



Rapid detection of *Escherichia coli* contamination in packaged fresh spinach using hyperspectral imaging

U. Siripatrawan^{a,*}, Y. Makino^{b,**}, Y. Kawagoe^b, S. Oshita^b

^a Department of Food Technology, Faculty of Science, Chulalongkorn University, Bangkok, Thailand

^b Department of Biological & Environmental Engineering, Graduate School of Agricultural & Life Science, The University of Tokyo, Japan

ARTICLE INFO

Article history:

Received 23 January 2011

Received in revised form 20 March 2011

Accepted 24 March 2011

Available online 5 April 2011

Keywords:

Hyperspectral imaging

Rapid detection

Packaged spinach

E. coli

Chemometrics

ABSTRACT

A rapid method based on hyperspectral imaging for detection of *Escherichia coli* contamination in fresh vegetable was developed. *E. coli* K12 was inoculated into spinach with different initial concentrations. Samples were analyzed using a colony count and a hyperspectroscopic technique. A hyperspectral camera of 400–1000 nm, with a spectral resolution of 5 nm was employed to acquire hyperspectral images of packaged spinach. Reflectance spectra were obtained from various positions on the sample surface and pretreated using Savitzky–Golay. Chemometrics including principal component analysis (PCA) and artificial neural network (ANN) were then used to analyze the pre-processed data. The PCA was implemented to remove redundant information of the hyperspectral data. The ANN was trained using Bayesian regularization and was capable of correlating hyperspectral data with number of *E. coli*. Once trained, the ANN was also used to construct a prediction map of all pixel spectra of an image to display the number of *E. coli* in the sample. The prediction map allowed a rapid and easy interpretation of the hyperspectral data. The results suggested that incorporation of hyperspectral imaging with chemometrics provided a rapid and innovative approach for the detection of *E. coli* contamination in packaged fresh spinach.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Fresh spinach and other raw or minimally processed vegetables are under increasing demand by consumers due to their popular perception as a healthy food. However, fresh vegetables are common vehicles for fresh produce-associated foodborne illness. Fresh spinach has been implicated in outbreaks of *Salmonella* spp. [1], *Escherichia coli* O157:H7, *Listeria monocytogenes*, and *Shigella flexneri* infections as well as other communicable diseases in recent years [2–5]. Although the precise means by which the bacteria spread to the spinach remain unknown, these contaminations may occurred through different steps and processes, including field contamination or cross-contamination during processing. Such incidents have caused considerable mortality, morbidity and financial losses [2,6].

Pathogenic detection is crucial to implement disease control measures. Early detection of foodborne contamination directly in packaged fresh vegetables can be useful to prevent the intake of contaminated products. Recently, research has focused on development of rapid and accurate techniques to identify

pathogens in food products [7,8]. Yang et al. [9] used a miniature microfluidic flow cytometer integrated with several functional micro-devices capable of sample purification and detection by utilizing a magnetic bead-based immunoassay to recognize and capture target microorganisms. Cheng et al. [10] used biofunctional magnetic nanoparticles combined with ATP bioluminescence for detection of *E. coli*. Hong et al. [11] developed QCM-DNA biosensor to detect a main viral RNA encoding G protein in viral haemorrhagic septicaemia infection. These current pathogenic detection techniques are rapid, specific and sensitive. However, most are technically complicated and costly, and require well-trained specialists. Therefore, a rapid, simple, and inexpensive detection and identification method of pathogens in food is still needed.

Near infrared (NIR) spectroscopy is a rapid and non-destructive multi-component analytical technique enabling several determinations to be made simultaneously without requiring extensive sample preparation [12]. NIR spectroscopy has been used for food quality evaluation and analyses of microorganisms including determination of the content of ergosterol and fumosin B1 in maize [13], evaluation of bacterial contamination in shredded cabbage [14], and detection and discrimination of *E. coli* K12 and *E. coli* 25922 [15]. However, conventional NIR spectroscopic instruments are considered as point-based scanning instruments as they provide one spectrum of the target sample without giving any information about the distribution of the chemical composition of a sample,

* Corresponding author. Tel.: +662 218 5536.

** Corresponding author. Tel.: +813 5841 5361.

E-mail addresses: ubonratana.s@chula.ac.th (U. Siripatrawan), amakino@mail.ecc.u-tokyo.ac.jp (Y. Makino).

and thus spectra obtained from NIR instruments do not give spatial information [16–18].

