



# Local anomaly detection and quantitative analysis of contaminants in soybean meal using near infrared imaging: The example of non-protein nitrogen

Guanghai Shen<sup>a,b</sup>, Juan Antonio Fernández Pierna<sup>b</sup>, Vincent Baeten<sup>b</sup>, Yaoyao Cao<sup>a</sup>, Lujia Han<sup>a</sup>, Zengling Yang<sup>a,\*</sup>

<sup>a</sup> College of Engineering, China Agricultural University, Beijing 100083, PR China

<sup>b</sup> Walloon Agricultural Research Centre (CRA-W), Valorisation of Agricultural Products Department, Chaussée de Namur 24, 5030 Gembloux, Belgium

## ARTICLE INFO

### Article history:

Received 26 April 2019

Received in revised form 20 August 2019

Accepted 28 August 2019

Available online 29 August 2019

### Keywords:

Untargeted detection

Near-infrared spectroscopy

Soybean meal

Local anomaly detection

Near-infrared hyperspectral/microscopic imaging

## ABSTRACT

The melamine scandal indicates that traditional targeted detection methods only detect the specifically listed forms of contamination, which leads to the failure to identify new adulterants in time. In order to deal with continually changing forms of adulterations in food and feed and make up for the inadequacy of targeted detection methods, an untargeted detection method based on local anomaly detection (LAD) using near infrared (NIR) imaging was examined in this study. In the LAD method, with a particular size of window filter and at a 99% level of confidence, a specific value of Global H ( $GH$ , modified Mahalanobis distance) can be used as a threshold for anomalous spectra detection and quantitative analysis. The results showed an acceptable performance for the detection of contaminations with the advantage of no need of building a 'clean' library. And, a high coefficient of determination ( $R_{LAD}^2 = 0.9984$  and  $R_{PLS-DA}^2 = 0.9978$ ) for the quantitative analysis of melamine with a limit of detection lower than 0.01% was obtained. This indicates that the new strategy of untargeted detection has the potential to move from passive to active for food and feed safety control.

© 2019 Elsevier B.V. All rights reserved.

## 1. Introduction

Food and feed safety is closely related to the quality of life and physical health of human beings. Consumers have been paying more and more attention to food and feed safety following numerous safety scandals in recent years. In 2007, pet food and animal feed contamination by melamine was confirmed and reported by US scientists, leading to the sickness and death of numerous pets (dogs and cats) [1]. In 2008, a more serious milk powder contamination scandal occurred in China, killing six infants and making more than 52,000 children ill [2,3]. Subsequently, melamine was detected in many food and feed products, which were also being exported to many countries worldwide. In addition, many researchers had shown that melamine contamination in milk and milk products might be related not only to direct external additions, but also to adulteration in animal feeds [4–7]. To ensure food and feed safety, many countries set the maximum permitted concentration for melamine in food and feed products at 2.5 mg/kg [8].

After the melamine scandal, various targeted analysis methods based on different platforms for detecting melamine were developed quickly. These included chromatography-based methods such as high performance liquid chromatography (HPLC) [9], gas-chromatography mass-spectrometry (GC-MS) [10] and LC-MS-MS [11], enzyme linked immunosorbent assays (ELISA) [12]; optical sensing technologies such as near- and mid-infrared (NIR/MIR) spectroscopy [13,14], Raman spectroscopy [15,16] and near-infrared and Raman imaging [17–20]. However, adulterations with new, unknown illegal ingredients continue to occur from time to time. If it relies on routine targeted methods, analysis will be trapped in a cycle of adulterations, followed by targeted analysis, followed by new adulterations, and the adulterations could spiral out of control [21,22]. A review of the whole melamine scandal shows that one main contributory factor was that melamine was not specifically listed as an illegal additive, which underscores the vulnerability and deficiencies of targeted methods. A new strategy of untargeted detection methods, which can respond quickly to suspicious contaminations by screening out all anomalous spectra, is therefore urgently required to deal with the ever-changing pattern of adulteration in food and feed products, in order to keep people from harm [23].

In recent years, untargeted detection methods based on NIR spectroscopy have shown great application potential for detecting contaminants in agro-food products. In 2012, Moore et al. [23] proposed to

\* Corresponding author at: China Agricultural University, No. 17 Qinghua Donglu Haidian District, Beijing 100083, PR China.

E-mail addresses: [shenguanghai@cau.edu.cn](mailto:shenguanghai@cau.edu.cn) (G. Shen), [j.fernandez@cra.wallonie.be](mailto:j.fernandez@cra.wallonie.be) (J.A. Fernández Pierna), [v.baeten@cra.wallonie.be](mailto:v.baeten@cra.wallonie.be) (V. Baeten), [caoyaoao@cau.edu.cn](mailto:caoyaoao@cau.edu.cn) (Y. Cao), [hanlj@cau.edu.cn](mailto:hanlj@cau.edu.cn) (L. Han), [yangzengling@cau.edu.cn](mailto:yangzengling@cau.edu.cn) (Z. Yang).

standardize the non-targeted screening method for detecting adulterations in skim milk powder using NIR spectroscopy and chemometrics; in 2013, one-class partial least squares (OCPLS) classifier [24] combined with NIR spectroscopy was investigated as a tool for untargeted detection of illegal adulterations in Chinese glutinous rice flour (GRF) [22], Chinese propolis [25] and yogurt [21]; in 2015, Fernández Pierna et al. [26] used multivariate moving window PCA for the untargeted detection of contaminants in milk using vibrational spectroscopy; in 2016, Shen et al. [19] proposed a non-targeted adulterant screening method based on the NIR microscopic spectral library of soybean meal to guarantee the quality of soybean meal, and in 2017, Fu et al. developed an untargeted screening method to detect maleic acid in cassava starch by Fourier transform near-infrared spectroscopy [27]. All the mentioned untargeted detection methods needed to collect representative 'clean' samples in order to set up a reference library or build statistical models based on the measured fingerprints, the unknown new sample could then be characterized and its purity estimated; however, for most of the mentioned untargeted methods it was almost impossible to quantify the amount of adulterants.

In the present study, a novel untargeted adulterant screening method based on local anomaly detection (LAD) by near-infrared (NIR) imaging is examined, which has a significant difference with our previous study [19] where the method needed a 'clean' and representative spectra library. The main drawback of that methodology was that sometimes it was not possible to build such library, and then the method failed. The method proposed in this study can also screen out the suspicious spectra without having a spectra library previously built. In this work, soybean meal has been used as one most of the important cereal products and in order to deal with the ever-changing pattern of adulteration in food and feed products. It is nearly impossible to make an exhaustive analysis of all potential adulterants in food and feed using targeted detection methods with the ever-changing pattern of adulteration. To a certain extent, methods based on vibrational spectroscopy as the NIR microscopy and hyperspectral imaging are more suitable than GC or LC-MS techniques for untargeted detection, especially for solid samples. To investigate the performance of the LAD method in distinguishing adulterants and quantitative analysis, one sample with soybean meal and melamine arranged in special shape, 21 adulterated samples (prepared using three kinds of yogurt soybean meal and six kinds of non-protein nitrogen) and 7 adulterated soybean meal samples with different percentages of melamine were prepared, respectively. The melamine concentrations selected were well above the common international limit of 2.5 mg/kg. The choice of a higher range was motivated by the fact that deliberate adulteration of feed/food with melamine should be economically viable and then mass contamination is expected. For ensuring the proposed LAD method could be used for analyzing the images acquired by different instruments, two different NIR imaging systems were used to collect spectral images in this study. All the adulterated samples were also analyzed by a targeted detection method based on partial least squares discriminant analysis (PLS-DA) and the results were compared.

## 2. Materials and methods

### 2.1. Samples collection

Three different soybean meal samples (hulls, full-fat and de-hulled soybean meal) were selected from the sample library at the China Agricultural University, one pure soybean meal sample was collected from the sample library of the Valorisation of Agricultural Products Department of the Walloon Agricultural Research Centre (CRA-W). Melamine, cyanuric acid, urea, and di-ammonium phosphate (DAP) were purchased from the Sinopharm Chemical Reagent Co., Ltd. (Shanghai, China). Biuret was bought from the Tianjin Fuchen chemical reagents factory (Tianjin, China). Mono-ammonium phosphate (MAP) was purchased from Beijing Chemical Reagent Company (Beijing, China).

