(\mathbf{A}\boldsymbol{x} + \boldsymbol{a}) + \boldsymbol{b}, \qquad (3)$$

where $\boldsymbol{x}$ is the input vector; $\boldsymbol{y}$ is the output vector; A and $\boldsymbol{a}$ are, respectively, the weight matrix and the bias vector of the hidden layer; B and $\boldsymbol{b}$ are, respectively, the weight matrix and the bias vector of the output layer; and φ is the activation function. The hyperbolic tangent sigmoid was used as an activation function for the neurons in the hidden layer.

The training method implemented was the Levenberg-Marquardt backpropagation, which, experimentally, has superior performance for regression problems (Hagan and Menhaj, 1994). Early stopping was used to prevent overfitting and to improve the generalization. The hyperspectral dataset was split into a new training set, a validation set and a test set. Training was stopped when one of three conditions occurred: the validation error rate started to grow, the maximum value of the damping factor λ was exceeded or the maximum number of epochs was reached.

Table 1
Parameters for the training of the MLP networks

Parameter	Value
Neurons in the hidden layer	10
Learning rate	0.01
Momentum	0.9
Initial λ [a]	0.001
λ Decrease factor	0.1
λ Increase factor	10
Maximum λ	1.0E+10
Maximum validation failures	5
Maximum number of epochs	1000

[a] Here λ is a scalar utilized by the training algorithm, not the wavelength.

Funahashi (1989) and Cybenko (1989) proved that an ANN with one hidden layer is capable of approximating any mapping to arbitrary accuracy as long as there is a sufficient number of hidden units, although considering more hidden layers can be useful in some cases (Bishop, 1995). The use of more than one hidden layer was reported by Kavzoglu and Mather (2003) to provide no significant improvement in the performance of classifiers and many empirical methods have been proposed to determine the optimal number of neurons. If the ANN has too many degrees of freedom, the generalization ability of the model may be compromised. The experimental approach used in this paper searched for the minimum size of the hidden layer, while also retaining performance. Table 1 shows a summary of the parameters utilized for the training of the MLP networks throughout this paper.

Despite the powerful modeling capability of the MLP, the relationship between the intricate problem at hand and how the network solves the problem is very difficult to understand. In order to obtain some insight about the underlying mechanism of the ANN regression model, a simpler architecture was also investigated, namely, a single-layer perceptron (SLP) network. The SLP network produces finite values as output enabling the construction of a linear regression model. The SLP output can be written as

$$\boldsymbol{y} = f(\boldsymbol{x}) = \boldsymbol{W}\boldsymbol{x} + b, \tag{4}$$

where, again, $\boldsymbol{x}$ and $\boldsymbol{y}$ are the input and output vectors, respectively; $\boldsymbol{W}$ is the weight vector; and the parameter b is the bias.

Considering the simplicity of the SLP architecture, the networks' weight values may be analyzed in order to indicate how each input node, corresponding to a wavelength band, contributes to the formation of the output, i.e. a chemical content prediction. The training method implemented for the SLP was the Widrow-Hoff learning algorithm that is based on an approximate steepest descent procedure, also known as least means squares algorithm (Widrow and Winter, 1988).

Table 2
Sizes of the experimental datasets

Dataset	Number of dimensions
Whole	121
Derivative	113
PCA whole	3
PCA derivative	23

The output layer, in both MLP and SLP architectures, was composed of only one neuron utilizing a linear transfer function. The number of neurons in the input layer was equal to the number of spectral bands or the number of components of the reduced dataset, after preprocessing by the second derivative and PCA. The ANNs were trained to minimize the mean squared error.

2.5. Performance evaluation metrics

The performance of the regression models built using ANNs was evaluated using two metrics, following the work of Murphy and Daan (1985).

