



UNIVERSITY OF LEEDS

This is a repository copy of *A rapid and nondestructive method to determine the distribution map of protein, carbohydrate and sialic acid on Edible bird's nest by hyper-spectral imaging and chemometrics.*

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/112728/>

Version: Accepted Version

Article:

Shi, J, Hu, X, Zou, X et al. (8 more authors) (2017) A rapid and nondestructive method to determine the distribution map of protein, carbohydrate and sialic acid on Edible bird's nest by hyper-spectral imaging and chemometrics. *Food Chemistry*, 229. pp. 235-241. ISSN 0308-8146

<https://doi.org/10.1016/j.foodchem.2017.02.075>

© 2017 Elsevier Ltd. This manuscript version is made available under the CC-BY-NC-ND 4.0 license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Reuse

Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

Accepted Manuscript

A rapid and nondestructive method to determine the distribution map of protein, carbohydrate and sialic acid on Edible bird's nest by hyper-spectral imaging and chemometrics

Jiyong Shi, Xuetao Hu, Xiaobo Zou, Jiewen Zhao, Wen Zhang, Mel Holmes, Xiaowei Huang, Yaodi Zhu, Zhihua Li, Tingting Shen, Xiaolei Zhang

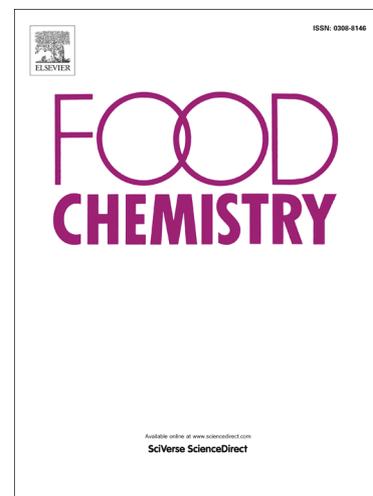
PII: S0308-8146(17)30277-7
DOI: <http://dx.doi.org/10.1016/j.foodchem.2017.02.075>
Reference: FOCH 20628

To appear in: *Food Chemistry*

Received Date: 29 April 2016
Revised Date: 10 September 2016
Accepted Date: 16 February 2017

Please cite this article as: Shi, J., Hu, X., Zou, X., Zhao, J., Zhang, W., Holmes, M., Huang, X., Zhu, Y., Li, Z., Shen, T., Zhang, X., A rapid and nondestructive method to determine the distribution map of protein, carbohydrate and sialic acid on Edible bird's nest by hyper-spectral imaging and chemometrics, *Food Chemistry* (2017), doi: <http://dx.doi.org/10.1016/j.foodchem.2017.02.075>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



A rapid and nondestructive method to determine the distribution map of
protein, carbohydrate and sialic acid on Edible bird's nest by
hyper-spectral imaging and chemometrics

Jiyong Shi^a, Xuetao Hu^a, Xiaobo Zou^{a*}, Jiewen Zhao^a, Wen Zhang^a, Mel Holmes^b

Xiaowei Huang^a, Yaodi Zhu^a, Zhihua Li^a, Tingting Shen^a, Xiaolei Zhang^a

^a School of Food and Biological Engineering, Jiangsu University, Zhenjiang 212013,

China

^b School of Food Science and Nutrition, the University of Leeds, Leeds LS2 9JT,

United Kingdom

*Corresponding author. Tel: +86 511 88780085; Fax: +86 511 88780201

Email address: zou_xiaobo@ujs.edu.cn (Z Xiaobo)

Abstract

Edible bird's nest (EBN) is a precious functional food in Southeast Asia. A rapid and nondestructive method for determining the distribution map of protein content (PC), carbohydrate content (CC) and sialic acid content (SAC) on EBN sample was proposed. Firstly, 60 EBNs were used for hyperspectral image acquisition, and components content (PC, CC and SAC) were determined by chemical analytical methods. Secondly, the spectral signals of EBN hyperspectral image and EBN components content were used to build calibration models. Thirdly, spectra of each pixel in EBN hyperspectral image were extracted, and these spectra were substituted in the calibration models to predict the PC, CC and SAC of each pixel in the EBN image, so the visual distribution maps of PC, CC and SAC on the whole EBN were obtained. It is the first time to show the distribution tendency of PC, CC and SAC on the whole EBN sample.

Keywords: edible bird's nest, distribution map, hyper-spectral imaging, nondestructive

1. Introduction

Edible bird's nests (EBN) are known as the Caviar of the East in Asia (Marcone, 2005). EBNs have long been regarded as a valuable food in China, and the consumption market has been expanding to Western countries (Lau & Melville, 1994). Based on modern research, the content of EBN components and its functions show the important value of the EBN. The proteins, carbohydrates, sialic acid are known to be the major compositional fraction of EBN – comprising of 40%-60%, 10%-30% and 6%-13% of the mass of the food item, respectively (Ma & Liu, 2012). Researchers have figured out the relationships of the components (proteins, carbohydrates, sialic acid and various kinds of elements) and functions of the EBN, such as chondro-protection ability on human articular chondrocytes (Chua et al., 2013), anti-inflammatory properties (Vimala, Hussain & Nazaimoon, 2012) and anti-aging properties (Kim et al., 2012).

Nowadays there is a broad and growing interest in knowing more about the distribution of major component in the whole EBN sample. To our knowledge, no article related to the distribution of major components on EBN sample has been reported in the current literature. Both chemical methods (high performance liquid chromatography (HPLC) (Guo et al., 2006), gas chromatographic (GC) (Chua, Chan, Bloodworth, Li & Leong, 2015), ultraviolet (UV) spectrometry (Saengkrajang, Matan & Matan, 2013) and near-infrared spectroscopy (NIRS) (Deng, Sun, Zhou & Li, 2006) may be used as analytical techniques for quantitative analysis, but these methods belong to “a single point/region” detection method which does not generally include consideration of the major components distribution map in the whole EBN sample.