### 2.2. Sample preparation

All the soybean meal samples were ground until they could pass through a 0.5-mm square mesh using a Retsch mill (Ultra centrifugal Mill ZM100; Retsch GmbH, Haan, Germany). Three different sample sets (A, B and C) were prepared in this study. One sample (set A) consisted of pure soybean meal and melamine arranged in a special shape, as indicated in Fig. 1A, was used to verify the LAD method. Set B: Two sets of 21 adulterated samples were prepared as described in Table 1, in set 1, each soybean meal sample was adulterated with one type of non-protein nitrogen at 0.5% (w/w), and in set 2, each soybean meal sample was adulterated with six types of non-protein nitrogen simultaneously, each at 0.5% (w/w). Set C: Seven adulterated soybean meal samples with different concentrations of melamine (0.01%, 0.05%, 0.1%, 0.5%, 1.0%, 1.5% and 2.0% (w/w)) were prepared using the soybean meal sample from the CRA-W, as shown in Table 2. All the samples of Set B and C were prepared in the laboratory using a mixer (REAX 20/8; Heidolph, Schwabach, Germany). In order to obtain sufficient homogeneity, we prepared the various materials by stepwise, diluting the contaminant with the soybean meal so that, in each dilution step, the ratio of the 2 different materials to be mixed did not exceed a factor of 3 [28].

### 2.3. Apparatus and data collection

Two different NIR imaging systems were used in this work. All the samples of set A and B were scanned using a Fourier transform infrared (FT-IR) imaging system (Spotlight 400 FTIR Imaging system, PerkinElmer Ltd., Beaconsfield, Bucks, UK). The instrument was equipped with a mercury cadmium telluride (MCT) line array detector, allowing the collection of spectra from arbitrary spatial regions with a high spectral resolution. A CCD detector collected a visible image from the same spatial region as the NIR microscopic image. Each sample was spread on the Teflon Spectralon plate with a height of approximately 1 mm, and then the sample surface was smoothed. The NIR microscopic image scanning area of the set A was  $5000\ \mu\text{m} \times 5000\ \mu\text{m}$ , the scanning areas of three pure samples (hulls, full-fat and de-hulled soybean meal), set 1 and set 2 from set B were of  $5000\ \mu\text{m} \times 5000\ \mu\text{m}$ ,  $5000\ \mu\text{m} \times 5000\ \mu\text{m}$  and  $8750\ \mu\text{m} \times 8750\ \mu\text{m}$ , respectively. The spatial resolution was  $25\ \mu\text{m} \times 25\ \mu\text{m}$ , and each spectrum was computed at  $32\ \text{cm}^{-1}$  resolution across the wavenumber range  $7800\text{--}4000\ \text{cm}^{-1}$  by combining 8 scans.

All the samples of set C and one melamine sample were scanned using a line-scan NIR hyperspectral or push-broom imaging system from BurgerMetrics SIA (Riga, Latvia) combined with a conveyor belt, for better and faster image acquisition. The acquisition was carried out using HyperProVB software from Burger Metrics SIA. In order to get high quality of images, the sample was spread on the conveyor belt to have a smooth surface. The conveyor belt speed was fixed at  $145\ \text{m s}^{-1}$ , and the spatial resolution was  $30\ \mu\text{m}$ , using a macro-lens at  $4\times$  magnification. All the images consisted of 200 lines of 320 pixels acquired at 209 wavelength channels ( $1100\text{--}2400\ \text{nm}$  at  $6.3\ \text{nm}$  intervals; i.e. 64,000 spectra by image). Each sample was divided into three parts and seven images were scanned for each part, then 21 images were obtained for each sample in total. The surface of all the scanned samples should be flat.

### 2.4. Local anomaly detection (LAD) method and judgment criteria

In this study, the LAD method was applied to detect suspicious non-soybean meal ingredients in the NIR microscopic images. The LAD method used a continuous-level sliding local window filter, which moved one pixel at a time from the top left corner to the bottom right corner, to analyze every pixel in the NIR microscopic image. In each local window, the pixel under test (the central pixel in the window) was regarded as a new unknown sample and the neighboring pixels

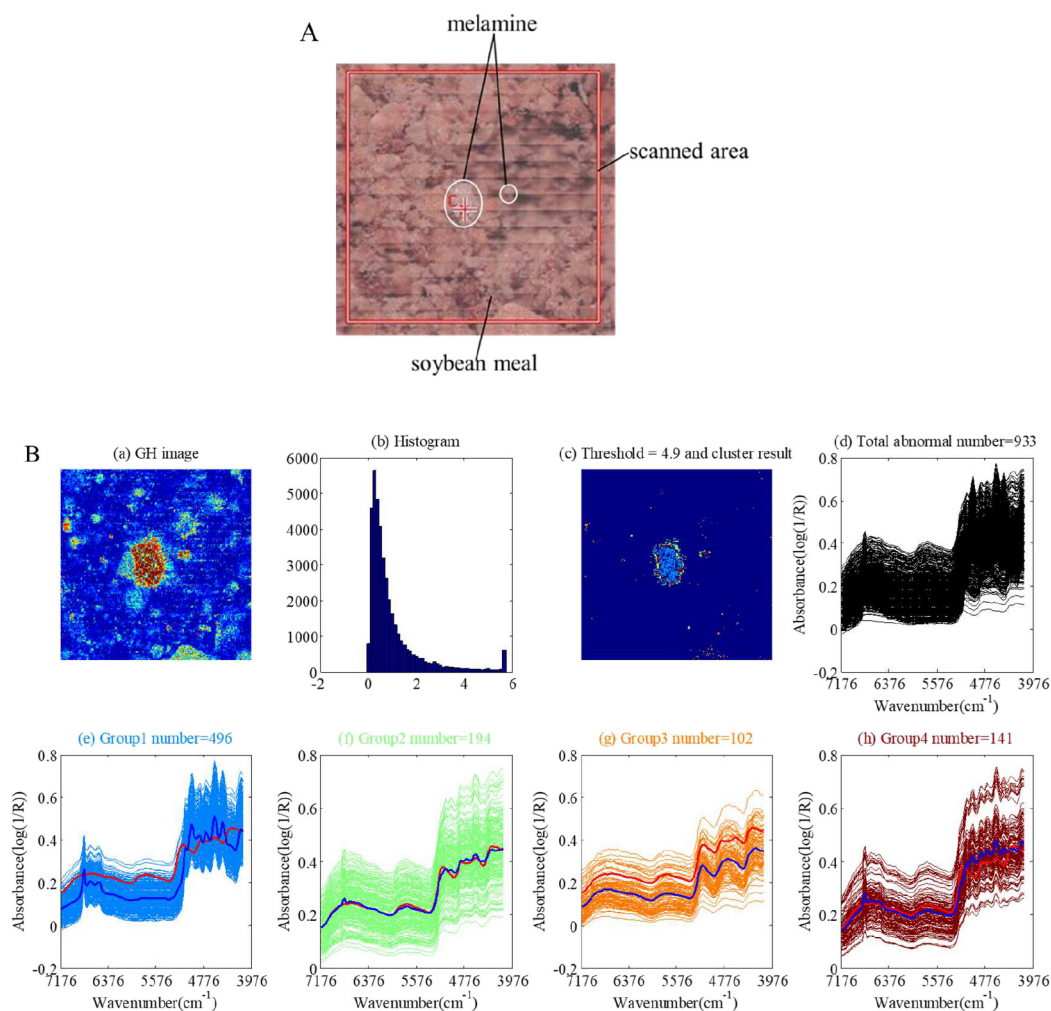


Fig. 1. (A) Preparation of sample set A, melamine was arranged in the white area in the image, (B) the detection results of sample set A by LAD method.

Table 1

Description of the sample set A.