Hyperspectral imaging (HSI) is an emerging technique that assimilates spectroscopy and imaging to provide both spectral and spatial information on the distribution of the components of an object. Compared to NIR, hyperspectral imaging is a more powerful technique. HSI generates data cubes from which the pertinent qualitative and quantitative information can be obtained [19–21]. HSI hypercubes are a three dimensional pattern of data, comprising two spatial and one wavelength dimension. Spatial information is important for monitoring of the sample as it can be used to extract the chemical mapping of the sample from a hypercube [18,20,22]. For food safety applications, HSI has been successfully used for detection of fecal contamination of apples [16], contaminant classification of poultry [23], and early detection of toxigenic fungi on maize [24].

Heretofore, no study has been reported on using HSI for detection of foodborne pathogens in fresh spinach. Therefore, the objective of this study was to develop a rapid, non-destructive, and accurate method based on HSI integrated with chemometrics for an early detection of *E. coli* in packaged fresh spinach. The principle underlying this approach was based on the assumption that the metabolic activity of microorganisms on spinach results in biochemical changes with the concurrent formation of metabolic by-products which potentially indicate the contamination. Hence, the HSI analysis of these metabolites would provide characteristic fingerprint which can be used to indicate the contamination of *E. coli* in the samples. This present research is an important step towards the development of a rapid and nondestructive method for detection of pathogenic contamination in packaged fresh vegetables with minimal sample manipulation.

2. Material and methods

2.1. Preparation of stock culture and inoculated vegetable

The nonpathogenic strain *E. coli* K12 (NBRC 3301) was obtained from the National Institute of Technology and Evaluation Biological Resource Center (Chiba, Japan). Before use, *E. coli* was cultured in Luria Bertani broth (Becton, Dickinson and Company, Sparks, MD, USA), and incubated at 37 °C for 8 h in a program incubator (IN801, Yamato Scientific Co., Ltd., Tokyo, Japan) with an agitation rate of 100 rpm using a desktop rotary shaker (Invitro, TAITEC Co., Ltd., Saitama, Japan) and centrifuged using a refrigerated micro-centrifuge (MX-301, Tomy Seiko Co., Ltd., Tokyo, Japan). Broth was poured from the culture and the sedimented pellet was resuspended in sterile Butterfield's phosphate buffer which was used as a dipping suspension. Preliminary experiments were conducted to determine the population of *E. coli* necessary in the dipping suspension to result in an initial population of ~5 to ~7 Log (CFU/g) on spinach leaves. According to our previous work [15], it has been suggested that NIR spectroscopy was able to detect *E. coli* when its number is higher than 5 Log (CFU/g).

Spinach (*Spinacia oleracea* L.) leaves were purchased from a local supermarket earlier on the day of the experiment and stored at 5 °C until use. Spinach leaves with decay, cuts, or bruises were discarded, and only fresh unblemished leaves were used. The leaves were washed and drained several times before use. They were then placed in screened baskets, and submerged in the suspension containing *E. coli* for 5 min. Different sample subgroups were prepared, including non-inoculated spinach (control) and spinach inoculated with different concentration of *E. coli* (K12A, K12B, K12C and K12D). The uninoculated control was similarly treated except sterile phosphate buffer was used in place of the inoculum. Fifty grams of spinach were then packed into a commercial

100 mm × 70 mm × 0.08 mm low density polyethylene (LDPE) bag and sealed. The samples were incubated at 10 °C for 48 h in order to allow *E. coli* to grow.

2.2. Microbiological analysis

The microbial cell count was determined on all packaged samples. Spinach leaves were aseptically removed from plastic bags and placed in phosphate buffer. Leaves were stomached for 1 min on high using a Stomacher Lab-blender (Pro-media SH-001, Elmex, Ltd., Tokyo, Japan) to remove microbes from the leaf surface. Serial dilutions were prepared from the stock suspension, and Petri plates were inoculated with those dilutions expected to give countable colonies. Inocula consisting of each of a dilution serials were deposited on prepared plates in duplicate using 3 M Petrifilm Aerobic Count Plates (3 M; St. Paul, MN, USA) for determining aerobic bacteria and 3 M Petrifilm *E. coli*/Coliform Count Plates (3 M; St. Paul, MN, USA) containing Violet Red Bile nutrient agar as an indicator of glucuronidase activity for *E. coli*. All plates were incubated at 37 °C for 48 h. Plate counts are recorded as colony forming units per gram (CFU/g).

