

Application of independent components analysis with the JADE algorithm and NIR hyperspectral imaging for revealing food adulteration

Puneet Mishra, Christophe B.Y. Cordella, Douglas N. Rutledge, Pilar Barreiro, Jean Michel Roger,

Belén Diezma

ABSTRACT

In recent years, Independent Components Analysis (ICA) has proven itself to be a powerful signal-processing technique for solving the Blind-Source Separation (BSS) problems in different scientific domains. In the present work, an application of ICA for processing NIR hyperspectral images to detect traces of peanut in wheat flour is presented. Processing was performed without *a priori* knowledge of the chemical composition of the two food materials. The aim was to extract the source signals of the different chemical components from the initial data set and to use them in order to determine the distribution of peanut traces in the hyperspectral images. To determine the optimal number of independent component to be extracted, the *Random ICA by blocks* method was used. This method is based on the repeated calculation of several models using an increasing number of independent components after randomly segmenting the matrix data into two blocks and then calculating the correlations between the signals extracted from the two blocks. The extracted ICA signals were interpreted and their ability to classify peanut and wheat flour was studied. Finally, all the extracted ICs were used to construct a single synthetic signal that could be used directly with the hyperspectral images to enhance the contrast between the peanut and the wheat flours in a real multi-use industrial environment. Furthermore, feature extraction methods (connected components labelling algorithm followed by flood fill method to extract object contours) were applied in order to target the spatial location of the presence of peanut traces. A good visualization of the distributions of peanut traces was thus obtained.

1. Introduction

In recent years, hyperspectral imaging (HSI) has emerged as a promising tool for monitoring the safety and quality of various food commodities. The ability of HSI to integrate the information of spectroscopy and imaging and to simultaneously acquire spectral and spatial information in a single system has made it a suitable choice for dealing with complex issues associated with evaluating individual food items (Wu and Sun, 2013). HSI often collects images with high spatial and spectral resolutions (Qin et al., 2013) from which the data can be further transformed into radiometric quantities such as reflectance, absorbance, and transmittance in order to relate them to the physical properties or chemical composition of samples (Wu and Sun, 2013). Different

supervised and unsupervised methods followed by feature extraction procedures are generally used to extract important spatial and spectral patterns. The main steps required to analyze hyperspectral images include pre-processing of data, dimensionality reduction, enhancement of spectral responses, and component detection or classification (Mahesh et al., 2015). HSI with advanced chemometrics methods is nowadays gaining in importance for detecting adulteration in various food products. Some recent works include detection of minced lamb meat adulteration (Kamruzzaman et al., 2013), gelatine adulteration in prawns (Wu et al., 2013), detection and quantification of melamine in milk powders (Fu et al., 2014; Santos et al., 2013), detection of microbial contamination in pork (Barbin et al., 2013) and faecal contamination in leafy greens (Kang et al., 2011). However, these works were mainly focussed on the use of statistical methods based on second-order moment such as PCA (Principal Component Analysis), PLS regression (Partial Least Square), Multi-Linear Regression (MLR), or on

spectral similarity measures like Spectral Angle Mapper (SAM), Euclidian Distance Measure (EDM) and Spectral Correlation Measure (SCM). The need to resolve the spectra using Blind Source Separation methods was not initiated for complex data processing.

Independent Components Analysis (ICA), first proposed by [Herauld and Jutten in 1986](#), is known to be a powerful signal-processing technique for solving Blind-Source Separation (BSS) problems in various scientific and engineering domains. In recent years, ICA has received wide attention in analytical chemistry for pre-processing, exploration, classification, regression and resolution of data. One major capability of ICA is to perform unsupervised classification (classify objects with unknown spectral signatures) by finding a linear representation of non-Gaussian data so that the components are statistically independent ([Du, 2004](#)). In contrast to the statistical methods based on second-order moments, ICA uses fourth-order moment of the signals to obtain the latent variables ([Wang et al., 2008](#)). Recent applications of ICA for spectroscopic data interpretation includes processing of infrared spectra of marine organic matter aggregates ([Monakhova et al., 2015](#)), detection of orange juice frauds using front-face fluorescence spectroscopy ([Ammari et al., 2015](#)), spectrophotometric analysis of polysaccharide/milk protein interactions with methylene blue ([Rohart et al., 2015](#)), near-infrared spectroscopy for analysis of bioactive components ([Chuang et al., 2014a](#)), near infrared spectroscopy for evaluation of rice freshness ([Chuang et al., 2014b](#)), analysis of Raman images of pharmaceutical drug product ([Boiret et al., 2014](#)), fluorescence spectroscopy for studying interaction between plastic food packaging and olive oil ([Kassouf et al., 2014](#)) and to characterise organic matter in soils ([Ammari et al., 2014](#)). The outcomes of these studies indicate that ICA simplified the interpretation of the results by decomposing the original spectral data into "source signals". The source signals extracted were further easily related to chemical compounds and for determining their spatial distribution.

The present paper deals with a combined approach of using independent components analysis to process NIR hyperspectral image in order to detect peanut fragments as a food adulterant in wheat flour. Peanut allergy is a potentially life-threatening condition accounting for the majority of frequent allergenic reactions to foods ([Finkelman, 2010](#)). Unlike other allergies such as milk and egg allergies, peanut allergy persists throughout life while others typically resolve themselves during early childhood. Peanut allergy has become a major health concern in developed countries. However, the reasons for its increasing prevalence are still not well understood ([Burks, 2008](#)). The potential for accidental exposure to peanuts is an important concern and the ubiquitous nature of peanut in the food industry makes dietary avoidance difficult. Furthermore, manufacturing and labelling errors and label misinterpretation can contribute to the risk of inadvertent exposure ([Yu et al., 2006](#)). To protect allergic consumers, *European Directive 2003/89/EC* makes the labelling of all ingredients mandatory, especially food allergens used in the recipes of packaged foods. However, as small amounts of adventitious allergens can be found in a large range of industrial food products, a risk still persists ([Crépet et al., 2015](#)).