Table 3
Performance of the MLP regression models

Chemical [a]	Dataset	Training [b]				Best models	
		μ(MSE)	σ(MSE)	$\mu(R)$	$\sigma(R)$	MSE	R
Glucose	Deriv.	0.122	0.047	0.766	0.118	0.039	0.934
	Whole	0.093	0.016	0.832	0.032	0.062	0.893
	PCA deriv.	0.517	0.095	0.695	0.071	0.337	0.815
	PCA whole	0.601	0.031	0.632	0.025	0.548	0.672
Sucrose	Deriv.	0.108	0.030	0.771	0.078	0.055	0.901
	Whole	0.104	0.017	0.782	0.041	0.075	0.849
	PCA deriv.	0.488	0.056	0.717	0.039	0.355	0.804
	PCA whole	0.572	0.010	0.654	0.007	0.550	0.671
Fructose	Deriv.	0.247	0.052	0.649	0.102	0.113	0.859
	Whole	0.232	0.028	0.678	0.051	0.181	0.772
	PCA deriv.	0.640	0.071	0.600	0.061	0.492	0.714
	PCA whole	0.752	0.036	0.496	0.037	0.692	0.554
Nitrogen	Deriv.	0.291	0.053	0.641	0.082	0.158	0.825
	Whole	0.270	0.027	0.670	0.040	0.224	0.736
	PCA deriv.	0.711	0.072	0.538	0.071	0.560	0.668
	PCA whole	0.728	0.011	0.522	0.010	0.708	0.541

[a] The results were sorted in descending order according to the best models' R-value.
[b] μ refers to the mean and σ refers to the standard deviation.