2.3. Hyperspectral imaging system

A hyperspectral imaging system (JFE, Techno-Research Corporation, Tokyo, Japan) was used to produce full contiguous spectral and spatial information and project it on a 12 bit charge coupled device (CCD) camera. The Pentax C33500 having a 16 mm focal length for 2/3-inch CCD C-Mount camera was used. The lens has an aperture range of f/1.6–f/16 which provides both low-light sensitivity and depth of field control, covering a narrow 14.76° horizontal view. The light sources consisting of 150 W Tungsten halogen lamp (ColdSpot PCS-UHX, NPI, Tokyo, Japan) and 150 W Xe lamp (Super Bright 152S, SAN-EI Electric, Osaka, Japan), fixed at 45° angles from the imaging area were used as illumination sources.

The system is attached to a stage control unit (Model SGSP 26-200, Sigma-Kaki Co., Ltd., Tokyo, Japan), so that a hyperspectral image cube can be build by scanning in the direction perpendicular to the spatial plane. The ImSpector (Model V10, Spectral Imaging Ltd., Oulu, Finland) which creates a spectrum from 400 to 1000 nm with a spectral resolution of 5 nm was used to align the imaging system, acquire images, and store hyperspectral image data in a 12-bit binary file.

The SpectrumAnalyzer (Version 1.8.5, JFE, Techno-Research Corporation, Tokyo, Japan) software was used to control the stepper table and to construct the hyperspectral images. The program stores the imaging system setup, wavelength range and number of wavelength slices along with the hyperspectral imaging data in a file. The hyperspectral data were acquired through a Dell Vostro Intel(R) computer (Dell Inc., Tokyo, Japan).

2.4. Data analysis

Data were made up of 150 samples from 5 subgroups, including non-inoculated spinach (Control) and spinach inoculated with *E. coli* (K12A, K12B, K12C and K12D). Each subgroup had 30 replicate samples collected from several cultivations. The detector signal intensity counts were transformed into reflectance units. Reflectance spectra of each sample treatment were obtained from various positions on the sample surface and pretreated using Sawitzky–Golay. Chemometrics including principal component analysis (PCA) and artificial neural network (ANN) were then used to analyze the outcome of hyperspectroscopic technique.

The PCA was used for compressing the wavelength variables. The ANN based on multilayer perceptrons (MLP) with back propagation algorithm was used to predict the number of *E. coli* from the

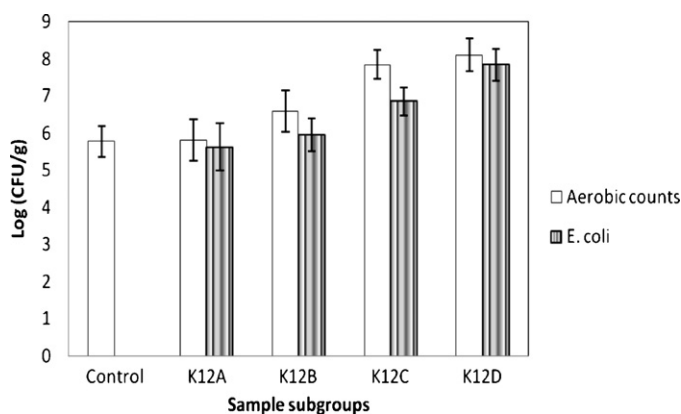


Fig. 1. Aerobic bacteria and *E. coli* (Log (CFU/g)) in the packaged spinach samples. Values represent mean (\pm SD) of measurements made on 30 independent samples per treatment.

hyperspectral reflectance. The data sets were partitioned into two subsets: a training set and a validation set. The training set is a set of samples used to adjust the network weights. The performance of the neural network was confirmed by measuring its performance on an unseen, independent validation set.

In order to facilitate a rapid and easy interpretation of the hyperspectral data, the prediction map of the number of *E. coli* contamination was performed with each pixel on the spatial plane of a hyperspectral image. The prediction map was constructed in accordance with sample subgroups using the developed ANN prediction algorithm and was performed on an unseen independent set of data. All matrix calculations were performed using MATLAB (Mathworks, Inc., Natick, MA, USA) routines written by the authors.

3. Results and discussion

3.1. Microbial cell counts

The cell counts of aerobic bacteria and *E. coli* on spinach are shown in Fig. 1. All samples had a high number of total aerobic bacteria. The numbers of *E. coli* from K12A to K12D were varied from 5.1 to 7.4 Log (CFU/g).