In this paper, adulterated mixtures of peanut traces in wheat flour were prepared and analysed by Near Infrared (NIR) HSI to test the potential of coupling HSI, ICA and image processing tools for detecting peanut traces in early stages of food manufacturing. As peanut and wheat flour contain a range of chemical constituents such as starch, fatty acid and proteins, the objective was to extract the interpretable source signals by applying ICA to the signal mixtures, so that they could then be used to enhance the contrast between the peanut traces and wheat flour in the hyperspectral images. As ICA is based on the decomposition of the original data

matrix into mutually independent source signals, the number of independent components to be extracted is of critical importance. To decide on the optimum number of ICs, *Random ICByBlocks* method was used. This method is based on the calculation of ICA models using an increasing number of independent components after segmenting the matrix data by randomly distributing the samples into two blocks and then calculating the correlations between the signals extracted from the blocks. After having obtained enhanced contrast images, a feature extraction method was used to detect the position of peanut fragments in the image.

2. Materials and methods

2.1. Samples

To obtain the hyperspectral images, four sets of samples were prepared: 100% ground peanuts, 100% wheat flour and 0.05% and 0.01% (w/w) ground peanuts in wheat flour. Wheat flour (*Triticum aestivum*), "Coeur de Blé" (MasterChef), with particle size 125–100 µm to 212–160 µm, was purchased from a local supermarket, Montpellier France. The list of ingredient on the pack does not indicate the presence of peanuts. Ground peanuts (*Arachis hypogaea*) (IRMM-481kit) were obtained from the European Commission Joint Research Centre, Institute for Reference Materials and Measurements. The kit (IRMM-481) with six different vials contains non-salted peanut powder with a particle size range from 500 µm to 1000 µm. Four of the six vials were used: i.e. IRMM-481a, origin from Runners Argentina, corresponded to blanched peanuts, air-roasted at 140 °C for 20 min; IRMM-481b origin from Common Natal South Africa refers to raw peanuts, air roasted at 160 °C for 13 min; IRMM-481c from Virginia USA, were blanched peanuts, oil roasted at 145 °C for 25 min; and IRMM-481e from Jumbo Runners, USA, were blanched peanuts without roasting. These samples were mixed manually with wheat flour to create the trace samples. These four peanut varieties with the different pre-processing procedures were used in order to take into account the variability of peanut traces to be found in an industrial environment. An analytical weighing balance with precision of 0.0001g (KERN 770, Balingen, Germany) was used to measure the different masses of peanut and wheat flour for subsequent mixing. An aluminium platform with an exposed sample surface area of 95 cm² was designed for performing the sample presentation for image acquisition. The depth of the surface used for sample presentation was 3 mm. Aluminium was chosen as a material because it is optically neutral, i.e., it presents constant absorption throughout the spectral range used. This property allows measuring the spectra of the samples without interferences from the platform. Samples were spread out and smoothed using a disposable wooden spatula for even distribution.

2.2. Hyperspectral imaging measurements

The images were acquired with a line-scan push broom hyperspectral imaging system. The system consists of a HySpex SWIR-320m-e hyperspectral push broom camera (Norsk Elektro Optikk, Lørenskog, Norway), translation stage operated by a stepper motor and a computer supported by Norsk software (Norsk Elektro Optikk, Lørenskog, Norway). An illumination unit including 150 W tungsten halogen source mounted at a 45° angle on side of the camera in respect to the vertical plane was used to illuminate the sample. The hyperspectral camera covers the spectral range from 1000 to 2500 nm in 256 bands (spectral resolution of 6 nm) with 320 pixels over the cross-track field of view (FOV). The camera was equipped with a 30 cm lens that produced a nominal pixel size of approximately 0.287 mm across and 0.427 mm along the

FOV respectively. Settings of acquisition parameters (integration time and dark current subtraction), data acquisition and recording were performed by means of the software provided by NEO. Prior to acquisition of images, the halogen bulb was switched on for a half hour in order to stabilize the light source temperature drift and improve the spatial uniformity of the lighting of the sample placed under the Field of View (FOV) (Piqueras et al., 2012). Samples were placed successively on the translation stage to be scanned line by line by the imaging sensors. Raw images were first corrected in radiance using sensor characteristics (e.g. spectral sensitivity) provided by the manufacturer. To obtain reflectance images (images independent of illumination conditions), a 2-point reflectance calibration was performed using a white diffuse reflectance standard (SPECTRALON[®] Labsphere, Martillac, France). Finally, absorbance images ($-\log_{10}$) were computed for use in further analysis.

2.3. Chemometric analysis

2.3.1. Preprocessing of hyperspectral images

Pre-processing of spectroscopic data sets are necessary to improve chemometric modelling (Rohart et al., 2015), and to remove any non-chemical biases such as scattering effects due to inhomogeneity of the surface, interference from external light sources or random noise (Boiret et al., 2014). SNV (Standard Normal Variate) pre-processing method (Barnes et al., 1989) is very helpful for spectroscopic techniques as it reduces effects of baseline shift and variations in global signal intensity (Boiret et al., 2014). Before performing the SNV pre-processing step, the spectral range of hyperspectral images was reduced to 1000–2200 nm (region of interest). Then the mean and standard deviations of each spectrum were calculated. Each mean was then subtracted from the data points of the corresponding spectrum, which was then divided by the corresponding standard deviation.

2.3.2. ICA by the JADE algorithm

ICA is a blind source separation method that has been used for multivariate resolution purposes in analytical chemistry. In ICA, the data matrix is assumed to be a collection of weighted sums of the source signals, the weights being proportional to the contribution of the corresponding source signal to each mixed signal. The aim of ICA is to obtain linear transformations of the data matrix that maximize the statistical independence (non-Gaussianity) among the extracted source signals. According to the Central Limit Theorem, the measured signals, which are mixtures of several independent sources, should be “more Gaussian” than the source signals. Hence, the objective of ICA is to search for the least Gaussian sources (Rutledge and Jouan-Rimbaud Bouveresse, 2013). Let us suppose a mixed signal gives a data matrix X ($r \times c$) where r is the number of observed mixed spectra in X and c is the number of points in the signal, or variables corresponding to the matrix columns. In ICA it is assumed that the matrix X can be decomposed as:

$$X = A \times S$$

where S is $k \times c$ matrix of k independent source signals (ICs) called the Independent Components, and A is the $r \times k$ mixing matrix (proportions of pure signals). ICA aims to determine both S (ICs) and A (proportions), with knowledge only of X . With higher order statistic, ICA aims to find a demixing matrix W in order to extract the source signals from the mixture, as per the equation:

$$S = W \times X$$

This demixing matrix approximates the pseudo inverse of mixing matrix. Once this demixing matrix is approximated, the source signals can be extracted from the observed signals.