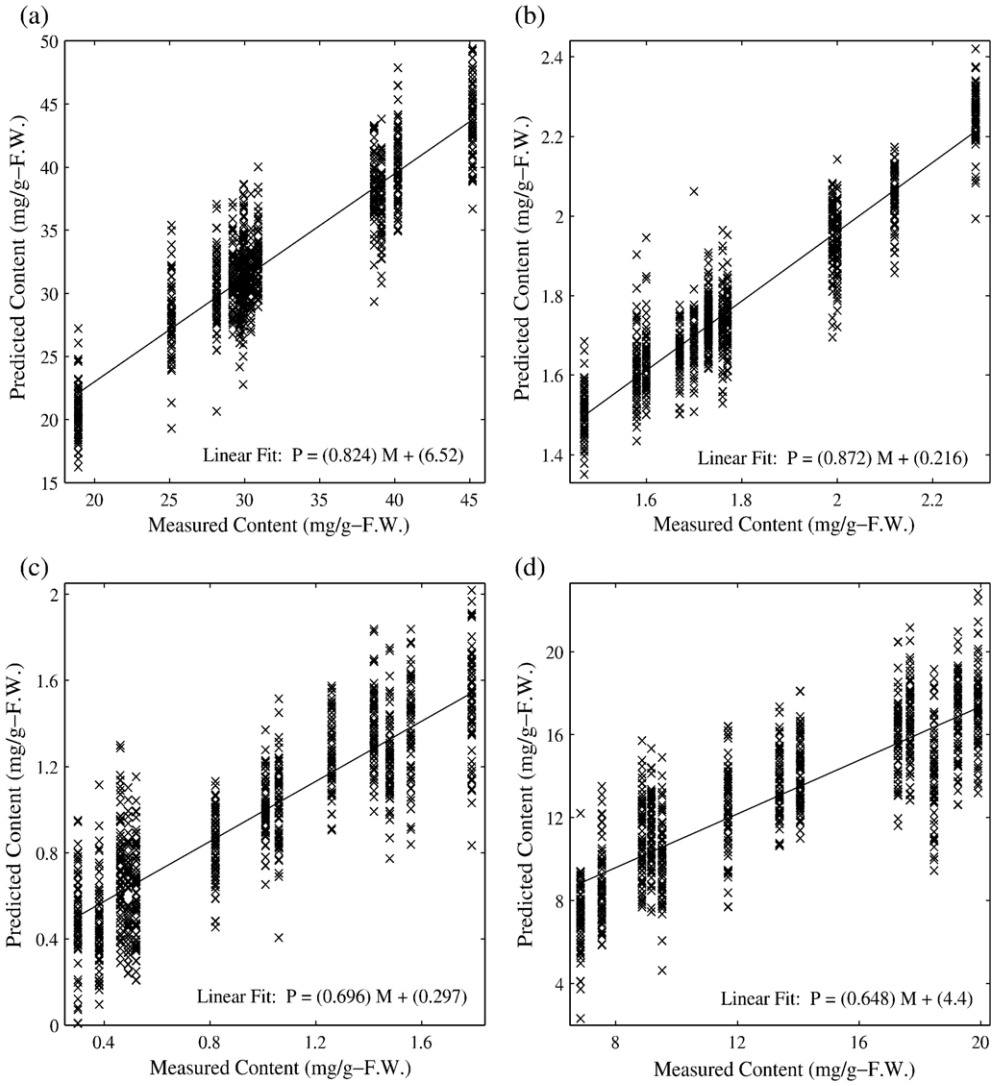

Fig. 3. Linear regression analysis between the predictions given by the MLP network and the ground reference measurements. The cross markers show the corresponding observations and the straight lines indicate the best linear fit of the data, as expressed by the linear equations. (a) Sucrose. (b) Glucose. (c) Fructose. (d) Nitrogen.

2.5.1. Mean of squared errors

One measure of network performance is the mean of squared errors (MSE) between corresponding elements of the network response (prediction) and the target (ground reference measurements). The MSE is defined by

$$\mathrm{MSE}(\boldsymbol{y}) = \frac{1}{N}\sum_{i=1}^{N} (\boldsymbol{y}^p - \boldsymbol{y}^m)^2, \tag{5}$$

where N is the total number of prediction comparisons, and p and m indicate, respectively, the predicted and measured values.

2.5.2. Correlation coefficient

Another way of evaluating network performance is to calculate the correlation coefficient (R-value) between the network response and the actual measurements. As the strength of the relationship between the predicted and measured values increases, so does the correlation coefficient value. A perfect prediction would give a coefficient of 1. The R-value is calculated by

$$R(\boldsymbol{y}) = \frac{C(\boldsymbol{y}^p, \boldsymbol{y}^m)}{\sqrt{C(\boldsymbol{y}^p, \boldsymbol{y}^p).C(\boldsymbol{y}^m, \boldsymbol{y}^m)}}, \tag{6}$$