Sample ID	Mixing amount (% w/w)								
	De-hulled	Full-fat	Hulls	Urea	Melamine	MAP <sup>a</sup>	DAP <sup>b</sup>	Cyanuric acid	Biuret
Set 1	1	99.5	–	–	0.5	–	–	–	–
	2	99.5	–	–	–	0.5	–	–	–
	3	99.5	–	–	–	–	0.5	–	–
	4	99.5	–	–	–	–	–	0.5	–
	5	99.5	–	–	–	–	–	–	0.5
	6	99.5	–	–	–	–	–	–	–
	7	–	99.5	–	0.5	–	–	–	–
	8	–	99.5	–	–	0.5	–	–	–
	9	–	99.5	–	–	–	0.5	–	–
	10	–	99.5	–	–	–	–	0.5	–
	11	–	99.5	–	–	–	–	–	0.5
	12	–	99.5	–	–	–	–	–	–
	13	–	–	99.5	0.5	–	–	–	–
	14	–	–	99.5	–	0.5	–	–	–
Set 2	15	–	–	99.5	–	–	0.5	–	–
	16	–	–	99.5	–	–	–	0.5	–
	17	–	–	99.5	–	–	–	–	0.5
	18	–	–	99.5	–	–	–	–	–
	19	97	–	–	0.5	0.5	0.5	0.5	0.5
	20	–	97	–	0.5	0.5	0.5	0.5	0.5
	21	–	–	97	0.5	0.5	0.5	0.5	0.5

<sup>a</sup> MAP: mono-ammonium phosphate.

<sup>b</sup> DAP: di-ammonium phosphate.

**Table 2**  
Description of the sample set B.

Sample ID	Mixing amount (% w/w)	
	Soybean meal	Melamine
1	99.99	0.01
2	99.95	0.05
3	99.90	0.10
4	99.50	0.50
5	99.00	1.00
6	98.50	1.50
7	98.00	2.00

were the calibration set. Multivariate distance measuring computations could then be used to assess whether the central pixel was an outlier in the current local window. Here, the modified Mahalanobis distance known as Global H (*GH*) (which was preferable to the Euclidean distance (*ED*) for measuring the suitability of samples for inclusion in the calibration population) was calculated for each sliding local window [29,30]. In order to resolve overlapped spectra, increase spectral resolution and reduce scattering effects, all the spectra used in this study were pre-treated with 1st derivative Savitzky-Golay and standard normal variate (SNV). *GH* was calculated according to Eqs. (1) and (2) as follows [19,31]:

$$GH = \frac{H^2}{f} \quad (1)$$

$$H^2 = (M - \bar{S}) * V^{-1} * (M - \bar{S})' \quad (2)$$

Here *S* is the  $n \times f$  score matrix of the current local window calculated by principal component algorithm (PCA), *n* is the number of spectra in the current window and *f* is the number of principal components selected. *V* is the covariance matrix of matrix *S*. *M* is the PCA score ( $1 \times f$ ) of the pixel under test.

According to Whitfield et al. [32], the *GH* criterion for deciding if the pixel under test belongs to the current window depends on the number of samples and the dimensions used. From a theoretical Mahalanobis distance criterion, it has been established that  $P * H^2$  can have an *F*-distribution with *f* and ( $n - f + 1$ ) degrees of freedom, where  $P = (n - f + 1) / (n * f)$ . It is therefore possible to obtain a reasonable *GH* threshold with a fixed level of confidence according to Eqs. (3) and (4):

$$P * H^2 = F_{\alpha(f, n-f+1)} \quad (3)$$

$$GH_{threshold} = \frac{F_{\alpha(f, n-f+1)}}{P * f} \quad (4)$$

where  $F_{\alpha(f, n-f+1)}$  is the upper  $100 * \alpha\%$  critical point of the *F*-distribution with (*f*,  $n - f + 1$ ) degrees of freedom.

Then, all the spectra with a *GH* value bigger than threshold will be screened and classified by *k*-means clustering, and the similarity between screened spectra and pure samples will be calculated for further analysis to avoid false judgment.

## 2.5. Problems in LAD method and solutions

In Du et al. [33], the LAD method was applied in NIR hyperspectral remote sensing imaging and some problems, which are common to the strategy proposed in this paper, were found, namely, the multi-pixel targets in the local window and the absence of edge detection.

### 2.5.1. Deal with multi-pixel targets in the local window

The adulterated components with various particle sizes and shapes usually occupied a small fraction of pixels in the NIR microscopic image. Some small particles might occupy only one or less than one

pixel, called a single-pixel target, but others occupied not just one pixel in the NIR microscopic image, but multiple particles gathered together; known as multi-pixel targets. The distribution of target pixels has a great influence on the LAD method, and in particular these multi-pixel targets. An example is provided to illuminate the two cases in Fig. 2A, in which a  $5 * 5$  sliding window filter (the red box) has been used to detect contaminants (white areas) in a simulated NIR microscopic image. With the local window filter moving in the image, the single-pixel/multi-pixel target will be the central pixel under test or the neighboring pixels as a calibration set in one window. The single-pixel target is easy to detect with sufficient 'clean' neighboring pixels in this window, and its existence has practically no impact on the detection of other 'clean' pixels in related windows. However, the multi-pixel target will lead to unsatisfactory local distribution which does not meet the basic requirements for outlier detection using the *GH* criterion. The 'clean' pixel will still be an outlier when the number of anomalous pixels are greater than the 'clean' pixels in one local window. To deal with this situation, all the pixels in one NIR microscopic image were arranged completely randomly to generate a randomized image, in which almost all the multi-pixels would be divided into single pixels and evenly distributed in the image [34], as shown in Fig. 2B. In this way, as the random image exhibits a totally random distribution, it can be regarded as a homogeneity reference [34–36]. After every pixel had been analyzed by the LAD method, the randomized image will be recovered and suspicious sites will be visualized on the original image according to the *GH* criterion. All anomalous spectra will also be screened out and displayed in several groups.

### 2.5.2. Absence of edge detection

The LAD method used a continuous-level sliding local window filter for the detection of the central pixel in the current window. This potentially means that an indefinite quantity of pixels on the four edges of the image will be undetectable due to the use of a different size of window filter, resulting in loss of information. It could lead to false judgments if the target pixels only exist on the edge of the image. In order to avoid this situation, edge expansion was a critical step in the LAD method. If the window filter size was  $w * w$ ,  $(w - 1) / 2$  layers of pixels were expanded on each edge of the image, corresponding to the blue area shown in Fig. 2C. There would be a small percentage of target pixels in an NIR microscopic image of a low-level contamination sample, so the pixels used for edge expansion were randomly selected from the original NIR microscopic image.

In summary, the procedure of the proposed LAD method is as shown in Fig. 3, and about 7 min would be taken for the analysis of each image.

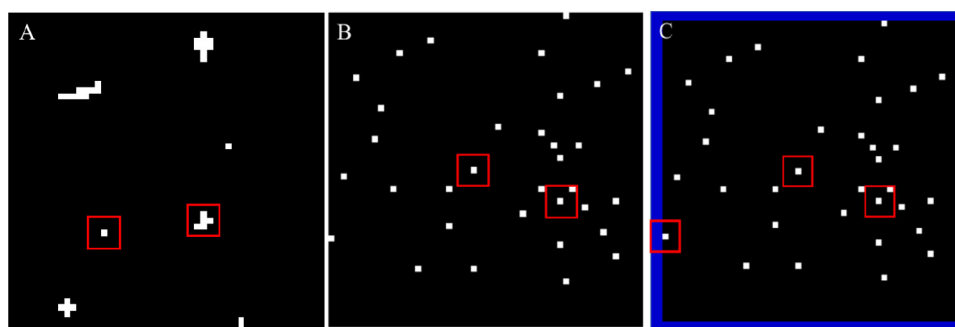
## 2.6. Partial least squares discriminant analysis (PLS-DA)

As a comparison of the detection results of the LAD method, PLS-DA, which is a most commonly used supervised classification method, is used for the detection of melamine in soybean meal samples. PLS-DA is a variation of the partial least squares (PLS) regression algorithm for discriminant analysis, which can seek a direct relationship between the spectral data and the reference values or categorical variables to classify samples into predefined groups. In this work, the representative pure soybean meal spectra and pure melamine spectra were extracted from the NIR hyperspectral images of soybean meal and melamine samples (set B), respectively, and used as the training set of PLS-DA model. All the samples are assigned a dummy variable (0 and 1) as the categorical variable. The PLS-DA model is built with 10-fold cross validation and the number of latent variables is chosen corresponding to the minimum classification error.