Hence, the mixing matrix (concentration) is calculated as:

$$A = X \times S^T \times (S \times S^T)^{-1}$$

The aim of ICA is to determine the ICs, which are as independent as possible. To perform this, there are different algorithms (FastICA (Hyvärinen and Oja, 1997) ?, InfoMax (Bell and Sejnowski, 1995) and JADE (Cardoso, 1993)). In the present work, the JADE algorithm was used to obtain this demixing matrix. The algorithm used in the experiment analysis has been explained in Rutledge and Jouan-Rimbaud Bouveresse (2013).

2.3.3. Deciding on the optimal number of ICs by random ICA-by-blocks

Before extracting the source signals, it is critical to determine the number of relevant components to be extracted so that it contains the useful information. For that purpose, an extension of the ICA-by-blocks (Jouan-Rimbaud Bouveresse et al., 2012) method was used. In the present work, ICA-by-blocks has been applied in modified form by performing *Random ICA-by-blocks* with repetitions. This method consists of splitting the data matrix into 2 blocks. The distribution of the samples into the 2 blocks was done randomly for each repetition. The 2 blocks contain approximately equal numbers of rows selected from the complete data matrix. Then, for each of these blocks, A_{\max} ICA models are computed. A_{\max} should exceed the expected optimal number of ICs. If the different blocks are representative of the whole data set, the true source signals should be found in both blocks. Calculating the absolute correlation between each pair of source signals can help in comparing the computed ICs for different blocks. The correlation of a source signal with itself, even when extracted from another block, will give a value close to one, but its correlation with a different source signal or a noise signal, will be close to zero. Hence, the decrease in correlations as a function of number of extracted ICs can provide appropriate information about the relevant number of ICs to be extracted.

2.4. Pixel detection using image segmentation

Imaging is an important way to characterise food quality. Image pre-processing (denoising, enhancement) and image analysis are two major steps in image processing (i.e. image segmentation, feature extraction), which have for goal the extraction of the useful information from images (Huang et al., 2001). In this work, image segmentation was used to subdivide images provided by the ICA treatment ('proportion' or 'score' images obtained from ICA) in order to isolate peanut particles. Once the images with enhanced contrast are obtained for different components present in a mixed image, different feature extraction algorithms can be applied to locate the exact area of the interesting pixels within the image. In the present work, the connected components labelling algorithm (Di Stefano and Bulgarelli, 1999) followed by the flood fill method to extract object contours (Lee and Kang, 2010), from the image processing toolbox of MATLAB (R2014a), was used to isolate the pixels of peanut particles from the background in the score images. The threshold used to separate the peanut particles from the background was 0.10 and was set by successive trials until the optimal value corresponding to the perfect detection of the overall peanut particles. Finally, the position of the detected peanut pixels was extracted as x and y coordinates with the help of the *regionprops* function in Matlab and represented in the corresponding image.

2.5. Data analysis

For performing ICA and *Random ICABlocks*, a data matrix (2000, 199) was created, consisting of 2000 spectra at 199 wavenumbers, randomly selected from the hyperspectral images of 100% peanut and 100% wheat flour. The first 1000 spectra were

of wheat flour whereas the second thousand spectra were of peanut. The *Random ICabyblocks* method distributed the spectra into two blocks randomly, so as to have two blocks representative of the whole data set. ICA models were computed for both blocks, with from 1 to 20 ICs. The random distribution of spectra, followed by *ICabyblocks*, was repeated 10 times. ICs were compared by calculating the correlation coefficients between the signals extracted from each blocks. The optimal numbers of ICs were decided upon by visualizing the decrease in the correlation. The ability of the extracted components to classify peanut and wheat flour was validated from the corresponding concentrations obtained for the same data. Finally, the independent components were merged to create a new signal with the aim of applying it to the hyperspectral images to maximize the contrast between peanut and wheat flour. To construct this new synthetic signal, all the ICs, which corresponded to higher proportions for the wheat flour, were added in negative sense to the ICs, which corresponded to higher proportions for the peanut. For validating the ability of resulting synthetic signal to detect peanuts, a hyperspectral image was simulated with known position of peanut at four different places. The simulated hyperspectral image was made with the random pure peanut and wheat flour spectra. Thus after the validation, the synthetic signal was used directly to enhance the contrast between peanut and wheat flour in the hyperspectral images. The enhanced contrast images were subsequently used for feature extraction in order to have an overview of the distribution of peanut traces in the images.

All data analyses were performed using MATLAB_R2014a version 8.3.0.532 (The MathWorks, Natick, USA) software. The Matlab codes of the ICA by JADE algorithm and *Random ICabyblocks* were downloaded from the web site in Cordella and Bertrand (2014).

3. Results and Discussion

3.1. Spectra

Fig. 1 shows the mean NIR absorbance spectra (before and after pre-treatment) of peanut and wheat flour extracted from hyperspectral images. Initially in the raw spectra (Fig. 1a), higher absorbance values were observed for peanut spectra compared to wheat flour. The smaller particle size of wheat flour (125–100 μm and 212–160 μm) caused a very high light scattering and consequently a low absorbance level whereas, larger particle size of peanut particles (500–1000 μm) caused low scattering and high absorbance levels. To have a more meaningful interpretation of the shape of spectra and to reduce the environmental and texture effects, the absorbance spectra were pre-processed with SNV. As can be seen in Fig. 1b, after pre-processing the spectra showed some clear stable peaks at different wavelengths. Absorbance peaks at 1200 nm, 1395 nm and 1734 nm in the peanut spectra can be related to the presence of long chain fatty acids (Osborne et al., 1993), which give rise to a CH_2 second overtone at 1200 nm, a combination of $2 \times \text{C-H}$ stretch + C-H deformation and C-H at 1395 nm and CH_2 first overtone at 1734 nm. On the wheat flour spectra, the two main peaks, related to OH bonds, are observed: at 1450 nm and 1940 nm. In this case, absorbance at 1450 nm is due to the first overtone of OH stretching and at 1940 nm to the combination of OH stretching and OH bending (Osborne et al., 1993). Another smaller negative peak, at 1580 nm, is caused by O–H stretching first overtone and is related to starch (Osborne et al., 1993). A positive peak at 2030 nm is caused by C=O stretching second overtone and is related to amide function (Osborne et al., 1993). Finally, a clear negative peak at 2100 nm is caused by the combination ($2 \times \text{O-H}$ deformation + $2 \times \text{C-O}$ stretching)