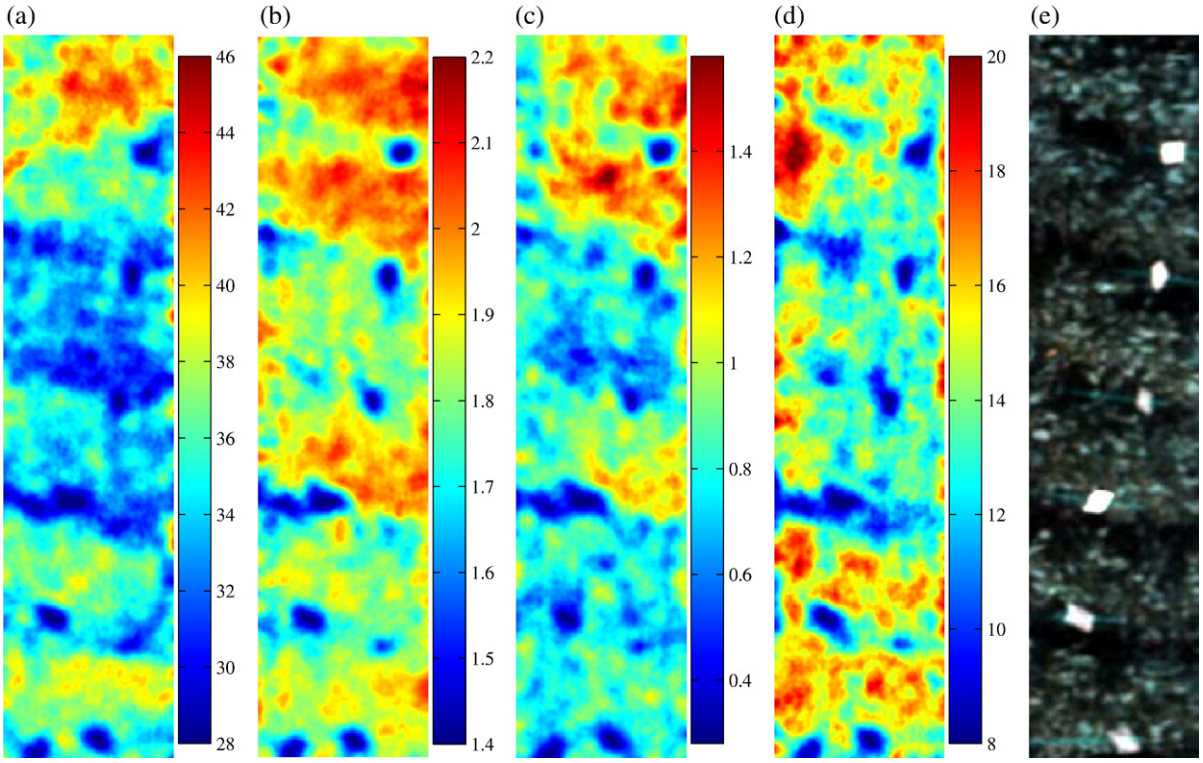

Fig. 4. Example of prediction maps of chemical contents by using ANN regression. The chemical's concentrations are expressed in mg/g-F.W. An RGB visualization of the soybean field's different varieties separated by white labels is displayed in (e), a textual description is shown in Table 4. (a) Sucrose. (b) Glucose. (c) Fructose. (d) Nitrogen. (e) Ground reference.

where, C is the covariance matrix, and p and m indicate the predicted and measured values.

3. Results

The data sample used in this study consists of diverse varieties of soybeans cultivated in an experimental field at Yamagata University, Japan. The hyperspectral image data and ground reference were acquired in the middle of summer on a sunny day from noon to early afternoon. Thirteen varieties of green vegetable soybeans were analyzed. From each labeled area (i.e., variety of soybean), 128 pixels in the images were randomly chosen creating 1664 observations, which when multiplied by the spectral dimension of 121 bands produced a training dataset composed of 201 344 points in total.

A performance analysis was carried out by comparing regression models obtained using input datasets that had passed through different preprocessing steps. Four variations of input datasets were tested:

(1) Whole: all spectral bands available were applied directly to the network.

Table 4
Measured values of chemical concentrations in the test case

Row [a]	Cultivar	Sucrose	Glucose	Fructose	Nitrogen
1	Kanro B	45.18	1.99	1.26	13.38
2	Shonai 4	29.93	2.12	1.56	8.87
3	Ina	29.19	1.60	0.30	11.69
4	Bansei-Sirayama-dadacha	30.89	2.00	1.48	9.17
5	Dadachamame	38.62	1.58	0.52	19.90
6	Wase-dadacha B	40.20	1.73	0.49	19.24

[a] Row numbers correspond to consecutive crop rows, from top to bottom, of the prediction maps in Fig. 4.

Table 5
Performance of the SLP regression models

Chemical	Dataset	MSE	R
Sucrose	Whole	0.121	0.737
	Derivative	0.125	0.727
Glucose	Whole	0.136	0.739
	Derivative	0.138	0.735
Fructose	Whole	0.266	0.615
	Derivative	0.272	0.604
Nitrogen	Whole	0.289	0.637
	Derivative	0.305	0.611

(2) Derivative: the second derivative was calculated from the whole set of spectral bands and applied to the network.
(3) PCA whole: PCA was calculated on the whole dataset and the retained components were then applied to the network.
(4) PCA derivative: PCA was calculated on the second derivative dataset and the retained components were then applied to the network.

In order to retain a reasonable number of principal components for both of the base datasets analyzed by this