### 2.7. Quantitative analysis

In order to evaluate the quantitative analysis performance obtained with the LAD method, the relationships between the numbers of pixels





**Fig. 2.** The  $5 \times 5$  sliding window filter (the red box) used by the LAD method in a simulated image. (A) Simulated original image with single-pixel and multi-pixel targets (white areas); (B) the randomized image generated by the original image with almost no multi-pixel targets; (C) edge expansion of image (blue areas).

detected as ‘contaminant’ by both LAD and PLS-DA methods and the reference values (known added concentrations) were studied, respectively.

All the data are processed using Matlab version 7.14 (The Mathworks, Inc., Natick, MA, USA) with the PLS\_Toolbox version 8.0 (Eigenvector Research, Wenatchee, WA, USA) in this study.

### 3. Results and discussion

#### 3.1. Analysis of pure samples

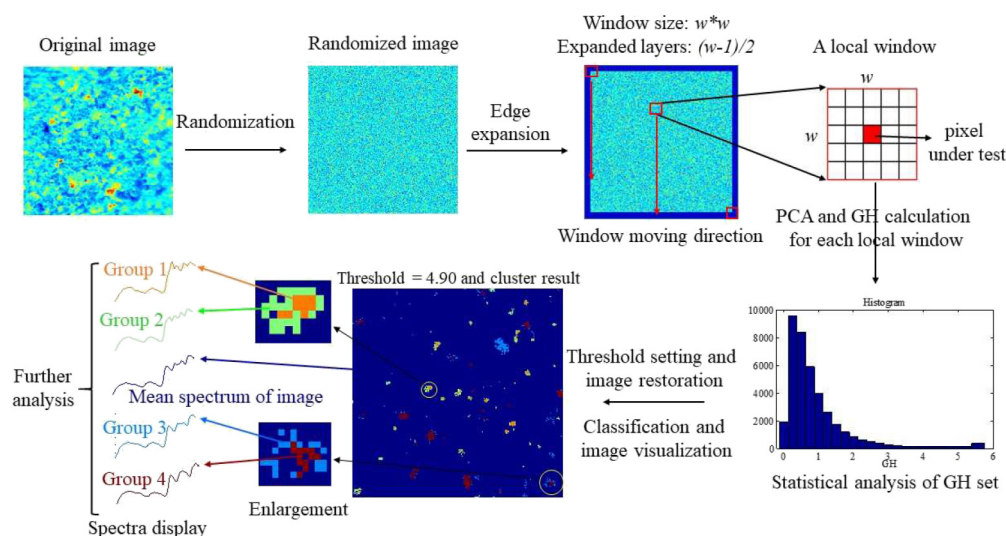
##### 3.1.1. NIR spectra of pure soybean meal and non-protein nitrogen samples

The mean spectra of the NIR images of pure soybean meal samples (de-hulled, hulls and full-fat soybean meal) and non-protein nitrogen (melamine, cyanuric acid, urea, biuret, MAP and DAP) were shown in Fig. 4. Comparison and analysis identified obvious differences in the spectrum of hulls from those of de-hulled and full-fat soybean meal samples around absorbance bands at  $4762\text{ cm}^{-1}$  (O—H bend and C—O stretch combination) and at  $4252\text{ cm}^{-1}$  ( $\text{CH}_2$  bend second overtone) associated with cellulose; this was consistent with the hulls high cellulose content. The spectra of de-hulled and full-fat soybean meal had common protein absorbance bands around  $6667\text{ cm}^{-1}$  (N—H stretch first overtone),  $4850\text{ cm}^{-1}$  (N—H bend second overtone or N—H bend and stretch combination) and  $4587\text{ cm}^{-1}$  (C—H stretch and C=O stretch combination); the difference between them was that full-fat soybean meal showed obvious fat absorption around  $5787\text{ cm}^{-1}$  and  $4261\text{ cm}^{-1}$  [37,38]. In this study, six kinds of non-

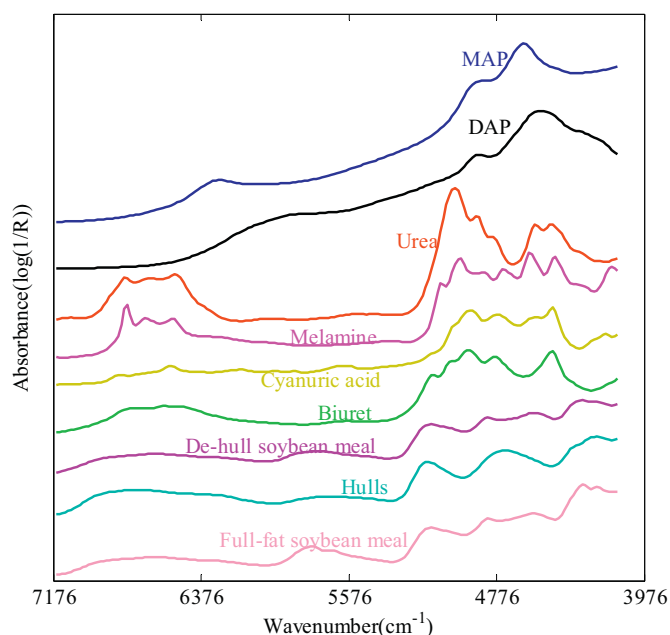
protein nitrogen (melamine, cyanuric acid, urea, biuret, MAP and DAP) were selected as examples of contamination in adulterated soybean meal samples. The spectra of the six kinds of non-protein nitrogen could be used as a reference to verify the anomalous spectra screened out by the LAD method in adulterated samples. As shown in Fig. 4, each non-protein nitrogen sample had its own near infrared characteristic absorption peaks and could be easily distinguished from soybean meal spectra.

##### 3.1.2. LAD analysis of pure soybean meal sample

Firstly, we should know that the composition of each kind of pure soybean meal sample is not single. There is a small amount of soybean hulls exist in de-hulled and full-fat samples, and soy embryo can also be found in soybean hulls sample. Then, the LAD method using a  $5 \times 5$  pixels window filter and four PCA component scores was applied to the NIR microscopic images of pure soybean meal samples. Different numbers of anomalous spectra (the spectra and mixed spectra of the ingredient with low content in pure samples, not contaminant) were detected and classified into four groups by the  $k$ -means method in each kind of pure soybean meal sample. For further analysis and judgment, the mean spectrum of each group and the mean spectrum of pure samples were compared; the results are shown in Table S-1 and Fig. 5. As can be seen in Fig. 5A, the mean spectra of groups 1, 2 and 4 were clearly different from the de-hulled sample and the mean spectrum of group 3 around absorbance bands at  $4762\text{ cm}^{-1}$ . Groups 2 and 4 showed high correlation coefficient with hulls, indicating that hulls remained in the de-hulled sample in the production process. However, the spectra of