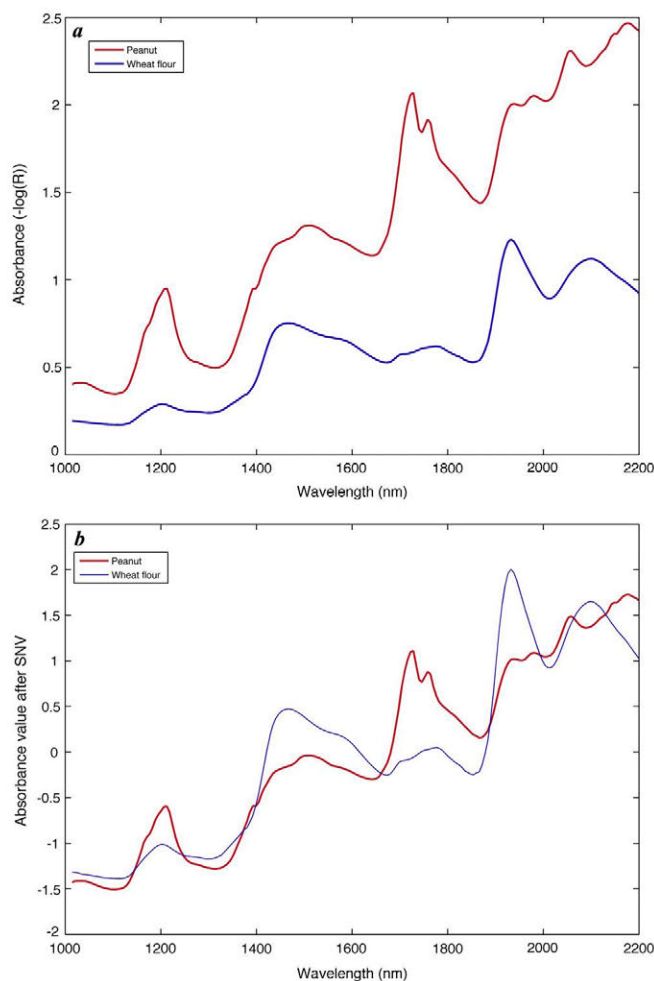


Fig. 1. Absorbance and SNV-treated mean spectra for peanut (thick red) and wheat flour (thin blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and is related to starch (Osborne et al., 1993). These observed spectral differences suggest that ICA could be useful to extract these different source signals. These signals could then be used for the detection of peanut traces in wheat flour samples.

3.2. Random ICA by blocks

Deciding the optimum number of ICs is a primary step to extract signals with ICA. A number larger than the optimal number of ICs can induce noise in the signals whereas extracting a number less than optimal can result in incomplete source separation. A representation of correlation between the blocks is presented in Fig. 2, the colours representing correlation values. In Fig. 2, it can be seen that after the 3rd IC, the correlation decreases for the next few ICs. However, it can also be seen that after the 4th IC, the correlation increases again to 0.90 until the 7th IC. A reason for initial decrease after 3rd ICs was assumed to be that the ICs are not extracted from the two data blocks in the same order (Boiret et al., 2014). After the 7th IC, the correlation further decreases significantly in a very step way. Hence, the optimum numbers of ICs to decompose the data was fixed to 7.

Since in the spectral range of 1000–2200 nm, the peanut particles and wheat flour contain four important compounds i.e. fatty acid, starch, amide and water, it was assumed that there would be four independent components and that the acquired spectra would be linear mixtures of these four independent components.

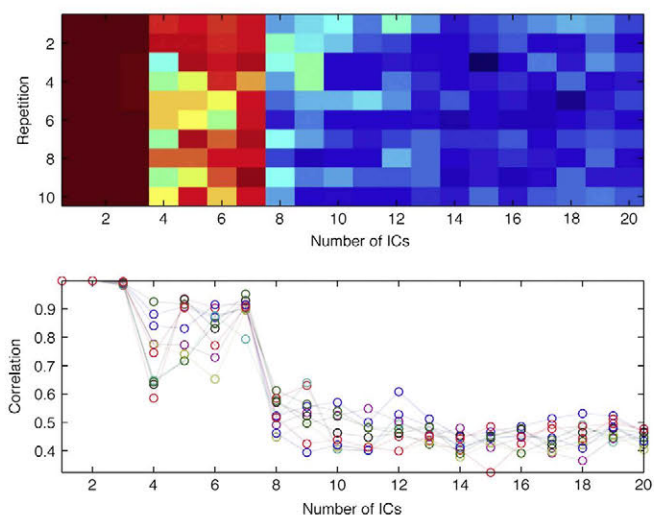


Fig. 2. (a) Representation of the results obtained from Random ICA-by-blocks (blue for the lowest and red for the highest) and correlation values (0–1), (b) correlation values between ICs extracted after 10 permutations of samples between the 2 blocks. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

However, in our work 7 ICs were detected by *Random ICabyblocks* to be optimum. The reason for obtaining extra ICs source signals could be the detection of physical and chemical effects such as the size of the particles and different varieties of peanuts used (Boiret et al., 2014).

3.3. Explanation of source signals and corresponding proportions

The decomposition of the data matrix (1000 wheat flour spectra and 1000 peanut spectra) resulted in 7 ICs. The extracted ICs also provide their proportion in each of the 2000 acquired spectra; hence, both the extracted IC signal and corresponding proportion are presented in Fig. 3.