3.2. Data pre-processing

Fig. 2 is shown the hyperspectral data of one spinach sample, covering the effective spectral range of 400–1000 nm and a

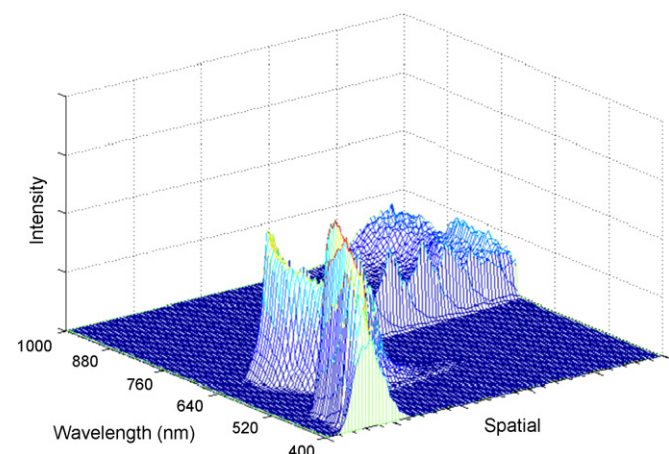


Fig. 2. Hyperspectral data covering the effective spectral range of 400–1000 nm and a total spatial distance.

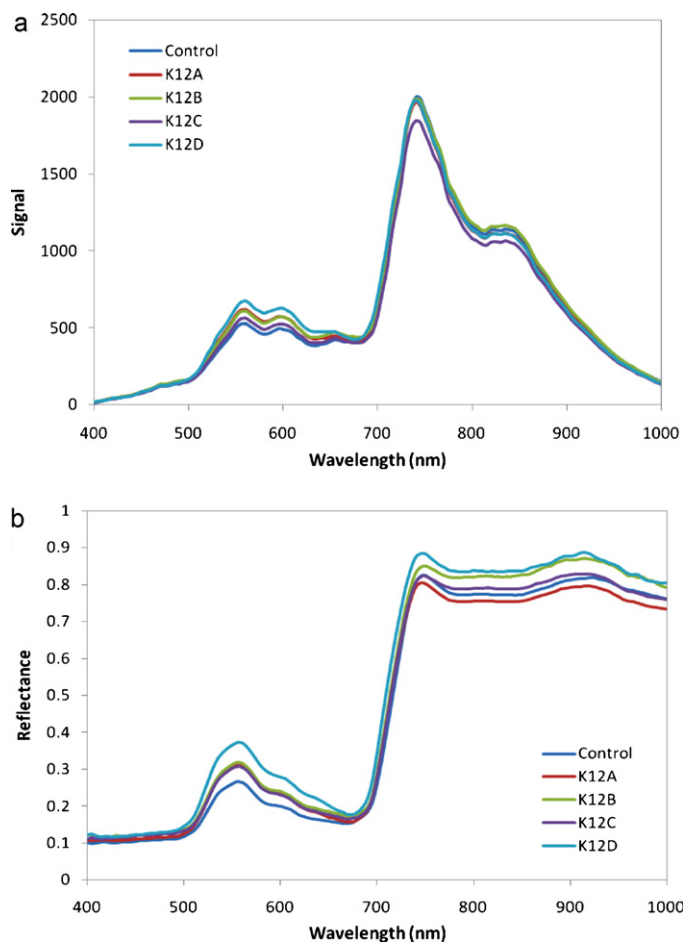


Fig. 3. Average spectra of all samples for each sample subgroup (a), Averaged pre-processed reflectance (b) using Sawitzky-Golay with a second order polynomial.

total spatial distance in which the intensity of the image is illustrated as the third dimension in addition to the spatial and spectral dimensions. A vertical line parallel to the spectral axis represents a spectral profile taken from a specific distance on the spatial horizontal axis. On the other hand, a horizontal line parallel to the spatial axis represents a spatial profile of a specific wavelength. Accordingly, each hyperspectral image is comprised of a number of spectra representing different spatial or pixels on the spinach surface, and thus each pixel in a hyperspectral image contains the spectrum of that specific position. The resulting spectrum can be thought of as a fingerprint which can be used to characterize the composition of that particular pixel.

Fig. 3(a) presents average raw signals of each sample subgroup. The detector signal intensity counts (S_D) were transformed into reflectance (R) units by comparing with spectra of dark current (D_k) and dividing by similarly corrected total reflectance spectrum (Rf_{100}) using Eq. (1). Reflectance spectra of each sample treatment were obtained from various positions on the sample surface and pretreated using Sawitzky-Golay. The averaged preprocessed reflectance using Sawitzky-Golay with a second order polynomial is shown in Fig. 3(b). The mean spectrum was used to represent the spectral features of the hyperspectral image. Mean reflectance of control sample was discernible from that contaminated with *E. coli*, probably due to metabolic compounds generated by inoculated *E. coli* during the incubation period.

$$R = \frac{(S_D - D_k)}{(Rf_{100} - D_k)} \quad (1)$$

3.3. Data processing

HSI generates data cubes from which the relevant qualitative and quantitative information could be extracted when optimal data analyses are applied. Such information could not be extracted by classical NIR spectroscopy because the quantification is performed on the integration of a signal over the sample surface and the spatial information is lost. The hyperspectral data have two pixel coordinates and a variables index (wavelength value) making a three-dimensional array (hypercube).