**Fig. 3.** Flowchart of the LAD method.



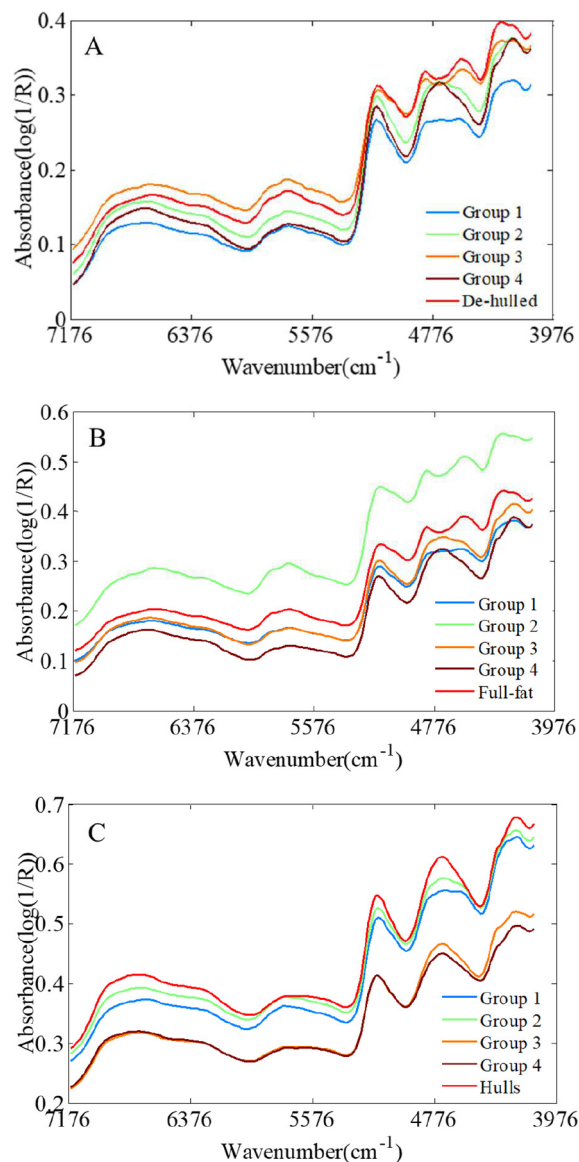
**Fig. 4.** NIR spectra of soybean meal samples (de-hulled, hulls and full-fat soybean meal) and non-protein nitrogen samples (melamine, cyanuric acid, urea, biuret, MAP and DAP).

group 1 without any obvious absorption peak around  $4726\text{ cm}^{-1}$ , was similar neither to the de-hulled sample nor to the hulls, and should be the resultant mixed spectra of de-hulled soybean meal and hulls. This was because each pixel collected by the FT-NIR imaging system was the mixed spectrum resulting from the constituent materials within the pixel. Analysis of Fig. 5B and Table S-1 showed that hulls were also present in the full-fat sample. Similarly, soy embryos inevitably appeared in the hulls sample because of the processing technology. As shown in Fig. 5C and Table S-1, the anomalous spectra of groups 1 and 2 detected in the hulls sample showed broad flat absorption peaks around  $4726\text{ cm}^{-1}$ , which were similar to the anomalous spectra of groups 1 and 2 in the de-hulled sample and of groups 1 and 3 in the full-fat sample, indicating that the presence of soy embryo residue in hulls. All the results indicated that, LAD methods can screen out a small amount of spectra which are different with the spectra of the main ingredient in the sample. Through comparison and analysis, we can know that the anomalous spectra detected in pure samples are the spectra of the low content ingredient which could not be regarded as the spectra of contaminants in future analysis.

### 3.2. LAD analysis of adulterated soybean meal samples

#### 3.2.1. Optimization of moving window filter size

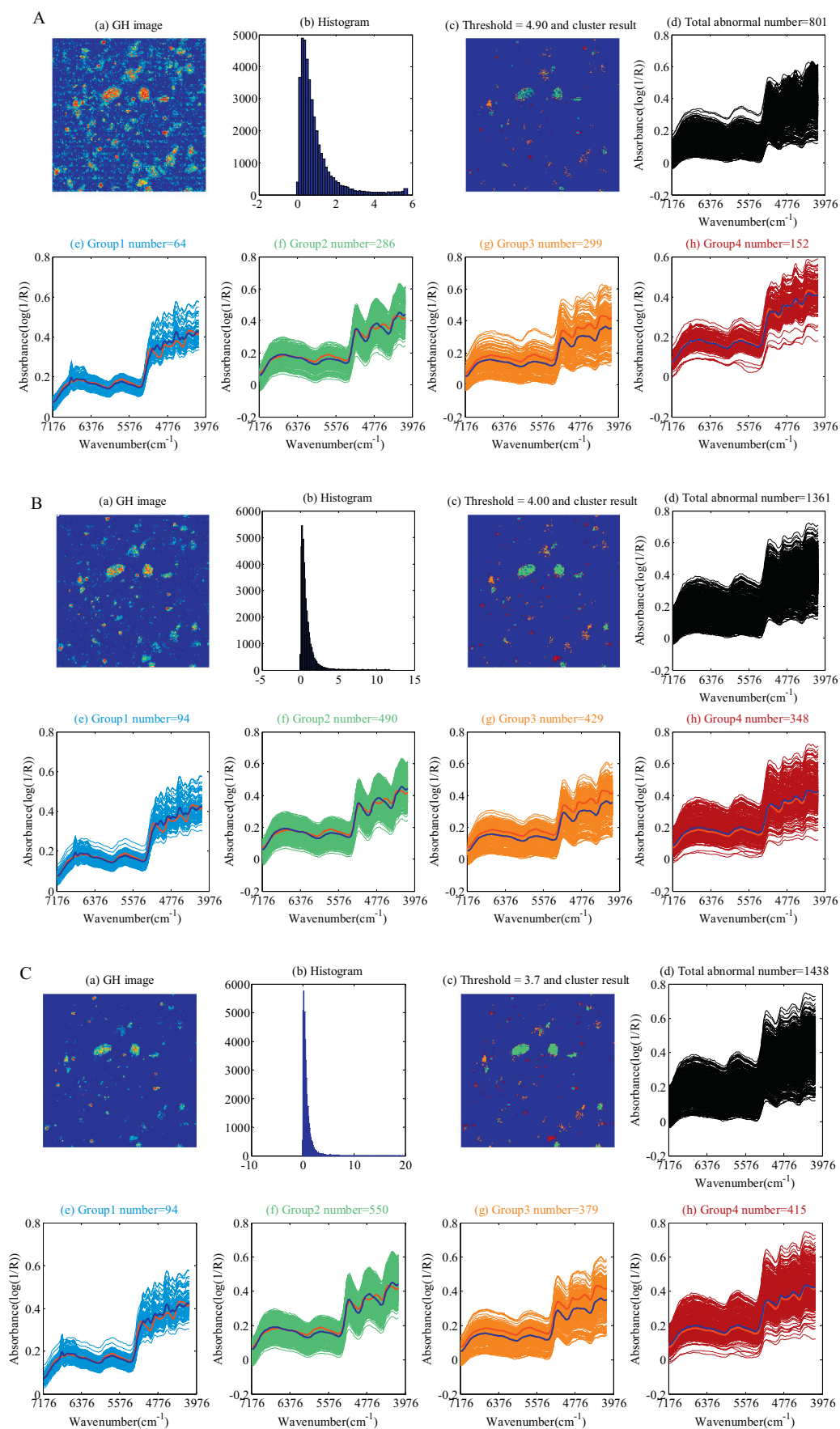
In the LAD method, every pixel in the images was analyzed with a window filter moving from the top left corner to the bottom right corner. For the calculation of  $GH$  between the central pixel and the mean spectrum of the current window, each  $P \times H^2$  had an  $F$ -distribution with  $f$  and  $(n - f + 1)$  degrees of freedom, so the  $GH$  criterion changed as the size of the moving window differed. Here, three different sizes ( $5 \times 5$  pixels,  $7 \times 7$  pixels and  $9 \times 9$  pixels) of window were used for untargeted detection of the same image using the  $GH$  criterion based on four PC scores with a 99% level of confidence; the threshold of  $GH$  was 4.90, 4.00 and 3.70 respectively. As can be seen in Fig. 6A–C, 801, 1361 and 1438 anomalous spectra were screened out using window



**Fig. 5.** Mean spectrum of each anomalous group detected in pure samples and the mean spectrum of pure samples: (A) de-hulled, (B) full-fat, (C) hulls.