As can be seen in Fig. 3a–g, all the ICs except IC1 (Fig. 3a) seems to be relevant and can be easily interpreted chemically. IC1 is not related to the chemical information that is represented by the peaks, but by a non-chemical variation, represented more by the baseline drift of the spectra. In fact, every time there is a peak, either in the wheat flour spectrum or peanut spectrum, the corresponding peak summit in IC1 tends towards zero.

IC2 (Fig. 3b) has a peak at 2100 nm caused by the combination ($2 \times \text{O-H}$ deformation + $2 \times \text{C-O}$ stretching) and is related to starch (Osborne et al., 1993). IC3 (Fig. 3c) has a peak at 1940 nm, which can be related to the combination of OH stretching, and OH bending. IC4 (Fig. 3d) presents a peak at 2030 nm, caused by C=O stretching second overtone and is related to the amide function (Osborne et al., 1993). IC5 (Fig. 3e) presents one high peak at 1734 nm and two low peaks at 1200 nm and 1395 nm. These peaks can be related to the presence of long chain fatty acids (Osborne et al., 1993) which give rise to a CH_2 second overtone at 1200 nm, a combination of $2 \times \text{C-H}$ stretch + C-H deformation and C-H at 1395 nm and CH_2 first overtone at 1734 nm. IC6 (Fig. 3f) presents a peak at 1450 nm due to first overtone of OH stretching; and IC7 (Fig. 3g) presents another smaller negative peak, at 1580 nm, caused by O-H stretching first overtone and is related to starch (Osborne et al., 1993).

IC2 and IC7 were related to the starch, and wheat flour contains more starch than do peanuts. This was in accordance from the corresponding proportions obtained for IC2 and IC7 (Fig. 3b and g) where the wheat flour has higher proportions of starch. Similarly, peaks found in IC3 and IC6 (Fig. 3c and f) provide information

regarding higher moisture content in wheat flour than in peanuts. IC5 provides information regarding the fatty acids, which are higher in peanut than in wheat flour. The corresponding concentrations for IC5 (Fig. 3e) shows a high concentration of fatty acid in peanut and shows almost null concentration in wheat flour. Similarly, IC4 (Fig. 3d) reflects a higher concentration of amide for peanuts than in wheat flour.

In an industrial environment, it is not wise to rely on just one chemical compound for classification. Hence, all 7 ICs were used to construct a new synthetic signal, which can simultaneously represent different chemical components and can be used for classification of pixel spectra. Fig. 3h presents the final synthetic signal constructed after taking into account all 7 ICs. The new signal was constructed as the sum of all the ICs with positive proportions for wheat flour, multiplied by -1 , and the all ICs with positive proportions for peanut, multiplied by $+1$. The resulting signal can then be used directly to enhance the contrast between peanut and wheat flour in the hyperspectral images. The final constructed signal contains information about all the main chemical components i.e. fatty acid (1734 nm) starch (1580 nm, 2100 nm), amide (2030 nm), water (1940 nm, 1450 nm) and overtones (1200 nm, 1395 nm). Hence, using this signal alone, a classification can be obtained, as can be seen in the corresponding concentrations obtained for the signal in Fig. 3h, when the signal was applied directly to the data used for decomposing.

3.4. Validation of peanut detection

Application of synthetic signal to the simulated hyperspectral image resulted in a clear contrast difference between peanut and wheat flour pixels. All the four locations where the peanut was added in a supervised way were identified. The peanut pixels appeared to be brighter than the wheat flours. Brightness of the peanut pixels was because of the higher score values obtained for the peanuts, compared to the wheat flour pixels (see Fig. 4).

3.5. Score images and images after feature extraction

ICA-score images were converted from grayscale images to binary (BW) images. The output BW image replaces all pixels in the input image with luminance greater than threshold level (0.10) with the value 1 (white) and replaces all other pixels with the value 0 (black). The next step consisted in image segmentation with a connected components labelling algorithm (Di Stefano and Bulgarelli, 1999) followed by a flood fill method (Lee and Kang, 2010) to extract object contours. The *regionprops* function of Matlab Image Processing Toolbox measures different set of properties for each connected component and allows one to retrieve the centroid coordinates of peanut-dots detected in the image. The use of ICA before applying the feature extraction procedure improves the performance of the image segmentation and contour detection processes, by increasing the contrast between the target points and the general background of the image.

It can be seen that the score values for peanut are higher than for the wheat flour. That is the reason why the peanut particle appears to be brighter than the wheat flour in the score images. After enhancing the contrast, images were used for feature extraction in order to locate the position of the pixels detected as peanut in the image. The results of the application of ICA followed by feature extraction are presented in Figs. 5 and 6. The total numbers of points detected as being peanut in 0.05% adulterated images were 12 and for 0.01% adulterated image were 3. It was logical that with reduction in the amount of adulteration the number of detected pixels decreased.

In our recent earlier work (Mishra et al., 2015), the hyperspectral images were analysed using PCA to detect the peanut traces in