PCA was implemented to reduce the spectral dimension of the hyperspectral data. The method generated a new set of variables named (PCs). In order to apply chemometric techniques to the hyperspectral data, the hypercube was first unfolded and then restructured into a two-dimensional matrix. The standardization was then applied to the unfolded spectra, followed by projection of the data into the direction defined by the PCs (Fig. 4a). The scree

plot of the PCs was generated and is shown in Fig. 4b. In this study, three PCs were extracted, as they explained a cumulative variance of 98% of all wavelength variables.

PCA searches for lower dimensional spaces which reflect as much as possible the original data. The projections of each sample onto the new space of the resulting PCs from the PCA analysis are shown in Fig. 4c. The principal component scores represent the transformation of the original observations into the new coordinate space defined by the principal components. They provide a measure of distance indicating how closely each sample conforms to the mode of variability represented by each principal component. Since each of the principal components captures variability across all wavelengths, not all wavelengths are required to reconstruct the original data set within a given level of accuracy. Therefore, it is possible to reduce the dimensionality of the data matrix and retain only those principal components that reflect the primary modes of variation.

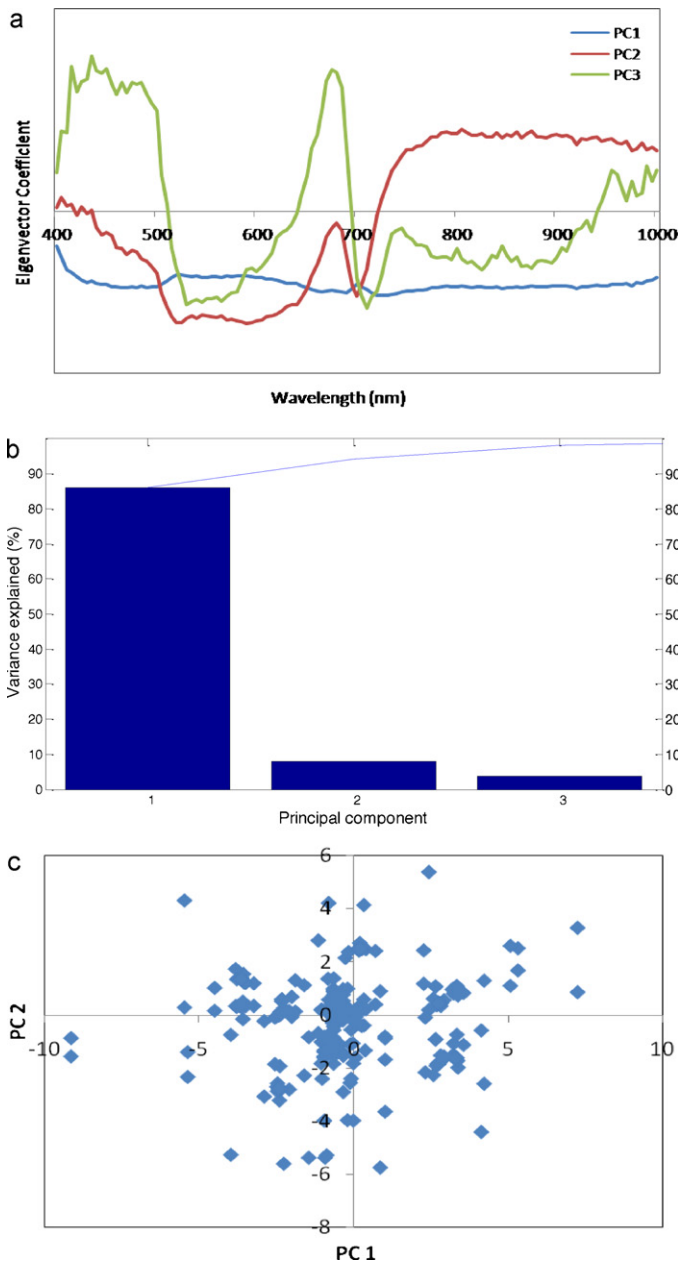


Fig. 4. PCA of hyperspectral data: (a) eigenvector coefficients for PC1 PC2 and PC3; (b) scree plot and percent variance explained of each PC; and (c) sample distribution in the new space obtained from the PCA.