filter sizes of  $5 \times 5$  pixels,  $7 \times 7$  pixels and  $9 \times 9$  pixels respectively. The anomalous spectra detected in each image were then classified into four groups. These are displayed in subfigures (e), (f), (g) and (h) and the location of each abnormal spectrum in the original image is shown in subfigure (c) using the same color as its spectra group. For further analysis and judgment of each group of anomalous spectra, the correlation coefficient between the mean spectra of each group obtained by different window sizes and the mean spectra of each pure sample (hulls, full-fat and de-hulled) were calculated. These are recorded in Table 3. In each subfigure (e) from Fig. 6, the difference between spectra of group 1 and the mean spectra of three kinds of pure samples in Fig. 4 were big enough to be determined as the contamination's spectrum by the naked eye, while all correlation coefficient values were less than 0.5. The distribution of contaminations (group 1) is shown in subfigure (c) with light blue pixels; further analysis found that brown pixels

**Fig. 6.** The results of LAD methods with different size of moving window filter for the same image: (A)  $5 \times 5$  pixels, (B)  $7 \times 7$  pixels and (C)  $9 \times 9$  pixels. Subfigure (a) image showed by  $GH$  values, (b) the frequency distribution of  $GH$  values, (c) visualization of different types of spectra screened out by specific  $GH$  threshold and k-means cluster, (d) all spectra screened out by specific  $GH$  threshold, (e), (f), (g) and (h) are four groups of spectra from (d) separated by k-means method. In each subfigure (e), (f), (g) and (h), the red spectrum is the mean of the whole image and the blue spectrum is the mean of current group.



**Table 3**

The correlation coefficient between the mean spectra of each group obtained by different window sizes and the mean spectra of each pure sample.

Pure sample	Groups from 5 * 5 pixels				Groups from 7 * 7 pixels				Groups from 9 * 9 pixels			
	1	2	3	4	1	2	3	4	1	2	3	4
De-hulled	0.39	0.86	0.94	0.81	0.46	0.86	0.94	0.93	0.46	0.86	0.94	0.94
Full-fat	0.32	0.60	0.71	0.66	0.38	0.61	0.70	0.76	0.38	0.61	0.71	0.76
Hulls	0.22	0.99	0.95	0.62	0.29	0.99	0.95	0.74	0.29	0.99	0.96	0.75

from group 4 usually appeared around light blue pixels (group 1) which indicated that the spectra of group 4 were mixed spectra of contaminations and soybean meal, as could also be verified through the correlation coefficient. As the size of window filter increased, more and more mixed spectra contained a small amount of exception information that was be screened out, and the correlation coefficient values between group 4 and the de-hulled sample increased from 0.81 to 0.94, making group 4 more similar to the de-hulled sample. However, a difference could still be found around  $5000\text{ cm}^{-1}$ , where the anomalous spectra had a small absorption peak, whereas the spectra of soybean meal did not exist. Through the correlation coefficient of group 2 and group 3 in Table 3 and the comparison of anomalous spectra screened out from pure samples in Fig. 4, these two groups appear to be the spectra of hulls and the mixed spectra as detected in pure samples, i.e. anomalous spectra of the sample itself rather than contaminants. In view of the calculation speed and the difficulty of identifying the spectra of contaminants, a 5 \* 5 pixels window filter was selected and used in this study.

### 3.2.2. Verification of the LAD method

In order to verify the proposed LAD method, sample set A consisted of pure soybean meal and melamine arranged in a special shape was prepared and analyzed by LAD in this study. As shown in Fig. 1B, all the anomalous spectra screened out by LAD were classified into 4 groups and shown in subfigure (e)–(f). It was obvious that the spectra of group 1 (light blue points in subfigure (c)) and group 3 (orange points in subfigure (c)) appeared in the center of the detected area by LAD were from melamine, and the anomalous spectra of group 2 (green points in subfigure (c)) could be determined as mixed spectra of soybean meal and melamine according to their locations. According to the analysis results of pure soybean meal sample by LAD method, the spectra of group 4 should be the spectra of hulls existed in soybean meal sample. Through comparison, we could find the location of detected melamine by LAD (light blue, orange and green points in subfigure (c)) was consistent with the arranged place of melamine (white area in Fig. 1A), indicating that LAD was a feasible and effective untargeted method.

### 3.2.3. LAD analysis of adulterated samples with single non-protein nitrogen

In order to test the identification feasibility of contaminations in soybean meal of the untargeted LAD method, 18 samples adulterated with single non-protein nitrogen were prepared using three kinds of soybean meal and six kinds of non-protein nitrogen according to the mixing ratio in Table 1. All the adulterated samples were analyzed using a 5 \* 5 pixels window filter, and all the results are shown in Fig. S-1, Tables 4 and S-2. As can be seen in Fig. S-1, hundreds of anomalous spectra were screened out in each adulterated sample (sample 1 to sample 18) using the untargeted LAD method proposed in this study. By analyzing pure samples, we knew that hulls would inevitably appear in the de-hulled and full-fat samples, and soy embryo are always found in hulls samples, the spectra of which were anomalous but unrelated to the contaminations to be detected. To distinguish contaminations from anomalous spectra, all the anomalous spectra screened out from each adulterated sample were classified into four groups by the *k*-means method and each anomalous spectrum was visualized in subfigure (c) using the color of its group in subfigure (e), (f), (g) and (h) respectively. The correlation coefficient between the mean spectra of each group and the mean spectra of each pure sample were then

calculated and recorded in Table S-2. Generally speaking, a small number of pure spectra of contaminations could be collected by the FT-NIR imaging system using a  $25\text{ }\mu\text{m} \times 25\text{ }\mu\text{m}$  spatial resolution in the center of the contamination's particle, and the difference between these and the spectra of the pure sample was usually big enough to be recognized by the naked eye. For example, the spectra of group 1 (light blue) in subfigure (e) from Fig. S-1 sample 1, showed obviously different absorption peaks from the mean spectrum of the whole image (red), and could be easily determined as contaminations, and the correlation coefficient between the mean spectrum of group 1 and the spectrum of hulls, full-fat and de-hulled was 0.34, 0.47 and 0.56 respectively. Further analysis found that most of the brown spectra from group 4 appeared around the light blue (group 1) pixels in subfigure (c), which indicated that the spectra of group 4 were mixed spectra of contaminations and soybean meal, while the correlation coefficient between the mean spectrum of group 4 and hulls, full-fat and de-hulled was 0.67, 0.75 and 0.89 respectively. The spectra of group 2 and group 3 were similar with anomalous spectra detected in the pure de-hulled sample as shown in Fig. 5, which could be classified as the spectra and mixed spectra of hulls by naked eye and using the correlation coefficients in Table S-2. Ultimately, therefore, 377 spectra of contaminations and 555 anomalous spectra of hulls were screened out in sample 1 by the LAD method. The rest of the adulterated samples were also analyzed in this way and the results of LAD method were compared with the *GH* method based on the spectral library and the PLS-DA method published in 2016 [19]. Through the comparisons in Table 4, we found that the LAD method could screen out more spectra of contaminations than the *GH* and the PLS-DA methods in most of the samples adulterated with 0.5% single non-protein nitrogen, which meant that the LAD method as untargeted detection pattern needs neither a spectral library nor a discriminant analysis model and is more rigorous and effective for contamination detection in soybean meal samples.

As referred to in previous work [19], DAP exposed to the air gradually loses ammonia and is converted into MAP, which makes it impossible for the PLS-DA model to recognize. However, such situation would not have a negative impact on untargeted detection methods, which would be able to screen out all the suspicious spectra without knowing what kind of illegal ingredient had been added to the soybean meal in advance. Sample 16 is a typical case, in which only 5 spectra of DAP were detected by the PLS-DA method, while 227 and 193 spectra of contaminations were screened out by the LAD and *GH* methods respectively. This shows that the untargeted detection method is more advantageous in adulteration detection, without any concern about the effect of the deterioration of adulterants.

### 3.2.4. LAD analysis of adulterated samples with multiple non-protein nitrogen

The LAD method has performed well in the analysis of single non-protein nitrogen adulterated samples, so soybean meal samples adulterated with six kinds of non-protein nitrogen at the same time were also prepared to test the LAD method in this study. As shown in Fig. S-1 (sample 19–21), all the anomalous spectra were classified into eight groups. Combining the spectra in subfigure (d)–(l) and the correlation coefficient in Table S-4, the spectra of contaminations could be easily identified and visualized on the image (subfigure (c)), which suggested that the LAD approach was also effective in the face of complex adulteration.