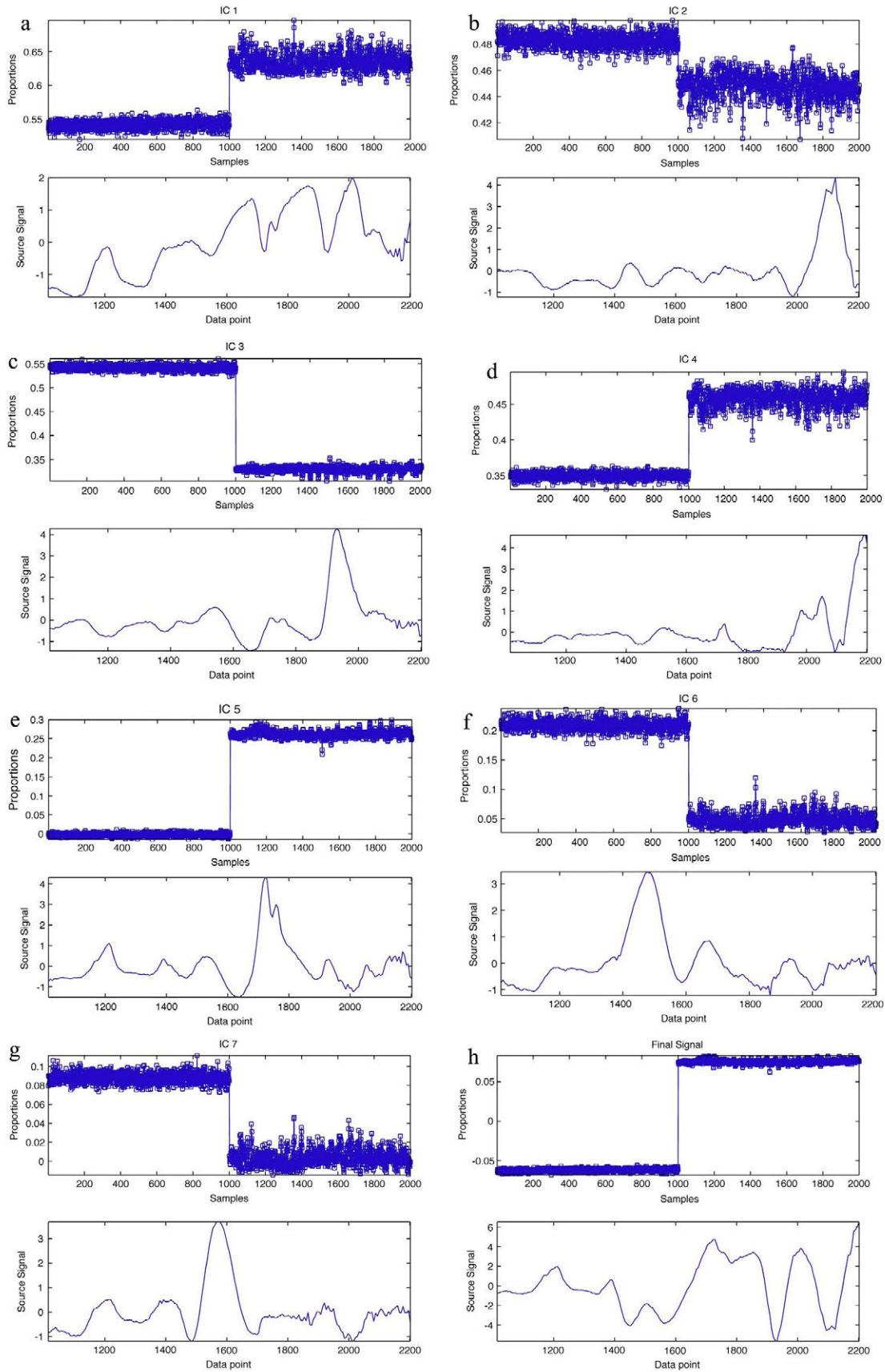


Fig. 3. Representation of IC signals extracted and the corresponding proportions (left block wheat flour and right block peanuts). (a) First Independent Component (IC1), (b) Second Independent Component (IC2), (c) Third Independent Component (IC3), (d) Fourth Independent Component (IC4), (e) Fifth Independent Component (IC5), (f) Sixth Independent Component (IC6), (g) Seventh Independent Component (IC7) and (h) synthetic signal.

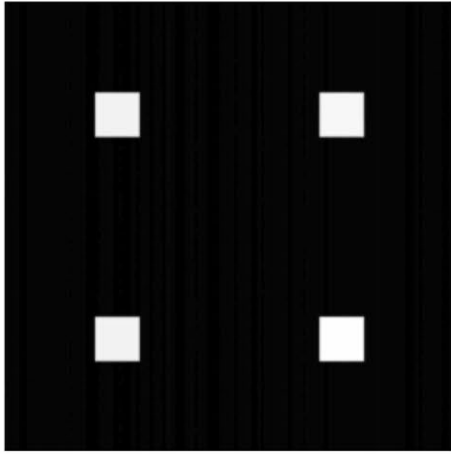


Fig. 4. Validation score image obtained from the application of synthetic signal.

wheat flour (down to 0.01%). The interpretation of the resulting loadings vector was not very straightforward as each PCA loadings vector represented combinations of different phenomena described by the data.

In this present work, ICA proves to be a better processing method as it extracts the different physico-chemical signatures with their corresponding proportion in the peanut and wheat flour mixes. This difference is due to the fundamental difference in the principles of the two methods: while PCA is based on determining the orthogonal directions of maximum dispersion of the samples in the multidimensional space defined by the variables, ICA aims to recover the source signals mixed together in the observed signals. Interpretation was therefore simpler as the ICs directly provided the underlying source signals.

One example of this was the extraction of the overtones of fatty acid as a single independent signal by ICA (Fig. 3e). As can be seen in Fig. 3e, IC5 presents one intense peak at 1734 nm and two small peaks at 1200 nm and 1395 nm. These peaks were related to the presence of long chain fatty acids which give rise to CH_2 first and second overtones and a combination of $2 \times \text{C-H}$ stretch + C-H deformation. The simplified ICs can then be easily used to

construct a synthetic vector as required to highlight a particular component of the system.

The main challenge raised was in using the feature extraction to decide the threshold in order to separate the peanut regions from the background. An incorrect threshold can result in false positives and can lead to misinterpretation of peanut distribution. However, iterative trials and comparison between the two images helped in deciding on the threshold value. In order to avoid this challenge for applying this method in real situations, random hyperspectral images with repetition of the samples could be tested to decide the optimum threshold for separating the peanut pixels from the background.

4. Conclusions

HSI with advanced chemometrics methods is nowadays gaining in importance for detecting of adulteration in various food products. Independent Components Analysis (ICA) has proved to be as a powerful signal-processing technique for solving the blind-source separation (BSS) problem in spectroscopy by resolving the observed signals into source signals, which provides better chemical interpretation of data. Processing of hyperspectral images with ICA can help in detection of foreign materials in food products.

In the present work, an application of ICA to process hyperspectral images is presented. The ability to resolve multivariate hyperspectral images was studied by detecting the adulteration of peanut traces in wheat flour. The *Random ICAblocks* method indicated the optimal number of independent signals to be extracted from the data set. The calculated ICA signals were interpreted and then used to detect peanut in wheat flour. Furthermore, a feature extraction method was applied to have an overview of the spatial location of the peanut traces. The results provided a clear detection of peanut traces. HSI with ICA can be applied for quantitative prediction of the chemical constituents and to study physical properties of samples, with simultaneous representation of their spatial distribution.