3.4. Prediction of *E. coli* contamination

A MLP neural network based on back propagation was used to predict number of *E. coli* from the hyperspectral data. The network architecture created for the *E. coli* data matrix comprised an input layer, one hidden layer of neurons, and an output layer. Based on preliminary analysis, the best performance of the network was obtained when there were 3 neurons in the input layer, 3 neurons in the hidden layer, and one neuron in the output layer.

In mathematical terms, a neuron k can be described by Eq. (2). The inputs to a neuron include its bias and the sum of its weighted input. The output of a neuron depends on the neuron's inputs and on its transfer function. The transfer function in the hidden layer was a hyperbolic tangent and a linear function was used in the output layer.

$$y_k = \varphi \left(\sum_{j=1}^m w_{kj} x_j + b_k \right) \quad (2)$$

where y_k is the output, $\varphi(\cdot)$ is the transfer function associated with the neuron k , b_k is the bias, and x_j is the input signal, and w_{kj} is the weight connection of neuron j and neuron k .

The retained components calculated from PCA were used as the inputs and the numbers of *E. coli* were used as the outputs. The difference between target value and actual neural output was propagated back through the network to the input. The error was minimized by adjusting the weight. In the learning process, the updated value of weight is computed by

$$w_{kj}(n+1) = w_{kj}(n) + \eta e_k(n) x_j(n) \quad (3)$$

where $w_{kj}(n+1)$ is the weight applied to the synaptic weight at time step $n+1$, $e_k(n)$ is the mean square error at time step n , $x_j(n)$ is the element of the input vector at time step n , and η is the learning rate parameter.

The neural network was trained using the Bayesian regularization which was used to avoid overfitting. The training started with different initial random weights, and was optimized during training. The learning process continued until the synaptic weights and bias level of the network stabilized and the average square error over the entire training set converged to the minimum value. The performance of the trained network was evaluated by measuring the errors in the validation sets. The performance function used during training of the feed-forward neural network was the mean sum of squares of the network errors (MSE).

$$MSE = \frac{1}{N} \sum_{i=1}^N (t_i - a_i)^2 \quad (4)$$

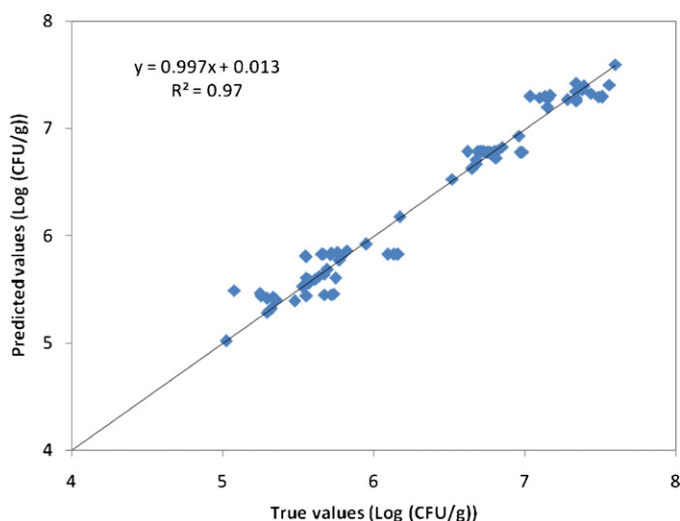


Fig. 5. Predicted versus true values of *E. coli* contamination.

where a is the network output, t is the targets, and N is the number of samples.

Regression analysis was performed between the predicted values and the true values. Fig. 5 shows the predicted versus true values of the numbers of *E. coli* contaminated on the packaged spinach. The coefficient of determination ($R^2 = 0.97$) between the outputs and targets indicates a very good fit between true and predicted data. The goodness of fit of the approach was also evaluated by the mean square error ($MSE = 0.038$), which measures the average deviation between observed and predicted values. The smaller the value of this index the better the fit of the model to the experimental data.

3.5. Prediction map

Hyperspectral data obtained from the spinach sample comprised contiguous wavebands for each spatial position of a target studied and each pixel of the image contained the spectrum of that specific position. The resulting pixel spectra can be thought of as a fingerprint which can be used to characterize the biochemical compositions and concentrations ascribed by *E. coli* present in the sample. Therefore, an individual hyperspectral image can be mapped to provide spatial information of *E. coli* contamination. The developed ANN algorithm was used to determine the number of *E. coli* contamination of each pixel on the spatial plane of a hyperspectral image in order to produce a prediction map in accordance with different sample subgroups.