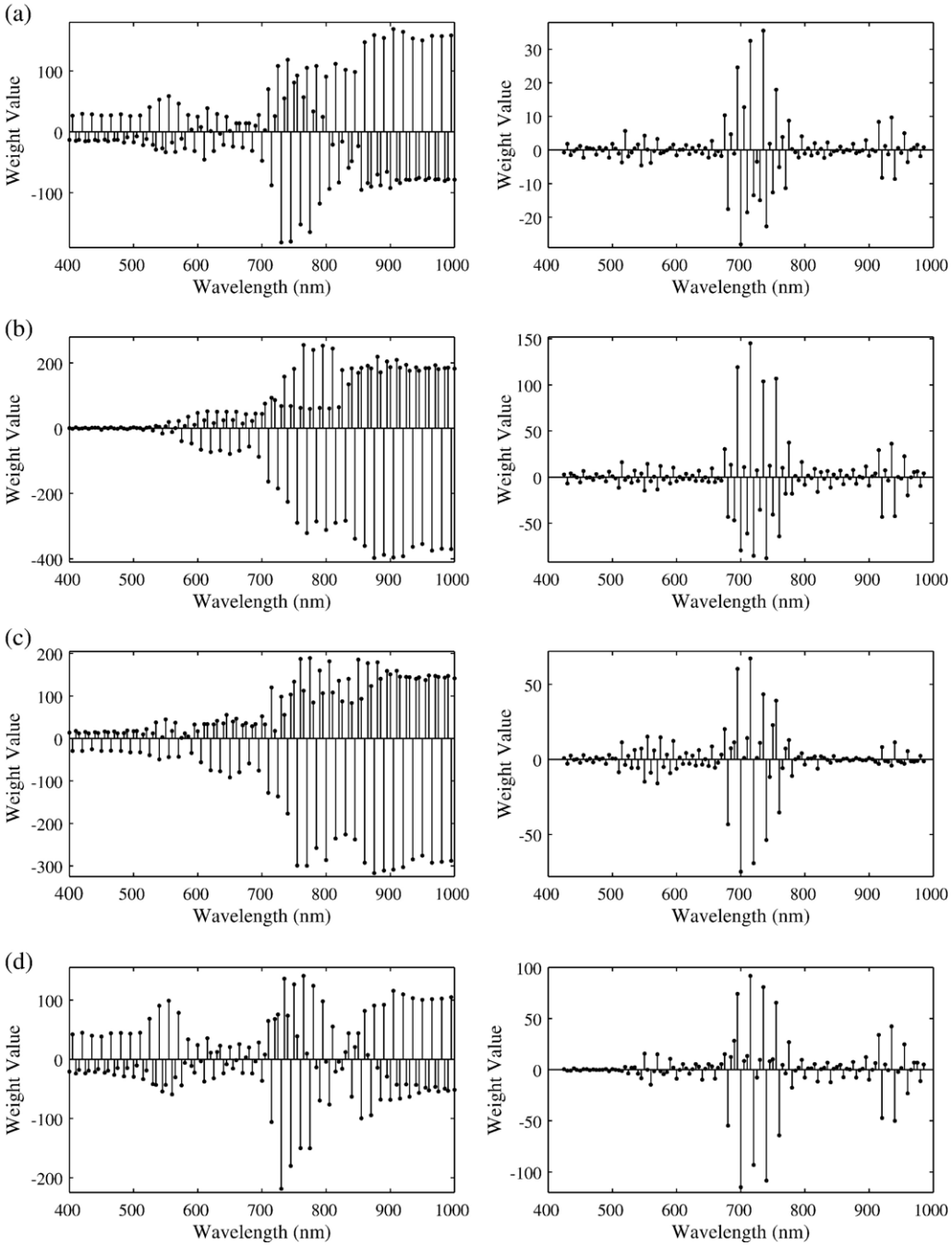

Fig. 5. SLP networks' weight vector. The plots on the left were calculated using the raw reflectance of all spectral bands and those on the right were calculated using the second derivatives. (a) Sucrose. (b) Glucose. (c) Fructose. (d) Nitrogen.

study, raw reflectance and second derivative, a variance threshold of 99.5% was selected for the PCA. The actual size of each type of dataset is shown in Table 2, which corresponds to the number of nodes in the input of the ANNs.

To account for the stochastic nature of the ANN training, each regression model was tested over 50 independent runs, each time starting the weights of the network with random values generated by a different random seed. The results of the MLP learning process of regression models for each chemical substance are summarized in Table 3.

Only the best models were kept, i.e. the ANNs that presented lower MSE and higher *R*-values. A linear regression analysis was then performed between the MLP networks' predicted output and the ground reference measurements obtained by laboratory analysis. The results are shown in Fig. 3 for the second derivative of each chemical substance. For a perfect prediction, each model's best linear fit should have an inclination of 45° (gradient $m = 1$).