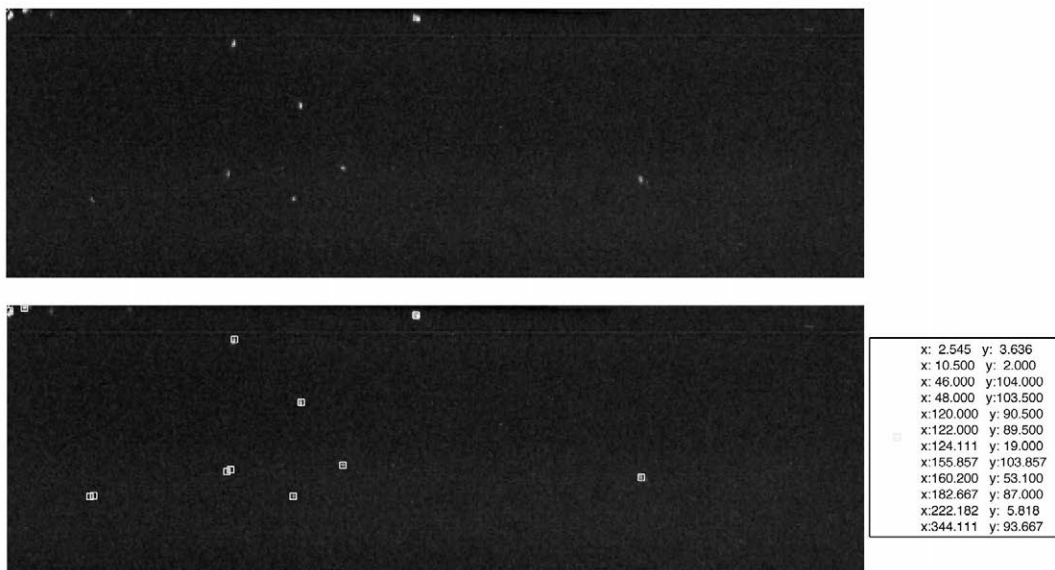


Fig. 5. Score images and extracted features image for 0.05% peanut traces in wheat flour. (146 * 464 pixels).

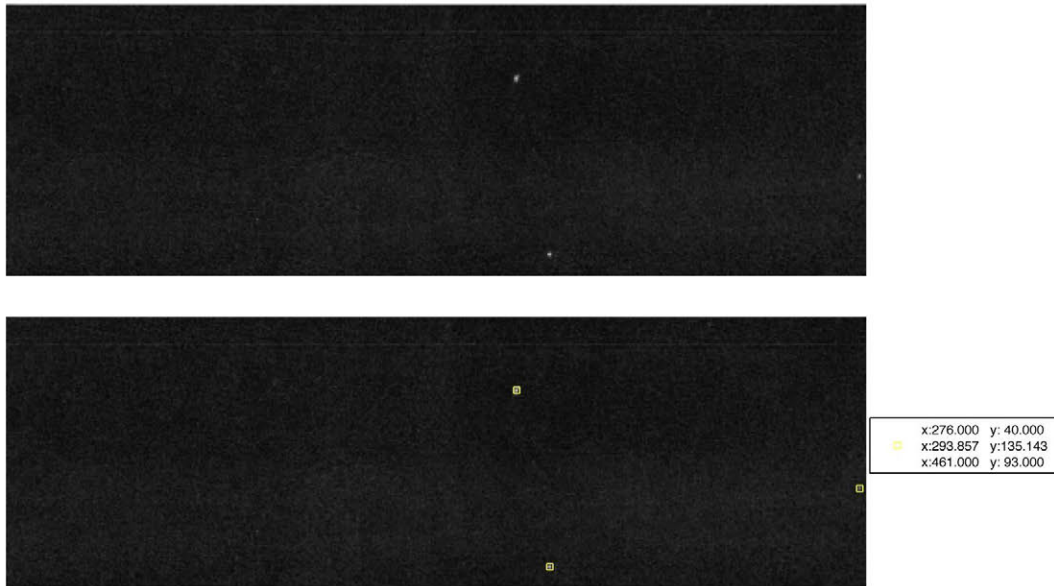


Fig. 6. Score images and extracted features image for 0.01% peanut traces in wheat flour. (146 * 464 pixels).

Acknowledgements

The authors are very grateful for the peanut samples sets used in the work provided by European Commission Joint Research Centre, Institute for Reference Materials and Measurements. We also gratefully acknowledge the time to time technical assistance provided by Daniel Moura (Irstea, UMR ITAP Montpellier) in setting up and implementation of the HSI-camera.