To construct the prediction map, a rectangular region-of-interest (ROI) of 240 pixels was selected. The hypercube was unfolded, all pixel spectra were computed, and the ANN prediction algorithm was applied to calculate number of *E. coli*. The color map was then constructed on the original spatial locations of the source pixel spectra. A visualization of different sample subgroups is displayed in Fig. 6. Each pixel was colored with respect to the calculated number of *E. coli*, expressed as Log (CFU/g), as shown on the color bar on the right of the prediction map. For comparison purposes, the target color of the expected values was depicted on the left of each prediction map. A wide range of target color was displayed, representing the numbers of *E. coli* depicted in the dataset. The prediction map with intensity scaling shows a direct correlation with the numbers of *E. coli* present in the sample. The color matching between expected and predicted values was agreeable. Slightly inhomogeneous areas could be easily noticed from changes in coloration of the images.

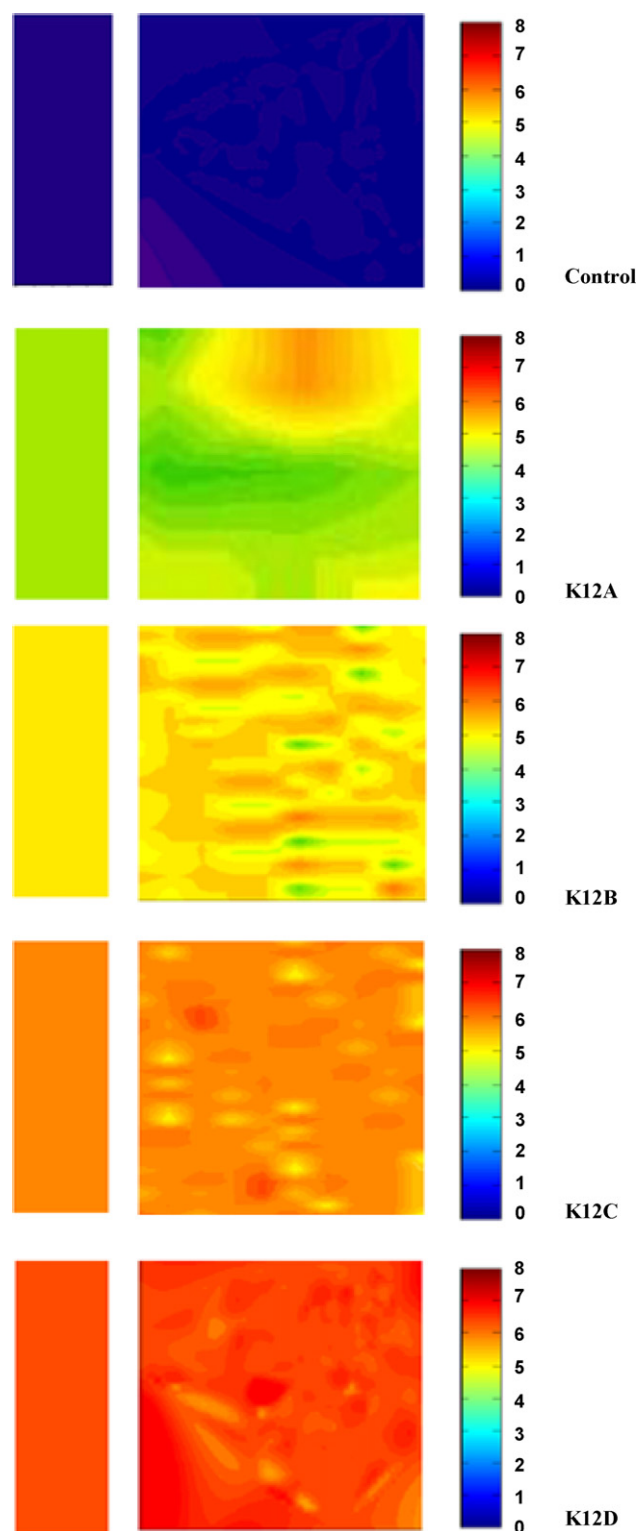


Fig. 6. Prediction map of each pixel on the spatial plane of the sample image indicating the number *E. coli* contamination determined using ANN algorithm. The color corresponds to the number of *E. coli*, expressed as Log (CFU/g), as shown on the color bar. Color intensities are mapped to the full scale values of RGB color space. The target color map of the expected values is depicted on the left of each prediction map.

Construction of prediction map based on chemometric techniques used in this study allowed a rapid and easy interpretation of the HSI data. The simplified visualization of results is encouraging and has potential to be complementary to food safety control system.