Once the ANN regression model is successfully trained, it can be used to predict the chemical content of each pixel on the spatial plane of a hyperspectral image of the crop field in order to produce a prediction map. As a test case, a portion of the experimental field having different varieties of soybeans in adjacent rows was processed. The prediction maps were calculated using the best models, which were obtained by the second derivatives. These are presented in Fig. 4. For display purposes, non-vegetative regions were assigned the lowest concentration values. The separation between the plant rows is indicated by the white labels in Fig 4(e). The mean value of the actual chemical concentrations measured in the laboratory for each respective row is displayed in Table 4.

The SLP network was tested to model all spectral bands and the second derivative dataset. The SLP network's weight vector gives an indication of each input entry's contribution to the output calculation. If the input data is physically meaningful, then it may offer an insight about which wavelengths are more important to the chemical content prediction. The complexity of calculating Eq. (3) denies the possibility of such analysis using the MLP. Since PCA transforms the spectral data into uncorrelated components without explicit physical meaning, it was not included in this analysis. The SLP models' performances are displayed in Table 5. The weight vectors $\boldsymbol{W}$ from Eq. (4) for each case of dataset and chemical substance are plotted in Fig. 5. The weights were matched to the wavelengths of the respective input spectral bands.

4. Discussion

In the range of wavelengths investigated, greater accuracy was obtained for the calculation of glucose, followed by sucrose, fructose, and nitrogen. The derivatives contributed a slight and gradual improvement to the prediction accuracy of the MLP models. The performance of the second derivative analysis may have been affected by errors present in the experimental dataset (low signal-to-noise ratio) in wavelengths greater than 900 nm. Some of these wavelength bands may be correlated with the chemical substance being targeted and thus necessary to improve the prediction accuracy. Nevertheless, the calculation of the second derivative could also contribute to improving the generalization of the regression model for diverse weather conditions other than those observed during the acquisition of the training data. This is due to its alleged robustness in accounting for variations in illumination intensity in the sample. Furthermore, the second derivative provided a more selective use of the spectrum as can be observed in Fig. 5.

The models based on the standard PCA generally presented inferior performances. This result may be due to the principal components' characteristic of retaining the main features of the dataset, but not necessarily those correlating with the chemical contents. Furthermore, the accuracy of the PCA-based models may have been affected by accumulation of errors in the calculation of the covariance matrix from the hyperspectral data (Chitroub et al., 2001). Ustin et al. (2004) also suggest that the measurement of a broader spectral range may possibly provide a better basis for predicting chemical concentrations.

The weight vector of the SLP model may indicate the significance of wave-length bands on the prediction calculation. In the case of all bands of the raw reflectance, i.e., the left plots in Fig. 5, a coincident spectral region ranging from 720 nm to 1000 nm presented higher weight values for sucrose, glucose and fructose. For nitrogen, the same region was also important, except for a trough of low weights around 820 nm. On the second derivative models, i.e., the right plots in Fig. 5, the coinciding regions were even more pronounced, presenting two distinctive regions in the range 675 nm to 775 nm and 915 nm to 960 nm. The former corresponds approximately to the "red edge" point, which is frequently used to characterize the water stress of vegetation (Filella and Penuelas, 1994). This analysis indicates that, although one chemical substance was predicted more accurately than others, they could all be calculated using wavelengths in a

narrow spectral range. Another interesting outcome was that the derivative-based models presented the best performance overall for the MLP networks, which are nonlinear by nature. This suggests that the second derivative of the spectral bands may present some nonlinear relationship correlating to sweetness and amino acid content. However, the MLP weights, which constitute a matrix processed by the activation function, could not provide an analysis of the input vector's physical importance as the SLP weights were capable of doing.