**Table 4**  
The LAD analysis results of adulterated samples set A.

Sample ID	Total number <sup>a</sup>	Number of contaminant spectra		
		LAD method	GH [19]	PLS-DA [19]
Set 1	1	40,000	377	165
	2	40,000	216	98
	3	40,000	211	87
	4	40,000	276	82
	5	40,000	302	117
	6	40,000	309	181
	7	40,000	330	126
	8	40,000	557	341
	9	40,000	112	171
	10	40,000	176	132
	11	40,000	150	123
	12	40,000	91	105
	13	40,000	173	110
	14	40,000	346	190
	15	40,000	111	119
	16	40,000	227	193
	17	40,000	163	20
	18	40,000	223	216
Set 2	19	122,500	1214	1549
	20	122,500	678	1781
	21	122,500	772	1181

<sup>a</sup> Total number of spectra from one image.

In summary, the LAD method proposed in this study showed acceptable good performance for the detection of 0.5% (w/w) single and 3.0% (w/w) multiple non-protein nitrogen adulteration in soybean meal samples. The untargeted detection pattern without the need for either

a spectral library or a discriminant analysis model showed great application potential for preventing the emergence of new adulterations to guarantee food and feed safety.

### 3.2.5. Repeatability of the LAD method

Because of the randomization step during data processing, the reproducibility of the LAD method was tested by analyzing one image by ten repetitions. As shown in Table S-3, the mean number of contaminant spectra screened out by the LAD method was 205.1 with a standard deviation (SD) of 7.46 and a relative standard deviation (RSD) of 3.64%, suggesting that the LAD method had good repeatability and could be used as an untargeted detection method for the detection of contaminants in soybean meal.

### 3.2.6. Quantitative performance of the LAD and PLS-DA method

The adulterated samples of set C were prepared to evaluate the quantitative performance of the LAD method. For this, all the samples of this set were analyzed by both LAD and PLS-DA methods, the number of detected contaminant pixels of each part is shown in Table S-4, which corresponds to the mean value of the seven images, and the mean values and standard deviation of the three parts of each sample were calculated. Both liner and quadratic regression equations between the real percentages of added contaminant and the pixels detected as melamine by LAD and PLS-DA methods were studied, respectively. As could be seen in Fig. 7(A) and (B), the quadratic regression equations (red) showed better performance than liner regression equations (black), and both quadratic regression lines showed a high  $R^2$  ( $R^2_{LAD} = 0.9984$  and  $R^2_{PLS-DA} = 0.9978$ ), which indicates that the LAD method has a similar performance as PLS-DA for quantitative assessment of melamine as adulterant concentrations in soybean meal. It could be seen from Table S-4 that samples contaminated at 0.01% level could still be detected indicating that the LOD (limit of detection) of the LAD method can be lower than 0.01%.

## 4. Conclusion

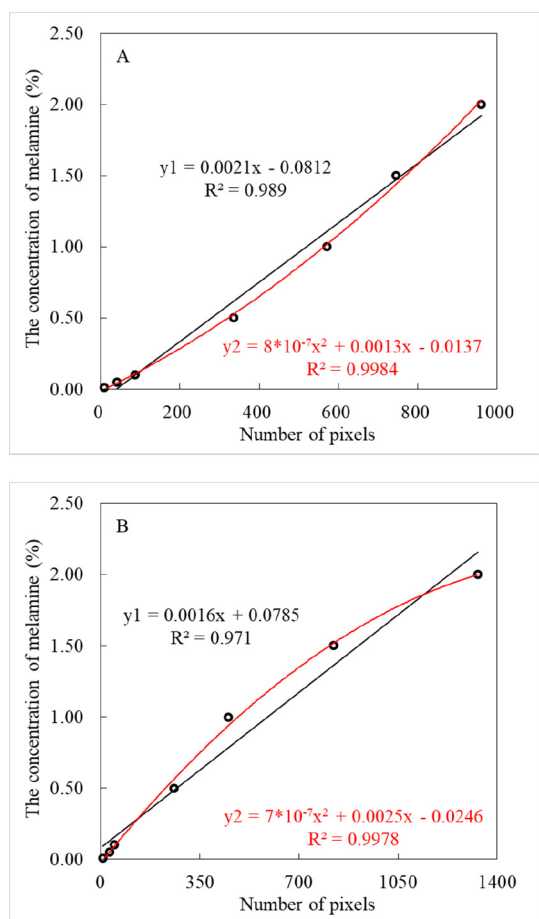
In this study, six kinds of contaminants and two different NIR imaging instruments were used to evaluate the performance of the new untargeted LAD method to detect and semi-quantify adulterants. The satisfactory results indicated that the untargeted LAD method could be used for the detection of many kinds of contaminants, and could also be applied to process different NIR images obtained by different types of NIR imaging instruments. The detection pattern of the untargeted LAD method, whose the LOD can be lower than 0.01%, should be extended to detect contaminants in other sample matrices, and there is reason to hope that this will make it possible to identify suspicious contaminants as early as possible to keep human being from injury and move food and feed safety control from passive to active.

## Acknowledgments

The authors gratefully acknowledge financial support from National Key R&D Program of China (2017YFE0115400) and Program for Changjiang Scholars and Innovative Research Team in University of Ministry of Education of China (IRT-17R105). The authors gratefully acknowledge also financial support from the China Scholarship Council that allowed the author to work at the CRA-W. The authors want to thank also the technical staff of the CRA-W, especially to Nicaise Kayoka Mukendi from the Valorisation of Agricultural Products Department (CRA-W) who helped with hyperspectral image acquisitions.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.saa.2019.117494>.



**Fig. 7.** Regression lines between the concentrations of added melamine and the number of pixels detected as melamine by: (A) LAD and (B) PLS-DA.