4. Conclusion

The results in this study demonstrated that hyperspectral imaging integrated with chemometrics could be used as a rapid and effective tool for detection of target bacteria in packaged fresh spinach. Hyperspectral imaging could provide spectral data and spatial image information to the spectra acquired. PCA was successfully implemented to remove redundant information from the hyperspectral images of all samples. The ANN developed in this study could predict number of *E. coli* from the HSI data. Prediction map of the *E. coli* number to pixel spatial information allowed rapid and convenient data interpretations.

The proposed HSI detection method has high potential to provide a simple, economical, and real-time detection system for pathogenic contamination in packaged fresh vegetables with minimal sample manipulation, and may become a powerful tool for monitoring the safety of packaged fresh vegetables in the food industry.

Acknowledgements

The authors gratefully acknowledge the full financial support provided by The Japan Society for the Promotion of Science (No. P 08441) and a Grant-in-Aid for Scientific Research (JSPS No. 21380155).

References

- [1] J.L. Guentzel, K.L. Lam, M.A. Callan, S.A. Emmons, V.L. Dunham, Food Microbiol. 25 (2008) 36–41.
- [2] Food and Drug Administration, FDA Finalizes Report on 2006 Spinach Outbreak, 27 March (2007), Available from: <http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/2007/ucm108873.htm>.
- [3] M.L. Kotewicz, M.K. Mammel, J.E. LeClerc, T.A. Cebula, Microbiol. (Read.) 154 (2008) 3518–3528.
- [4] B.S.M. Mahmoud, G. Bachman, R.H. Linton, Food Microbiol. 27 (2010) 24–28.
- [5] The Center for Disease Control, Morb. Mortal. Wkly. Rep. (2006) 1045–1046.
- [6] F. Pérez Rodríguez, D. Campos, E.T. Ryser, A.L. Buchholz, G.D. Posada-Izquierdo, B.P. Marks, G. Zurera, E. Todd, Food Microbiol. xxx (2010) 1–8.
- [7] H. Lin, Q. Lu, S. Ge, Q. Cai, C.A. Grimes, Sens. Actuators B: Chem. 147 (2010) 343–349.
- [8] V.C.H. Wu, S. Chen, C.S. Lin, Biosens. Bioelectron. 22 (2007) 2967–2975.
- [9] S. Yang, K. Lien, K. Huang, H. Lei, G. Lee, Biosens. Bioelectron. 24 (2008) 855–862.
- [10] Y. Cheng, Y. Liu, J. Huang, K. Li, W. Zhang, Y. Xian, L. Jin, Talanta 77 (2009) 1332–1336.
- [11] S. Hong, H. Jeoung, S. Hong, Talanta 82 (2010) 890–903.
- [12] H. Cen, Y. He, Trends Food Sci. Technol. 18 (2007) 72–83.
- [13] V. Fernández-Ibañez, A. Soldado, A. Martínez-Fernández, B. de la Roza-Delgado, Food Chem. 113 (2009) 629–634.
- [14] P. Suthiluk, S. Saranwong, S. Kawano, S. Numthuan, T. Satake, Int. J. Food Sci. Technol. 43 (2008) 160–165.
- [15] U. Siripatrawan, Y. Makino, Y. Kawagoe, S. Oshita, Sens. Actuators B: Chem. 148 (2010) 366–370.
- [16] M.S. Kim, A.M. Lefcourt, Y. Chen, Y. Tao, J. Food Eng. 71 (2005) 85–91.
- [17] A.A. Gowen, C.P. O'Donnell, P.J. Cullen, S.E.J. Bell, Eur. J. Pharm. Biopharm. 69 (2008) 10–22.
- [18] A.A. Gowen, M. Taghizadeh, C.P. O'Donnell, J. Food Eng. 93 (2009) 7–12.
- [19] J. Cruz, M. Bautista, J.M. Amigo, M. Blanco, Talanta 80 (2009) 473–478.
- [20] S. Sankaran, R. Ehsani, E. Etxeberria, Talanta 83 (2010) 574–581.
- [21] C. Gendrin, Y. Roggo, C. Collet, J. Near Infrared Spec. 16 (2008) 151–157.
- [22] C.B. Singh, D.S. Jayas, J. Paliwal, N.D.G. White, Comput. Electron. Agric. 73 (2010) 118–125.
- [23] B. Park, W.R. Windham, K.C. Lawrence, D.P. Smith, Biosystems Eng. 96 (2007) 323–333.
- [24] A. Del Fiore, M. Reverberi, A. Ricelli, F. Pinzari, S. Serranti, A.A. Fabbri, G. Bonifazi, C. Fanelli, Int. J. Food Microbiol. 144 (2010) 64–71.