For the past few years, hyperspectral imaging technology has been used to determine internal and external attributes of biological products (Cen & Lu, 2010; Kamruzzaman, Makino & Oshita, 2016; Konda Naganathan et al., 2016; Zou & Zhao, 2015). Hyperspectral imaging technology combines conventional spectroscopy and

imaging techniques to acquire both spectral and spatial information from an object. In comparison with conventional spectroscopy and imaging techniques, the instruments of hyperspectral imaging technology are much more expensive, the analytical accuracy of hyperspectral imaging technology may be slightly low, and sophisticated mathematical methods are indispensable for data processing due to the extremely large hyperspectral image data. As hyperspectral image data contains both spectral and spatial information simultaneously, the hyperspectral imaging technology has some unique advantages compared with conventional spectral/imaging technologies (Cheng, Sun, Pu & Zhu, 2015; Pu & Sun, 2015; Shi, Zou, Zhao & Wang et al., 2012). In order to analyze the physical and/or chemical properties of the biological products, the whole surface of the individual items must be evaluated to achieve a full assessment. The hyperspectral imaging technique meets these requirements and it has been used to analyze chemical properties of various biological products successfully. Examples include Total acid content in vinegar (Zhu et al., 2016), chlorophyll determination in cucumber plants (Zou et al., 2011), moisture content in mango (Pu & Sun, 2015), and total volatile basic nitrogen contents in prawns (Dai, Cheng, Sun, Zhu & Pu, 2016).

Hyperspectral imaging data contains a spectrum with a specific wavelength range for each pixel in a 2-dimensional image of the sample. Research has demonstrated there is a good correlation between spectral data and chemical composition content in fruit (Kumar, McGlone, Whitworth & Volz, 2015), vegetables (Sridhar, Witter, Wu, Spongberg & Vincent, 2014), herbs (Saltas, Pappas, Daferera, Tarantilis & Polissiou, 2013) and other samples (Amneh & Mohammed, 2011; Lebot, Champagne, Malapa & Shiley, 2009; Shi et al., 2013; Ziemons et al., 2010). Moreover, published papers reported that Carbohydrates components in foxtail millet (Chen, Ren, Zhang, Diao & Shen, 2013), Proteins components in protein powder products (Ingle et al., 2016), and Sialic acid components (monosialotetrahexosyl) in medical injections (Ma et al., 2014)

could be determined by NIR. These studies indicated that there are characteristic absorbances in NIR region that could be used to determine Carbohydrates/Proteins/Sialic acid, and therefore it is possible to detect the distribution of major components (Carbohydrates/Proteins/Sialic acid) in the whole EBN sample using hyperspectral imaging technology.

The objectives of this study are to: (1) provide a rapid and nondestructive method for determining the distribution map of major components content (PC, CC and SAC) on EBN sample; (2) analyze the distribution tendency of components content (PC, CC and SAC) on the whole EBN sample.

2. Materials and methods

2.1 materials

60 white EBN samples were provided by the Edible Bird's Nest Market Committee of China Agricultural Wholesale Markets Association (EBMC).

2.2 hyperspectral image acquisition and pre-processing

A hyperspectral imaging system in the Vis/NIR (430-960nm) was used to image the EBN sample (Zou & Zhao, 2015). This system consisted of a linescan spectrograph (ImSpector, V10E, Spectra Imaging Ltd., Finland), a CMOS camera (Bci4-1300, C-Cam Ltd., Belgium), a standard C-mount lens, a DC illuminator (2900, Illumination Technologies Inc., USA), a conveyer (Zolix TS200AB, Zolix. Corp., China), an enclosure, a data acquisition and pre-processing software (SpectraCube, Auto Vision Inc., USA), and a PC as shown in Fig. 1. Based on the hyperspectral imaging system, a hyperspectral image of the EBN sample was acquired. The hyperspectral imaging data cube of EBN sample was shown in Fig. 2.

Fig.1 goes here

Fig.2 goes here

As shown in Fig. 2(a), the EBN hyperspectral image can be considered as a 3 dimension data cube. X axis and y axis are used to indicate the location of each pixel in hyperspectral data, and λ axis is used to indicate the wavelength of image signal (Zou et al., 2010). While x, y equal to a fixed value specified x_j, y_k ($1 \leq x_j \leq 1024, 1 \leq y_k \leq 618$), λ equals to any value available ($\lambda \in [430, 960]$), the data cube of a specified pixel is obtained as shown in Fig. 2(b). The intensity of the images vary according to their wavelength, signals in Fig. 2(b) are presented in a line chart, then the spectral data of the pixel (x_j, y_k) are obtained, as shown in Fig.2(c). While λ equal to a fixed value specified λ_i ($430 \leq \lambda_i \leq 960$), x, y equals to any value available ($x \in [1, 1024], y \in [1, 618]$), an EBN image at the specific wavelength λ_i is obtained. Therefore the EBN hyperspectral image combines conventional spectroscopy and imaging techniques (Zou, Shi, Min, Zhao, Mao, Chen, Li & Mel, 2011). It can be used to acquire both spectral and spatial information from an EBN sample, which makes it possible to determine the distribution of major components in the whole EBN sample.

2.3 extraction of spectral data from hyperspectral image

Hyperspectral imaging systems acquire abundant spatial information during the process of collecting spectral information. The hyperspectral data cube obtained from an EBN sample is shown in Fig. 1. The appropriate selection of a ROI for a sample image becomes critical and has profound impacts on the performance of prediction models (Zou, Shi, Min, Zhao, Mao, Chen, Li & Mel, 2011). In this study, the center part of EBN sample is defined as the location of the ROI (50×50pixels) in EBN hyperspectral image. The average intensity of ROI in images of the specified wavelengths (430-960nm) was extracted, so the raw spectra of EBN samples was obtained. Each spectra was smoothed with an 11 point mean filter and Standard Normal Variate (SNV) to eliminate variations in the baseline promoted by light scattering (Guo, Wu & Massart, 1999).

2.4 determination of PC, CC and SAC

Immediately after hyperspectral image acquisition, samples were used to determine the PC, CC and SAC. PC was determined by Kjeldahl's method, using 6.25 as a conversion factor (Saengkrajang, Matan & Matan, 2013). CC was determined by subtraction method, presented in Saengkrajang's paper with slight modification (Saengkrajang, Matan & Matan, 2013). CC was obtained by subtracting the percent of moisture, protein, fat, fibre and ash from total EBN mass ($CC = \text{Total mass} - \text{moisture} - \text{protein} - \text{fat} - \text{fibre} - \text{ash}$). Moisture content was determined by drying the EBN sample in an oven at 105°C until a constant weight was obtained (Saengkrajang, Matan & Matan, 2013). Fat content was calculated from a fraction of lipid extracted from the hydrolysed EBN sample (Wrolstad et al., 2005). Fibre was determined after digesting a known weight of a fat-free sample in refluxing 1.25% sulfuric acid and 1.25% sodium hydroxide (Saengkrajang, Matan & Matan, 2013). Ash contents were determined by dry ashing in a furnace at 550°C for 18 h (Saengkrajang, Matan & Matan, 2013).