The graphical representation of chemical contents calculated from the hyperspectral image data provides a valuable visualization of the spatial distribution of variations in the state of the crop field. To achieve greater accuracy of the prediction maps, the training data could be improved by increasing the number of ground reference measurements. Another approach could be to stratify the training data by soybean variety, which would possibly produce ANN models that are more accurate. However, due to the limited sample size and poor diversity of such stratified training datasets, the ANN models would probably have difficulty in generalizing the results under different conditions (Niyogi and Girosi, 1996), thus limiting the models' usefulness. On the other hand, since the diversity of the training dataset is also important, not only larger, but also more reliable prediction maps could be generated if specific experiments were conducted to acquire data from the soybean field under different environmental and temporal conditions.

5. Conclusions

The ANN-based approach provides a reasonably accurate regression model for the estimation of the sweetness and amino acid concentrations in green vegetable soybeans from hyperspectral image data. The nonlinear regression model of the second derivative produces the best predictions for glucose, sucrose, fructose, and nitrogen concentrations using wavelength bands from the visible to the near infrared. The regression models obtained from the hyperspectral dataset reduced by PCA present the worst correlations. The results of the linear regression model suggest that spectral regions around the red edge are especially significant for the prediction of the chemical concentrations.

The proposed method permits one to exploit the capabilities of high-resolution hyperspectral imagery for estimating the chemical contents of soybean crops. The system could be used to monitor the conditions of localized crop fields before harvesting and without interfering in the growth process. The method could also be applied for predicting the chemical contents of large soybean fields by acquiring hyperspectral data from higher altitudes, perhaps airborne or spaceborne. Nevertheless, if the data contain mixed pixels, careful preprocessing would be required to estimate the pure spectra before generating the prediction maps.

Acknowledgment

This research was partly supported by the Grant-in-Aid for Scientific Research number 30108237 of the Japan Society for the Promotion of Science.

References

Ardrey, R.E., 2003. Liquid chromatography. In: Ando, D.J. (Ed.), Liquid Chromatography Mass Spectrometry: An Introduction. John Wiley & Sons, Inc., New York, pp. 7–31.

Bishop, C.M., 1995. Neural Networks for Pattern Recognition. Oxford University Press, New York.

Chitroub, S., Houacine, A., Sansal, B., 2001. Principal component analysis of multispectral images using neural network. Proc. ACS/IEEE Intl. Conf. Computer Systems and Application, pp. 89–95.

Cihlar, J., 2000. Land cover mapping of large areas from satellites: status and research priorities. International Journal of Remote Sensing 21 (6-7), 1093–1114.

Cybenko, G., 1989. Approximation by superpositions of a sigmoid function. Mathematics of Control, Signals and Systems 2, 303–314.

Du, P., Chen, Y., Yang, Y., Zhang, H., 2005. On the filtering of hyperspectral remote sensing image. In: Li, D., Ma, H. (Eds.), Proc. of the SPIE, MIPPR 2005: Image Analysis Techniques, vol. 6044, pp. 347–353.

Filella, I., Penuelas, J., 1994. The red edge position and shape as indicators of plant chlorophyll content, biomass and hydric status. International Journal of Remote Sensing 15 (7), 1459–1470.

Fukunaga, K., 1990. Introduction to Statistical Pattern Recognition, 2nd Edition. Academic Press, San Diego, CA.

Funahashi, K., 1989. On the approximate realization of continuous mappings by neural networks. Neural Networks 2, 183–192.

Hagan, M.T., Menhaj, M.B., 1994. Training feedforward networks with the Marquardt algorithm. IEEE Transactions on Neural Networks 5 (6), 989–993.

Haykin, S., 1999. Neural Networks: A Comprehensive Foundation, 2nd Edition. Prentice Hall, Englewood Cliffs, NJ.

Hruschka, W.R., 2001. Spectral reconstruction, In: Burns, D.A., Ciurczak, E.W. (Eds.), Handbook of Near-infrared Analysis, 2nd Edition. Marcel Dekker Inc., New York, pp. 401–419.

Jolliffe, I.T., 1988. Principal Component Analysis. Springer-Verlag, New York.

Kavzoglu, T., Mather, P.M., 2003. The use of backpropagating artificial neural networks in land cover classification. International Journal of Remote Sensing 24 (23), 4907–4938.

Koehler IV, F.W., Lee, E., Kidder, L.H., Lewis, E.N., 2002. Near infrared spectroscopy: the practical chemical imaging solution. Spectroscopy Europe 14 (3), 12–19.