## References

- [1] W.C. Andersen, S.B. Turnipseed, C.M. Karbiwnyk, S.B. Clark, M.R. Madson, C.M. Gieseker, R.A. Miller, N.G. Rummel, R. Reimschuessel, Determination and confirmation of melamine residues in catfish, trout, tilapia, salmon, and shrimp by liquid chromatography with tandem mass spectrometry, *J. Agr. Food Chem.* 56 (12) (2008) 4340–4347.
- [2] X. Xu, Y. Ren, Y. Zhu, Z. Cai, J. Han, B. Huang, Y. Zhu, Direct determination of melamine in dairy products by gas chromatography/mass spectrometry with coupled column separation, *Anal. Chim. Acta* 650 (1) (2009) 39–43.
- [3] C.M. Gossner, J. Schlundt, E.P. Ben, S. Hird, D. Lo-Fo-Wong, J.J. Beltran, K.N. Teoh, A. Tritscher, The melamine incident: implications for international food and feed safety, *Environ. Health Perspect.* 117 (12) (2009) 1803–1808.
- [4] J.S. Shen, J.Q. Wang, H.Y. Wei, D.P. Bu, P. Sun, L.Y. Zhou, Transfer efficiency of melamine from feed to milk in lactating dairy cows fed with different doses of melamine, *J. Dairy Sci.* 93 (5) (2010) 2060–2066.
- [5] C.W. Cruywagen, M.A. Stander, M. Adonis, T. Calitz, Hot topic: pathway confirmed for the transmission of melamine from feed to cow's milk, *J. Dairy Sci.* 92 (5) (2009) 2046–2050.
- [6] P. Sun, J.Q. Wang, J.S. Shen, H.Y. Wei, Residues of melamine and cyanuric acid in milk and tissues of dairy cows fed different doses of melamine, *J. Dairy Sci.* 94 (7) (2011) 3575–3582.
- [7] M. Battaglia, C.W. Cruywagen, T. Bertuzzi, A. Gallo, M. Moschini, G. Piva, F. Masoero, Transfer of melamine from feed to milk and from milk to cheese and whey in lactating dairy cows fed single oral doses, *J. Dairy Sci.* 93 (11) (2010) 5338–5347.
- [8] C. Decision, Commission Decision 2008/757/EC imposing special conditions governing the import of products containing milk or milk products originating in or consigned from China (2008a), *Off. J. Eur. Union* 51 (2008) 10–11 (Commission Decision, No. L 259/10–11).
- [9] G. Venkatasami, J.R. Sowa, A rapid, acetonitrile-free, HPLC method for determination of melamine in infant formula, *Anal. Chim. Acta* 665 (2) (2010) 227–230.
- [10] H. Miao, S. Fan, Y.N. Wu, L. Zhang, P.P. Zhou, J.G. Li, H.J. Chen, Y.F. Zhao, Simultaneous determination of melamine, ammeline, and cyanuric acid in milk and milk products by gas chromatography-tandem mass spectrometry, *Biomed. Environ. Sci.* 22 (2) (2009) 87–94.
- [11] M.S. Filigenzi, B. Puschner, L.S. Aston, R.H. Poppenga, Diagnostic determination of melamine and related compounds in kidney tissue by liquid chromatography/tandem mass spectrometry, *J. Agr. Food Chem.* 56 (17) (2008) 7593–7599.
- [12] H. Lei, Y. Shen, L. Song, J. Yang, O.P. Chevallier, S.A. Haughey, H. Wang, Y. Sun, C.T. Elliott, Hapten synthesis and antibody production for the development of a melamine immunoassay, *Anal. Chim. Acta* 665 (1) (2010) 84–90.
- [13] R.M. Balabin, S.V. Smirnov, Melamine detection by mid- and near-infrared (MIR/NIR) spectroscopy: a quick and sensitive method for dairy products analysis including liquid milk, infant formula, and milk powder, *Talanta* 85 (1) (2011) 562–568.
- [14] O. Abbas, B. Lecler, P. Dardenne, V. Baeten, Detection of melamine and cyanuric acid in feed ingredients by near infrared spectroscopy and chemometrics, *J. Near Infrared Spec.* 21 (3) (2013) 183–194.
- [15] M. Lin, L. He, J. Awika, L. Yang, D.R. Ledoux, H. Li, A. Mustapha, Detection of melamine in gluten, chicken feed, and processed foods using surface enhanced Raman spectroscopy and HPLC, *J. Food Sci.* 73 (8) (2008) T129–T134.
- [16] S. Okazaki, M. Hiramatsu, K. Gonmori, O. Suzuki, A.T. Tu, Rapid nondestructive screening for melamine in dried milk by Raman spectroscopy, *Forensic Toxicol* 27 (2) (2009) 94–97.
- [17] J. Qin, K. Chao, M.S. Kim, Simultaneous detection of multiple adulterants in dry milk using macro-scale Raman chemical imaging, *Food Chem.* 138 (2–3) (2013) 998–1007.
- [18] X. Fu, M.S. Kim, K. Chao, J. Qin, J. Lim, H. Lee, A. Garrido-Varo, D. Pérez-Marín, Y. Ying, Detection of melamine in milk powders based on NIR hyperspectral imaging and spectral similarity analyses, *J. Food Eng.* 124 (2014) 97–104.
- [19] G. Shen, X. Fan, Z. Yang, L. Han, A feasibility study of non-targeted adulterant screening based on NIRM spectral library of soybean meal to guarantee quality: the example of non-protein nitrogen, *Food Chem.* 210 (2016) 35–42.
- [20] Q. Li, X. Kang, D. Shi, Q. Liu, Determination of melamine in soybean meal by near-infrared imaging and chemometrics, *Anal. Lett.* 49 (10) (2016) 1564–1577.
- [21] L. Xu, S.M. Yan, C.B. Cai, Z.J. Wang, X.P. Yu, The feasibility of using near-infrared spectroscopy and chemometrics for untargeted detection of protein adulteration in yogurt: removing unwanted variations in pure yogurt, *J. Anal. Methods Chem.* 2013 (2013) 1–9.
- [22] L. Xu, S.M. Yan, C.B. Cai, X.P. Yu, Untargeted detection of illegal adulterations in Chinese glutinous rice flour (GRF) by NIR spectroscopy and chemometrics: specificity of detection improved by reducing unnecessary variations, *Food Anal. Method.* 6 (6) (2013) 1568–1575.
- [23] J.C. Moore, A. Ganguly, J. Smeller, L. Botros, M. Mossoba, M.M. Bergana, Standardisation of non-targeted screening tools to detect adulterations in skim milk powder using NIR spectroscopy and chemometrics, *NIR News* 23 (5) (2012) 9–11.
- [24] L. Xu, S. Yan, C. Cai, X. Yu, One-class partial least squares (OCPLS) classifier, *Chemometr. Intell. Lab.* 126 (2013) 1–5.
- [25] L. Xu, S.M. Yan, C.B. Cai, X.P. Yu, Untargeted detection and quantitative analysis of poplar balata (PB) in Chinese propolis by FT-NIR spectroscopy and chemometrics, *Food Chem.* 141 (4) (2013) 4132–4137.
- [26] J.A. Fernández Pierna, D. Vincke, V. Baeten, C. Grellet, F. Dehareng, P. Dardenne, Use of a multivariate moving window PCA for the untargeted detection of contaminants in agro-food products, as exemplified by the detection of melamine levels in milk using vibrational spectroscopy, *Chemometr. Intell. Lab.* 152 (2016) 157–162.
- [27] H.Y. Fu, H.D. Li, L. Xu, Q.B. Yin, T.M. Yang, C. Ni, C.B. Cai, J. Yang, Y.B. She, Detection of unexpected frauds: screening and quantification of maleic acid in cassava starch by Fourier transform near-infrared spectroscopy, *Food Chem.* 227 (2017) 322–328.
- [28] G. Gizzi, C. von Holst, V. Baeten, G. Berben, L. van Raamsdonk, Determination of processed animal proteins, including meat and bone meal, in animal feed, *J. AOAC Int.* 87 (6) (2004) 1334–1341.
- [29] O.E. De Noord, Multivariate calibration standardization, *Chemometr. Intell. Lab.* 25 (2) (1994) 85–97.
- [30] H. Mark, Normalized distances for qualitative near-infrared reflectance analysis, *Anal. Chem.* 58 (2) (1986) 379–384.
- [31] J.A. Guthrie, Robustness of NIR Calibrations for Assessing Fruit Quality, Central Queensland University, Rockhampton, Queensland, Australia, 2005.
- [32] R.G. Whitfield, M.E. Gerger, R.L. Sharp, Near-infrared spectrum qualification via Mahalanobis distance determination, *Appl. Spectrosc.* 41 (7) (1987) 1204–1213.
- [33] B. Du, R. Zhao, L. Zhang, L. Zhang, A spectral-spatial based local summation anomaly detection method for hyperspectral images, *Signal Process.* 124 (2016) 115–131.
- [34] J.G. Rosas, M. Blanco, A criterion for assessing homogeneity distribution in hyperspectral images. Part 1: homogeneity index bases and blending processes, *J. Pharmaceut. Biomed.* 70 (2012) 680–690.
- [35] P. Sacré, P. Lebrun, P. Chavez, C.D. Bleye, L. Netchacovitch, E. Rozet, R. Klinkenberg, B. Streel, P. Hubert, E. Ziemons, A new criterion to assess distributional homogeneity in hyperspectral images of solid pharmaceutical dosage forms, *Anal. Chim. Acta* 818 (2014) 7–14.
- [36] J.G. Rosas, M. Blanco, A criterion for assessing homogeneity distribution in hyperspectral images. Part 2: application of homogeneity indices to solid pharmaceutical dosage forms, *J. Pharmaceut. Biomed.* 70 (2012) 691–699.
- [37] J.A.S. Del Rivero, E. Montañés-Roces, B. De La Roza-Delgado, A. Soldado, O. Luaces, J.R. Quevedo, A. Bahamonde, Feature selection for classification of animal feed ingredients from near infrared microscopy spectra, *Inform. Sciences* 241 (2013) 58–69.
- [38] D.A. Burns, E.W. Ciurczak, *Handbook of Near-infrared Analysis*, CRC press, 2007.