SAC was determined by high performance liquid chromatography (Shimadzu Co., Kyoto, Japan) with ultraviolet detection (Hurum & Rohrer, 2012). EBN samples were dissolved in 0.5 mol/L sodium bisulfate aqueous solution and kept for 30 min in 80°C water bath. After cooling the derivatization was carried out using O-phenylenediamine 2HCl as derivative. The chromatographic separation was achieved on a ZORBAX SB-C₁₈ (4.6mm × 150mm, 5 μm; Sigma Chemical Company, St Louis, MO) column using a mobile phase composed of 1.0% tetrahydrofuran aqueous solution (containing phosphoric acid and 1- butylamine at the levels of 0.5% and 0.15%, respectively) and acetonitrile (95:5, V/V) at a flow rate of 1.0 mL/min in the isocratic elution mode. The column temperature was kept at 35 ± 0.5°C using a column oven. Sialic acid separations are detected by the SPD-20A UV-detector, which was set at 230 nm. Standards of the N-Acetylneuraminic acid (Neu5Ac,

Product#A0812, Sigma Chemical Company) were dissolved in water. The linear portions of the standard curves were used to convert the integrated areas to $\mu\text{g}/\text{mg}$ EBN weight.

2.5 Chemometrics methods

In this study, Genetic Algorithm-interval Partial Least Squares (GA-iPLS) and Genetic Algorithm- Partial Least Squares (GA-PLS) were used to select the most informative wavelengths correlated with PC/CC/SAC. PLS was used to build calibration models based on the selected wavelengths, and leave-one-out cross-validation (LOOCV) was employed to evaluate the established calibration model. The performance of the calibration models was back-evaluated according to the root mean square error of calibration (RMSEC), the root mean square error of cross-validation (RMSECV) and the correlation coefficient in the calibration set (R_c). The optimal model was also tested by an independent prediction set. The performance of optimal model for the prediction set was evaluated according to the root mean square error of prediction (RMSEP) and the correlation coefficient in the prediction set (R_p) (Shi, Zou, Zhao & Holmes et al., 2012; Zou, Zhao, Malcolm, Mel & Mao, 2010).

2.5.1 Genetic algorithm iPLS (GA-iPLS)

The GA-iPLS algorithm which combines the advantages of GA and PLS described in this paper was an evolution of the GA algorithm and the iPLS algorithm. The GA algorithm was used to select spectral regions, the PLS algorithm was used to establish regression model using the selected spectral regions, and leave-one-out cross-validation (LOOCV) was employed to evaluate the established calibration model. The combination of intervals with the lowest RMSECV was chosen. The GA-iPLS was repeated ten times in order to avoid its stochastic influence. Details of the GA-iPLS algorithm can be found in our published literatures (Shi, Zou, Zhao & Holmes, 2012).

2.5.2 Genetic algorithm PLS (GA-PLS)

In order to reduce the number of variables and simplify the calibration model, the GA-PLS algorithm was used to select most informative wavelengths correlated with PC/CC/SAC from those wavelength regions that selected by GA-iPLS. Therefore, the GA-PLS algorithm described in this paper was similar to the GA-iPLS algorithm. The GA algorithm was used to select spectral wavelengths from the specific wavelength regions, then the PLS algorithm was used to established regression model using the selected spectral regions. The GA-PLS was repeated ten times in order to avoid its stochastic influence. Details of the GA-PLS algorithm can be found in our published literatures (Zou, Zhao, Huang & Li, 2007).

2.6 Estimating major components distribution map

The main steps of estimating major components distribution map including: (1) building calibration models, (2) testing the calibration models, (3) estimating distribution map. The flow chart of determining PC/CC/SAC map on an EBN sample is shown in Fig. 3.

2.6.1 Building calibration models

As shown in Fig.3, 40 EBN samples in the calibration set were used to build PC, CC and SAC calibration models. Firstly, after hyperspectral image acquisition, components content (PC, CC and SAC) of ENB samples were determined by chemical analytical methods. Secondly, the center part of EBN sample is defined as the location of the ROI (50×50pixels) in EBN hyperspectral image. The average intensity of ROI in images of the specified wavelengths (430-960nm) was extracted, so the spectra of calibration set was obtained. Thirdly, GA-iPLS algorithm and GA-PLS algorithm were used to select most informative wavelengths correlated with PC/CC/SAC and build calibration models based on selected wavelengths. The whole spectrum was divided into 30 equidistant subintervals, the number of the generations is equal to 60, crossover probability (p_c) is equal to 0.50, mutation probability (p_m) is

equal to 0.05, a random population (population size 40, the average percentage of variables selected in the chromosomes of the starting population was 10%) was used as the initialized population.

2.6.2 Testing the calibration models

As shown in Fig.3, 20 EBN samples in the prediction set were used to test the PC/CC/SAC calibration models. Firstly, after hyperspectral image acquisition, components content (PC, CC and SAC) of EBN samples were determined by chemical analytical methods. Secondly, the spectral of prediction set was extracted according to the ROI that has been defined in section 2.6.1. Thirdly, the spectral of prediction set was substituted in the PC/CC/SAC calibration models to calculate the PC/CC/SAC of the prediction samples. Finally, the root mean square error of prediction (RMSEP) and the correlation coefficient in the prediction set (R_p) were used to evaluate the capability of the PC/CC/SAC calibration models, so the optimal PC/CC/SAC calibration model could be obtained.

2.6.3 Estimating distribution map

As we known, hyperspectral imaging data contains a spectrum with a specific wavelength range for each pixel in a 2-dimensional image of the sample. In the optimal PC/CC/SAC calibration model, a relationship between EBN spectra and PC/CC/SAC was defined. Therefore, it possible to estimate the distribution of major components in the whole EBN sample using hyper-spectral imaging technology.

Firstly, after hyperspectral image acquisition, spectra of all pixels was extracted from the hyperspectral of an EBN sample. Secondly, the spectral of each pixel was substituted in the PC/CC/SAC calibration models to estimate the PC/CC/SAC in each pixel of the EBN sample. Thirdly, the PC/CC/SAC of the pixels were displayed in two dimension spastically, then the distribution maps of PC/CC/SAC were obtained, as shown in Fig.3.