Liu, K., 1997. Soybeans: Chemistry, Technology, and Utilization. Chapman and Hall, New York.

Minekawa, Y., Uto, K., Kosugi, Y., Oda, K., 2004. Development of crane-mounted hyperspectral imagery system for stable analysis of paddy field. Proc. International Symposium on Remote Sensing, Jeju, Korea.

Monteiro, S.T., Minekawa, Y., Kosugi, Y., Akazawa, T., Oda, K., 2006. Prediction of sweetness and nitrogen content in soybean crops from high resolution hyperspectral imagery. Proc. 2006 IEEE International Geoscience and Remote Sensing Symposium, Denver, Colorado, vol. 5, pp. 2263–2266.

Murphy, A.H., Daan, H., 1985. Probability, statistics, and decision making in the atmospheric sciences. In: Murphy, A., Katz, R. (Eds.), Forecast Evaluation. Westview Press, Boulder, Colorado, pp. 379–437.

Myneni, R.B., Hall, F.G., Sellers, P.J., Marshak, A.L., 1995. The interpretation of spectral vegetation indexes. IEEE Transactions on Geoscience Remote Sensing 33 (2), 481–486.

Niyogi, P., Girosi, F., 1996. On the relationship between generalization error, hypothesis complexity, and sample complexity for radial basis functions. Neural Computation 8 (4), 819–842.

Osborne, B.G., Fearn, T., 1986. Near infrared spectroscopy in food analysis. In: Meyers, R.A. (Ed.), Encyclopedia of Analytical Chemistry. John Wiley & Sons, New York, pp. 86–103.

Petisco, C., García-Criado, B., Vázquez de Aldana, B.R., Zabalgogeazcoal, I., Mediavilla, S., García-Ciudad, A., 2005. Use of near-infrared reflectance spectroscopy in predicting nitrogen, phosphorus and calcium contents in heterogeneous woody plant species. Analytical and Bioanalytical Chemistry 382 (2), 458–465.

Richards, J.A., Jia, X., 1999. Remote Sensing Digital Image Analysis, An Introduction, 3rd Edition. Springer-Verlag, New York.

Rouse, J.W., Hass, R.H., Schell, J.A., Deering, D.W., 1973. Monitoring vegetation systems in the great plains with ERTS. Third ERTS Symposium, NASA SP-351, vol. 1, pp. 309–317.

Schowengerdt, R.A., 1997. Remote Sensing: Models and Methods for Image Processing, 2nd Edition. Academic Press, San Diego, CA.

Siesler, H.W., Ozaki, Y., Kawata, S., Heise, H.M., 2002. Near-Infrared Spectroscopy Principles, Instruments, Applications. Wiley-VCH, Weinheim, Germany.

Slaughter, D.C., Barrett, D., Boersig, M., 1996. Nondestructive determination of soluble solids in tomatoes using near infrared spectroscopy. Journal of Food Science 61 (4), 695–697.

Spectral Imaging Ltd., 2003. ImSpector Imaging Spectrograph User Manual, 2nd Edition.

Toko, K., 1998. A taste sensor. Measurement Science & Technology 9, 1919–1936.

Tsai, F., Philpot, W., 1998. Derivative analysis of hyperspectral data. Remote Sensing of Environment 66, 41–51.

Tsuta, M., Sugiyama, J., Sagara, Y., 2002. Near-infrared imaging spectroscopy based on sugar absorption band for melons. Journal of Agricultural and Food Chemistry 50 (1), 48–52.

Ustin, S.L., Roberts, D.A., Gamon, J.A., Asner, G.P., Green, R.O., 2004. Using imaging spectroscopy to study ecosystem processes and properties. Bioscience 54 (6), 523–534.

Widrow, B., Winter, R., 1988. Neural nets for adaptive filtering and adaptive pattern recognition. IEEE Computer 21 (3), 25–39.

Zude, M., 2003. Comparison of indices and multivariate models to non-destructively predict the fruit chlorophyll by means of visible spectrometry in apples. Analytica Chimica Acta 481, 119–126.