Fig.3 goes here

2.6 Software

All the image processing and data analysis procedures described above were executed using programs developed in Matlab 7.0 (MathWorks, Natick, MA, USA). Extraction of reflectance spectra from the hyper-spectral images was accomplished using ENVI 4.3 (ITT Visual Information Solutions, Boulder, CO, USA).

3. Results and discussion

3.1 PC, CC and SAC in EBN samples

As illustrated in Table 1, the descriptive statistics for the PC/CC/SAC in EBN samples were presented. The min values of the PC, CC and SAC was 470.12mg/g, 189.47 mg/g, and 83.35 mg/g for all EBN samples (including both calibration set and validation set). The max values of the PC, CC and SAC was 600.05 mg/g, 446.49 mg/g, and 113.58 mg/g for all EBN samples (including both calibration set and validation set). The 60 EBN samples were divided into a calibration set and a validation set. To avoid bias in subset selection, this division was made as follows: all samples had been sorted according to their respective y-value (viz. the reference measurement value of PC/CC/SAC). A 2/1 division of calibration/validation samples was chosen, thus two samples out of every three samples were randomly selected into the calibration set, so that the final calibration set contains 40 samples and the validation set contains 20 samples (Shi, Zou, Zhao, Holmes, Wang, Wang & Chen, 2012). The descriptive statistics for the PC, CC and SAC in calibration set and validation set were shown in Table 1.

Insert table 1 here

3.2 Estimating the distribution map of PC/CC/SAC

As shown in Fig.3, the main steps of estimating major components distribution map including: (1) building calibration models, (2) testing the calibration models, (3) estimating distribution map. In this section, relevant results and discussion to these steps were presented.

3.2.1 Building calibration models

After hyperspectral image acquisition, spectral data of calibration set were extracted from the hyperspectral images of 40 EBN samples in calibration set, as shown in Fig.3. Then spectra data and EBN components content (PC, CC and SAC) determined by chemical analytical methods were used to build calibration models. In order to obtain good and simple calibration models, GA-iPLS algorithm and GA-PLS algorithm were used to select most informative wavelengths correlated with PC/CC/SAC and build calibration models based on selected wavelengths.

Firstly, GA-iPLS was employed to select most informative spectral regions correlated with PC/CC/SAC, and calibration models based on selected spectral regions were shown in Table 2. 82 wavelengths were identified as the optimal wavelengths for PC prediction. Based on the optimal wavelengths, a PC calibration was built and yielded acceptable results ($R_c = 0.95$, $RMSEC = 2.02$ mg/g, $R_{cv} = 0.93$, $RMSECV = 2.29$ mg/g). 103 wavelengths were identified as the optimal wavelengths for CC prediction. Based on the optimal wavelengths, a CC calibration was built and yielded acceptable results ($R_c = 0.93$, $RMSEC = 15.78$ mg/g, $R_{cv} = 0.92$, $RMSECV = 16.10$ mg/g). 83 wavelengths were identified as the optimal wavelengths for SAC prediction. Based on the optimal wavelengths, a SAC calibration was built and yielded acceptable results ($R_c = 0.96$, $RMSEC = 0.38$ mg/g, $R_{cv} = 0.95$, $RMSECV = 0.39$ mg/g).

Secondly, In order to reduce the number of variables and simplify the calibration model, the GA-PLS algorithm was used to select most informative wavelengths

correlated with PC/CC/SAC from those wavelength regions that selected by GA-iPLS. Calibration models based on spectral wavelengths selected by GA-PLS were shown in Table 2. 12 wavelengths were identified as the optimal wavelengths for PC prediction. Based on the optimal wavelengths, a new PC calibration was built and yielded acceptable results ($R_c = 0.88$, $RMSEC = 3.28$ mg/g, $R_{cv} = 0.86$, $RMSECV = 3.49$ mg/g). 9 wavelengths were identified as the optimal wavelengths for CC prediction. Based on the optimal wavelengths, a new CC calibration was built and yielded acceptable results ($R_c = 0.90$, $RMSEC = 18.42$ mg/g, $R_{cv} = 0.89$, $RMSECV = 18.76$ mg/g). 10 wavelengths were identified as the optimal wavelengths for SAC prediction. Based on the optimal wavelengths, a new SAC calibration was built and yielded acceptable results ($R_c = 0.90$, $RMSEC = 0.43$ mg/g, $R_{cv} = 0.87$, $RMSECV = 0.51$ mg/g).

Insert table 2 here

3.2.2 Testing the calibration models

After hyperspectral image acquisition, spectral data of predication set were extracted from the hyperspectral images of 20 EBN samples in predication set, as shown in Fig.3. Then the spectral of prediction set was substituted in the calibration models based on GA-iPLS/GA-PLS to predicate the PC/CC/SAC of the EBN samples, and the root mean square error of prediction (RMSEP) and the correlation coefficient in the prediction set (R_p) were used to evaluate the capability of the PC/CC/SAC calibration models for predication set. Results of ‘Testing the calibration models’ were presented in Table 2. RMSEP of PC, CC and SAC based on GA-iPLS calibration models was 2.81 mg/g, 19.04 mg/g and 0.42 mg/g, respectively. R_p of PC, CC and SAC based on GA-iPLS calibration models was 0.90, 0.88 and 0.91, respectively. RMSEP of PC, CC and SAC based on GA-PLS calibration models was 3.51 mg/g,

19.30 mg/g and 0.54, respectively. Rp of PC, CC and SAC based on GA-PLS calibration models was 0.85, 0.86 and 0.85, respectively.

Table 2 summarizes the results of PC/CC/SAC calibration models developed by GA-iPLS and GA-PLS. By comparing the PC/CC/SAC calibration models developed by GA-iPLS, PC/CC/SAC calibration models developed by GA-PLS yielded results with lower PLS factors and higher RMSEC/RMSECV/RMSEP values. Results indicated that although accuracy of the GA-PLS models was decreased due to the higher RMSEC/RMSECV/RMSEP values; and dimensionality of GA-PLS models was reduced due to the lower PLS factors. Usually dimensionality reduction could be beneficial to develop a multispectral system for on/in-line application, and could also make the calibration models easier to interpret.

3.2.3 Estimating distribution map

A relationship between EBN spectra and PC/CC/SAC have been built by the established GA-PLS calibration models in section 3.2.1. Then these calibration models were used to estimate the PC/CC/SAC at each pixel of the EBN hyperspectral image. Distribution maps of PC/CC/SAC can be obtained by displaying PC/CC/SAC at all pixels as a 2D image, as shown in Fig.3. Fig.4 shows the distribution maps of PC/CC/SAC as predicted by the optimal GA-PLS calibration models. These distribution maps are coloured according to the band intensity indicating the relative PC/CC/SAC [mg/g]. With the PC/CC/SAC distribution maps, it is possible to observe the levels of PC/CC/SAC in the different regions directly. This highlights the advantages of hyperspectral imaging technology compared with point/region analysis technologies such as HPLC, GC, and near infrared spectroscopy.

PC distribution map in EBN sample was shown in Fig.4 (a), the PC of EBN sample was in the range of 450-650 mg/g, and is concentrated in 500-600 mg/g. CC distribution map in EBN sample was shown in Fig.4 (b), the CC of EBN sample was in the range of 200-450 mg/g, and is concentrated in 280-350 mg/g. SAC distribution

map in EBN sample was shown in Fig.4 (c), the SAC of EBN sample was in the range of 80-110 mg/g, and is concentrated in 85-95 mg/g. The acquired PC/CC/SAC distribution of EBN sample in general is in agreement with the results of chemical analysis (shown in Table 1) and the published paper (Ma & Liu, 2012).

PC and CC are not evenly distributed throughout the EBN sample as shown in Fig.4 (a) and Fig.4 (b). It could be also found that the areas with high PC in Fig.4 (a) correspond to low CC in Fig.4 (b), and the areas with low PC in Fig.4 (a) correspond to high CC in Fig.4 (b). These maybe resulting from the PC/CC/SAC in saliva of swiftlets. As we know, EBN is made from the saliva of swiftlets (Ma & Liu, 2012). Usually, the construction process of an edible bird's nest may take the birds about 35 days (Marcone, 2005). During this period, daily consumption of food may affect the protein content and carbohydrate content in the saliva of swiftlets. The consumption of insects and small fish increase PC in the saliva of swiftlets, while the consumption of seaweed increase CC in the saliva of swiftlets. Fig.4 (c) shows that SAC is distributed evenly on the EBN sample. Sialic acids are a family of nine-carbon acidic monosaccharides that occur naturally at the end of sugar chains attached to the surfaces of cells and soluble proteins (Wang & Brand-Mille, 2003). Sialic acid disorders will cause serious problems to humans and animals (Sillanaukee, Pönniö & Jääskeläinen, 1999). Therefore sialic acid content in saliva of swiftlets remains relatively stable level, which is the main reason why sialic acid component of the EBN sample is distributed evenly. According to the published papers, Sialic acids have been identified as one of the special nutrient components in EBNs due to its anti-virus and immune-enhancing properties (Guo, Takahashi, Bukawa, Takahashi, Yagi, Kato, Hidari, Miyamoto, Suzuki & Suzuki, 2006). Fig.4 (c) indicates that nutrient values of different parts of EBN sample are the same in the aspects of anti-virus and immune-enhancing properties.

Fig.4 goes here

EBN samples with different Carbohydrates/Proteins/Sialic acid content can cause changes in the reflectance of hyperspectral image, this is the fundamental reason why Carbohydrates/Proteins/Sialic acid content can be predicated by hyperspectral data. However, EBN samples with different shape and surface angle can also cause changes in the reflectance of hyperspectral image. Changes in the reflectance caused by sample shape or surface result in errors in predicating Carbohydrates/Proteins/Sialic acid content. In order to eliminate the negative effects of sample shape and surface angle, some measures were employed in hyperspectral image collection. (1) EBN is in its natural round shaped form, resembling the shape of “cupped hand”. Unlike samples with flat surface, the shape and surface angle of EBN samples can affect the quality of hyperspectral image acquisition. Usually, in order to obtain high quality images, the height of a sample should be less than the Depth of field (DOF) of an imaging system. In this study, A proper position with low height (EBN samples were put on a platform with its big side down the ground) was chosen for EBN samples during hyperspectral image collection. So the height (4-6 cm) of EBN samples is lower than DOF (10 cm) of hyperspectral imaging system. (2) At the same time, EBN surface angle and light source (type, distribution) decide the directions of reflected light. We noticed that EBN is dome-shaped. In order to reduce the effects of surface angle, two directional lights (45° and 135°) in symmetric distribution were employed as lighting source system during hyperspectral image collection.

4. Conclusions

In this paper, PC/CC/SAC and its distribution on the whole EBN sample were determined using hyperspectral imaging. The results presented illustrate that hyperspectral imaging is a powerful tool for PC/CC/SAC analysis in EBN sample. After hyperspectral image acquisition and pre-processing, average spectra obtained

from the ROI of EBN images were used for calibration model development. GA-iPLS, GA-PLS were constructed for the prediction of the PC/CC/SAC. When the calibration model was applied to an independent validation set, PC/CC/SAC was reasonably well predicted ($R_p = 0.85, 0.86, 0.85$). Application of the calibration models to the spectra of each pixel in hyperspectral image enabled the PC/CC/SAC distribution map to be estimated. The acquired PC/CC/SAC distribution of EBN sample in general is in agreement with the results based on chemical analysis. With the PC/CC/SAC distribution maps, it is possible to observe the levels of PC/CC/SAC in the different regions directly.

Acknowledgements

The authors gratefully acknowledge the financial support provided by the national science and technology support program (2015BAD17B04, 2015BAD19B03, 2016YFD0401104), the national natural science foundation of China (61301239), the natural science foundation of Jiangsu province (BK20130505), China postdoctoral science foundation (2013M540422, 2014T70483), the Jiangsu province science fund for distinguished young scholars (BK20130010), Science foundation for postdoctoral in Jiangsu province (1301051C), Suzhou science and technology project (SNG201503), Research foundation for advanced talents in Jiangsu University (13JDG039), Priority Academic program development of Jiangsu higher education institutions (PAPD).

References

- Amneh, A., & Mohammed, B. (2011). Determination of glucose concentration from NIR spectra using independent component regression. *Chemometrics and Intelligent Laboratory Systems*, 105(1), 131-135.
- Cen, H., & Lu, R. (2010). Optimization of the hyperspectral imaging-based spatially-resolved system for measuring the optical properties of biological materials. *Optics Express*, 18(16), 17412-17432.
- Chen, J., Ren, X., Zhang, Q., Diao, X. M., & Shen, Q. (2013). Determination of protein, total carbohydrates and crude fat contents of foxtail millet using effective wavelengths in NIR spectroscopy. *Journal of Cereal Science*, 58(2), 241-247.
- Cheng, J., Sun, D., Pu, H., & Zhu, Z. (2015). Development of hyperspectral imaging coupled with chemometric analysis to monitor K value for evaluation of chemical spoilage in fish fillets. *Food Chemistry*, 185, 245-253.
- Chua, K., Lee, T., Nagandran, K., Yahaya, N. H. M., Lee, C., Tjih, E. T. T., & Aziz, R. A. (2013). Edible Bird's nest extract as a chondro-protective agent for human chondrocytes isolated from osteoarthritic knee: in vitro study. *BMC Complementary and Alternative Medicine*, 13(1), 19-25.
- Chua, Y. G., Chan, S. H., Bloodworth, B. C., Li, S. F. Y., & Leong, L. P. (2015). Identification of Edible Bird's Nest with Amino Acid and Monosaccharide Analysis. *Journal of Agricultural and Food Chemistry*, 63(1), 279-289.
- Dai, Q., Cheng, J., Sun, D., Zhu, Z., & Pu, H. (2016). Prediction of total volatile basic nitrogen contents using wavelet features from visible/near-infrared hyperspectral images of prawn (*Metapenaeus ensis*). *Food Chemistry*, 197(A), 257-265.
- Deng, Y., Sun, S., Zhou, Q., & Li, A. (2006). Analysis and discrimination of *Collocalia esculenta* L. via FTIR spectroscopy. *Spectroscopy and Spectral Analysis*, 26(7), 1242-1245.
- Guo, C. T., Takahashi, T., Bukawa, W., Takahashi, N., Yagi, H., Kato, K., Hidari, K., Miyamoto, D., Suzuki, T., & Suzuki, Y. (2006). Edible bird's nest extract inhibits influenza virus infection. *Antiviral Research*, 70(3), 140-146.
- Guo, Q., Wu, W., & Massart, D. L. (1999). The robust normal variate transform for pattern recognition with near-infrared data. *Analytica Chimica Acta*, 382(1-2), 87-103.
- Hurum, D. C., & Rohrer, J. S. (2012). Determination of sialic acids in infant formula by chromatographic methods: A comparison of high-performance anion-exchange chromatography with pulsed amperometric detection and ultra-high-performance liquid chromatography methods. *Journal of Dairy Science*, 95(3), 1152-1161.
- Ingle, P. D., Christian, R., Purohit, P., Zarraga, V., Handley, E., Freel, K., & Abdo, S. (2016). Determination of Protein Content by NIR Spectroscopy in Protein Powder Mix Products. *Journal of Aoac International*, 99(2), 360-363.
- Kamruzzaman, M., Makino, Y., & Oshita, S. (2016). Rapid and non-destructive detection of chicken adulteration in minced beef using visible near-infrared hyperspectral imaging and machine learning. *Journal of Food Engineering*, 170, 8-15.
- Kim, K. C., Kang, K. A., Lim, C. M., Park, J. H., Jung, K. S., & Hyun, J. W. (2012). Water extract of edible bird's nest attenuated the oxidative stress-induced matrix metalloproteinase-1 by regulating the mitogen-activated protein kinase and activator protein-1 pathway in human keratinocytes. *Journal of the Korean*, 55, 347-354.
- Konda Naganathan, G., Cluff, K., Samal, A., Calkins, C. R., Jones, D. D., Meyer, G. E., & Subbiah, J. (2016). Three dimensional chemometric analyses of hyperspectral images for beef tenderness

- forecasting. *Journal of Food Engineering*, 169, 309-320.
- Kumar, S., McGlone, A., Whitworth, C., & Volz, R. (2015). Postharvest performance of apple phenotypes predicted by near-infrared (NIR) spectral analysis. *Postharvest Biology and Technology*, 100, 16-22.
- Lau, A. S. M., & Melville, D. S. (1994). *International Trade in Swiftlet Nests with Special Reference to Hong Kong*. Cambridge (UK): Traffic Network.
- Lebot, V., Champagne, A., Malapa, R., & Shiley, D. (2009). NIR Determination of Major Constituents in Tropical Root and Tuber Crop Flours. *Journal of Agricultural and Food Chemistry*, 57(22), 10539-10547.
- Ma, F., & Liu, D. (2012). Sketch of the edible bird's nest and its important bioactivities. *Food Research International*, 48, 559-567.
- Ma, J., Yin, L., Ding, Y., Xiao, X., Ma, S., Ren, L., & Yang, X. (2014). Construction of Characteristic Spectral Correlation Coefficient Model for Monosialotetrahexosyl Ganglioside Sodium Injections by Near-Infrared Spectroscopy. *China Pharmacist*, 17(4), 537-540.
- Marcone, M. F. (2005). Characterization of the edible bird's nest the "Caviar of the East". *Food Research International*, 38(10), 1125-1134.
- Pu, Y., & Sun, D. (2015). Vis-NIR hyperspectral imaging in visualizing moisture distribution of mango slices during microwave-vacuum drying. *Food Chemistry*, 188, 271-278.
- Saengkrajang, W., Matan, N., & Matan, N. (2013). Nutritional composition of the farmed edible bird's nest (*Collocalia fuciphaga*) in Thailand. *Journal of Food Composition and Analysis*, 31(1), 41-45.
- Saltas, D., Pappas, C. S., Daferera, D., Tarantilis, P. A., & Polissiou, M. G. (2013). Direct Determination of Rosmarinic Acid in Lamiaceae Herbs Using Diffuse Reflectance Infrared Fourier Transform Spectroscopy (DRIFTS) and Chemometrics. *Journal of Agricultural and Food Chemistry*, 61(13), 3235-3241.
- Shi, J. Y., Zou, X. B., Zhao, J. W., Holmes, M., Wang, K. L., Wang, X., & Chen, H. (2012). Determination of total flavonoids content in fresh Ginkgo biloba leaf with different colors using near infrared spectroscopy. *Spectrochimica Acta Part A-Molecular and Biomolecular Spectroscopy*, 94, 271-276.
- Shi, J. Y., Zou, X. B., Zhao, J. W., Wang, K. L., Chen, Z. W., Huang, X. W., Zhang, D. T., & Holmes, M. (2012). Nondestructive diagnostics of nitrogen deficiency by cucumber leaf chlorophyll distribution map based on near infrared hyperspectral imaging. *Scientia Horticulturae*, 138, 190-197.
- Shi, J. Y., Zou, X. B., Zhao, J. W., & Holmes, M. (2012). Near infrared quantitative analysis of total flavonoid content in fresh Ginkgo biloba leaves based on different wavelength region selection methods and partial least squares regression. *Journal of Near Infrared Spectroscopy*, 20(2), 295-305.
- Shi, J., Zou, X., Huang, X., Zhao, J., Li, Y., Hao, L., & Zhang, J. (2013). Rapid detecting total acid content and classifying different types of vinegar based on near infrared spectroscopy and least-squares support vector machine. *Food Chemistry*, 138(1), 192-199.
- Sillanauke, Pönniö, & Jääskeläinen. (1999). Occurrence of sialic acids in healthy humans and different disorders. *European Journal of Clinical Investigation*, 29(5), 413-425.
- Sridhar, B. B. M., Witter, J. D., Wu, C., Spongberg, A. L., & Vincent, R. K. (2014). Effect of Biosolid Amendments on the Metal and Nutrient Uptake and Spectral Characteristics of Five Vegetable Plants. *Water Air and Soil Pollution*, 225(9), 1-14.

- Vimala, B., Hussain, H., & Nazaimoon, W. M. W. (2012). Effects of edible bird' s nest on tumour necrosis factor-alpha secretion, nitric oxide production and cell viability of lipopolysaccharide-stimulated RAW 264.7 macrophages. *Food and Agricultural Immunology*, 23(4), 303-314.
- Wang, B., & Brand-Mille, J. (2003). The role and potential of sialic acid in human nutrition. *European Journal of Clinical Nutrition*, 57, 1351-1369.
- Wrolstad, R. E., Acree, T. E., Decker, E. A., Penner, M. H., Reid, D. S., Schwartz, S. J., Shoemaker, C. F., Smith, D. M., & Sporns, P. (2005). *Handbook of Food Analytical Chemistry, Volume 1: Water, Proteins, Enzymes, Lipids, and Carbohydrates*. New Jersey: John Wiley & Sons, Inc.
- Zhu, Y., Zou, X., Shen, T., Shi, J., Zhao, J., Holmes, M., & Li, G. (2016). Determination of total acid content and moisture content during solid-state fermentation processes using hyperspectral imaging. *Journal of Food Engineering*, 174, 75-84.
- Ziemons, E., Mantanus, J., Lebrun, P., Rozet, E., Evrard, B., & Hubert, P. (2010). Acetaminophen determination in low-dose pharmaceutical syrup by NIR spectroscopy. *Journal of Pharmaceutical and Biomedical Analysis*, 53(3), 510-516.
- Zou, X. B., Shi, J. Y., Min, H. L., Zhao, J. W., Mao, H. P., Chen, Z. W., Li, Y. X., & Mel, H. (2011). In vivo noninvasive detection of chlorophyll distribution in cucumber (*Cucumis sativus*) leaves by indices based on hyperspectral imaging. *Analytica Chimica Acta*, 706, 105-112.
- Zou, X. B., Zhao, J. W., Holmes, M., Mao, H. P., Shi, J. Y., Yin, X. P., & Li, Y. X. (2010). Independent component analysis in information extraction from visible/near-infrared hyperspectral imaging data of cucumber leaves. *Chemometrics and Intelligent Laboratory Systems*, 104(2), 265-270.
- Zou, X. B., Zhao, J. W., Huang, X. Y., & Li, Y. X. (2007). Use of FT-NIR spectrometry in non-invasive measurements of soluble solid contents (SSC) of 'Fuji' apple based on different PLS models. *Chemometrics and Intelligent Laboratory Systems*, 87(1), 43-51.
- Zou, X. B., Zhao, J. W., Malcolm, P. J. W., Mel, H., & Mao, H. (2010). Variables selection methods in near-infrared spectroscopy. *Analytica Chimica Acta*, 667(1-2), 14-32.
- Zou, X., & Zhao, J. (2015). *Nondestructive Measurement in Food and Agro-products*: Springer.

Figure 1 Hyperspectral imaging system

Figure 2 EBN hyperspectral image data cube

Figure 3 Process flowchart for estimating PC/CC/SAC content distribution in EBN sample

Figure 4 Distribution maps of protein content (a), carbohydrate content (b) and sialic acid content (c)

ACCEPTED MANUSCRIPT

Table 1 descriptive statistics for the PC, CC, and SAC in EBN samples

components	group	number	Min (mg/g)	Mean (mg/g)	Max (mg/g)	SD (mg/g)
PC	All samples	60	470.12	568.82	600.05	15.43
	Calibration	40	470.12	568.33	600.05	15.47
	Validation	20	480.02	569.79	596.19	15.63
CC	All samples	60	189.47	347.12	446.49	57.78
	Calibration	40	189.47	344.46	446.49	57.35
	Validation	20	200.08	352.45	441.85	60.64
SAC	All samples	60	83.35	100.66	113.58	5.41
	Calibration	40	83.35	100.45	113.58	5.51
	Validation	20	90.27	101.07	112.62	5.29

Table 2 performance of PC, CC, and SAC models based on GA-iPLS and GA-PLS

models	component	Number of wavelength	PLS factors	RPD value	Calibration		Cross-validation		Prediction	
					Rc	RMSEC	Rcv	RMSECV	Rp	RMSEP
GA-iPLS	PC	82	15	5.56	0.95	2.02	0.93	2.29	0.90	2.81
	CC	103	12	3.18	0.93	15.78	0.92	16.10	0.88	19.04
	SAC	83	17	12.60	0.96	0.38	0.95	0.39	0.91	0.42
GA-PLS	PC	12	7	4.45	0.88	3.28	0.86	3.49	0.85	3.51
	CC	9	6	3.14	0.90	18.42	0.89	18.76	0.86	19.30
	SAC	10	5	9.80	0.90	0.43	0.87	0.51	0.85	0.54

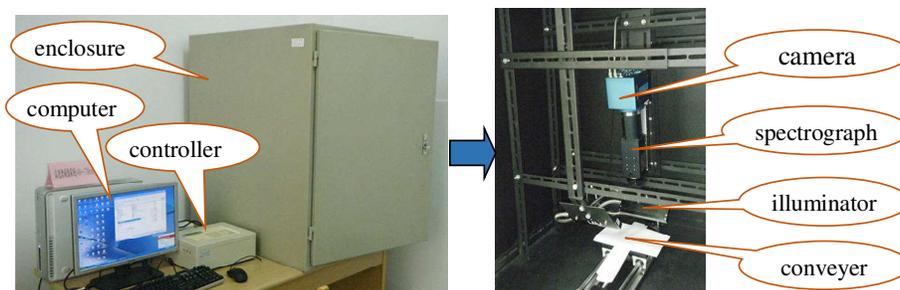


Figure 1

ACCEPTED MANUSCRIPT

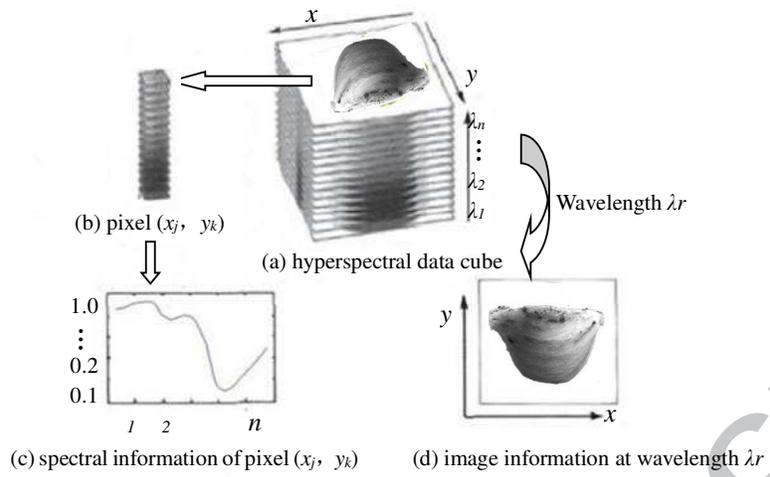
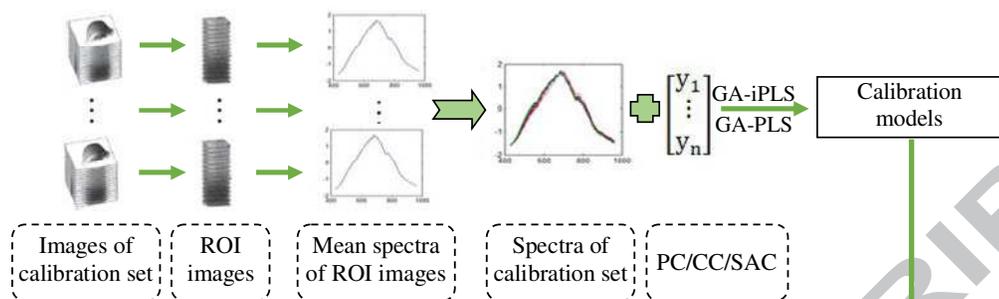
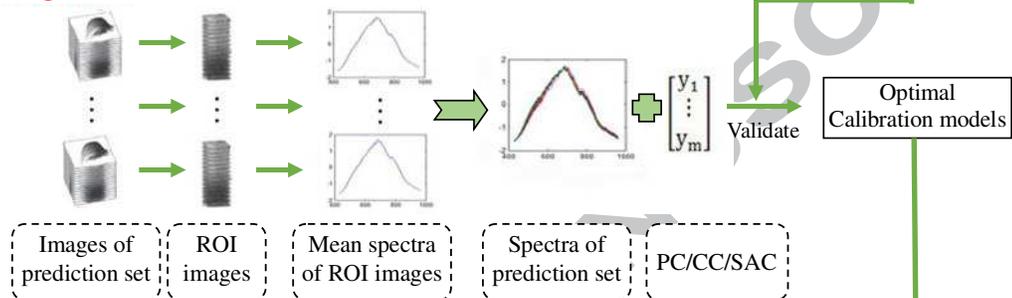


Figure 2

Building calibration models



Testing the calibration models



Estimating distribution map

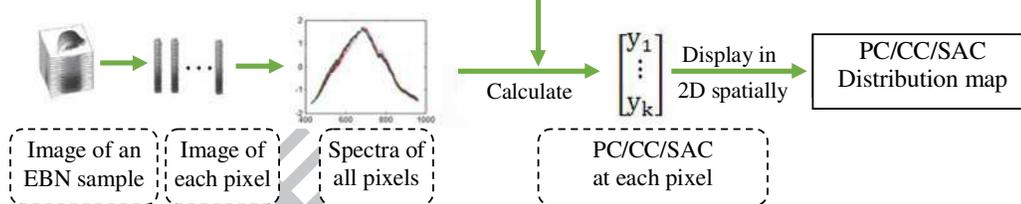


Figure 3

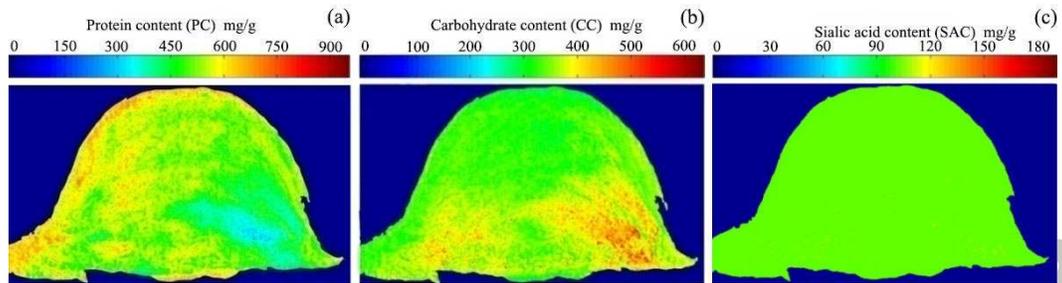


Figure 4

Highlights

- A rapid and nondestructive method for determining the distribution of EBN

components were first proposed.

- Distribution maps of three EBN components content (PC, CC and SAC) were determined.

- Distribution tendency of PC, CC and SAC on the whole EBN sample was first analyzed.

ACCEPTED MANUSCRIPT