References

- Ammari, F., Bendoula, R., Jouan-Rimbaud Bouveresse, D., Rutledge, D.N., Roger, J.-M., 2014. 3D front face solid-phase fluorescence spectroscopy combined with independent components analysis to characterize organic matter in model soils. *Talanta* 125, 146–152. <http://dx.doi.org/10.1016/j.talanta.2014.02.049>.
- Ammari, F., Redjdal, L., Rutledge, D.N., 2015. Detection of orange juice frauds using front-face fluorescence spectroscopy and independent components analysis. *Food Chem.* 168, 211–217. <http://dx.doi.org/10.1016/j.foodchem.2014.06.110>.
- Barbin, D.F., ElMasry, G., Sun, D.-W., Allen, P., Morsy, N., 2013. Non-destructive assessment of microbial contamination in porcine meat using NIR hyperspectral imaging. *Innovative Food Sci. Emerg. Technol.* 17, 180–191. <http://dx.doi.org/10.1016/j.ifset.2012.11.001>.
- Barnes, R., Dhanoa, M., Lister, S., 1989. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Appl. Spectrosc.* 43, 772–777.
- Bell, A.J., Sejnowski, T.J., 1995. An information-maximisation approach to blind separation and blind deconvolution. *Neural Comput.* 7, 1129–1159.
- Boiret, M., Rutledge, D.N., Gorretta, N., Ginot, Y.-M., Roger, J.-M., 2014. Application of independent component analysis on Raman images of a pharmaceutical drug product: pure spectra determination and spatial distribution of constituents. *J. Pharm. Biomed. Anal.* 90, 78–84. <http://dx.doi.org/10.1016/j.jpba.2013.11.025>.
- Burks, A.W., 2008. Peanut allergy. *Lancet* 371 (9623), 1538–1546. [http://dx.doi.org/10.1016/S0140-6736\(08\)60659-5](http://dx.doi.org/10.1016/S0140-6736(08)60659-5).
- Cardoso, S., 1993. IEE proceedings-F 140, vol. 6, 1993, pp. 362–370.
- Chuang, Y.-K., Hu, Y.-P., Yang, I.-C., Delwiche, S.R., Lo, Y.M., Tsai, C.-Y., Chen, S., 2014a. Integration of independent component analysis with near infrared spectroscopy for evaluation of rice freshness. *J. Cereal Sci.* 60 (1), 238–242. <http://dx.doi.org/10.1016/j.jcs.2014.03.005>.
- Cordella, C.B.Y., Bertrand, D., 2014. SALSIR: a new general chemometric toolbox. *TrAC Trends Anal. Chem.* 54, 75–82. <http://dx.doi.org/10.1016/j.trac.2013.10.009>.
- Chuang, Y.-K., Yang, I.-C., Lo, Y.M., Tsai, C.-Y., Chen, S., 2014b. Integration of independent component analysis with near-infrared spectroscopy for analysis of bioactive components in the medicinal plant *Gentiana scabra* Bunge. *J. Food Drug Anal.* 22 (3), 336–344. <http://dx.doi.org/10.1016/j.jfda.2014.01.021>.
- Crépet, A., Papadopoulos, A., Elegbede, C.F., Ait-Dahmane, S., Loynet, C., Millet, G., Moneret-Vautrin, D.A., 2015. Mirabel: an integrated project for risk and cost/benefit analysis of peanut allergy. *Regul. Toxicol. Pharmacol.* 71 (2), 178–183. <http://dx.doi.org/10.1016/j.yrtph.2014.12.006>.
- Di Stefano, L., Bulgarelli, A., 1999. A simple and efficient connected components labeling algorithm. In: *Image Analysis and Processing. Proceedings International Conference on*, 1999, pp. 322–327. doi: <http://dx.doi.org/10.1109/iciap.1999.797615>.
- Du, Q., 2004. Independent component analysis to hyperspectral image classification, vol. 5546, pp. 366–373, doi: <http://dx.doi.org/10.1117/12.557129>.
- Finkelman, F.D., 2010. Peanut allergy and anaphylaxis. *Curr. Opin. Immunol.* 22 (6), 783–788. <http://dx.doi.org/10.1016/j.coi.2010.10.005>.
- Fu, X., Kim, M.S., Chao, K., Qin, J., Lim, J., Lee, H., Ying, Y., 2014. Detection of melamine in milk powders based on NIR hyperspectral imaging and spectral similarity analyses. *J. Food Eng.* 124, 97–104. <http://dx.doi.org/10.1016/j.jfoodeng.2013.09.023>.
- Herauld, J., Jutten, C., 1986. Space or time adaptive signal processing by neural network models. *AIP Conf. Proc.* 151, 206–211. <http://dx.doi.org/10.1063/1.36258>.
- Huang, Y., Whittaker, A.D., Lacey, R.E., 2001. *Automation for food engineering: food quality quantization and process control. CRC Series in Contemporary Food Science.* CRC Press, Boca Raton, Fla.
- Hyvärinen, A., Oja, E., 1997. A fast fixed-point algorithm for independent component analysis. *Neural Comput.* 9, 1483–1492.
- Jouan-Rimbaud Bouveresse, D., Moya-González, A., Ammari, F., Rutledge, D.N., 2012. Two novel methods for the determination of the number of components in independent components analysis models. *Chemom. Intell. Lab. Syst.* 112, 24–32. <http://dx.doi.org/10.1016/j.chemolab.2011.12.005>.
- Kamruzzaman, M., Sun, D.-W., ElMasry, G., Allen, P., 2013. Fast detection and visualization of minced lamb meat adulteration using NIR hyperspectral imaging and multivariate image analysis. *Talanta* 103, 130–136. <http://dx.doi.org/10.1016/j.talanta.2012.10.020>.
- Kang, S., Lee, K., Son, J., Kim, M.S., 2011. Detection of fecal contamination on leafy greens by hyperspectral imaging. *Proc. Food Sci.* 1, 953–959. <http://dx.doi.org/10.1016/j.profoo.2011.09.143>.
- Kassouf, A., El Rakwe, M., Chebib, H., Ducruet, V., Rutledge, D.N., Maalouly, J., 2014. Independent components analysis coupled with 3D-front-face fluorescence spectroscopy to study the interaction between plastic food packaging and olive oil. *Anal. Chim. Acta* 839, 14–25. <http://dx.doi.org/10.1016/j.aca.2014.06.035>.
- Lee, J., Kang, H., 2010. Flood fill mean shift: a robust segmentation algorithm. *Int. J. Control, Autom. Syst.* 8 (6), 1313–1319. <http://dx.doi.org/10.1007/s12555-010-0617-6>.
- Mahesh, S., Jayas, D.S., Paliwal, J., White, N.D.G., 2015. Hyperspectral imaging to classify and monitor quality of agricultural materials. *J. Stored Products Res.* 61, 17–26. <http://dx.doi.org/10.1016/j.jspr.2015.01.006>.
- Mishra, P., Herrero-Langreo, A., Barreiro, P., Roger, J., Diezma, B., Gorretta, N., Lleó, L., 2015. Detection and quantification of peanut traces in wheat flour by near infrared hyperspectral imaging spectroscopy using principal-component analysis. *J. Near Infrared Spectrosc.* 23 (1), 15–22. <http://dx.doi.org/10.1255/jnirs.1141>.
- Monakhova, Y.B., Tsikin, A.M., Mushtakova, S.P., Mecozzi, M., 2015. Independent component analysis and multivariate curve resolution to improve spectral interpretation of complex spectroscopic data sets: application to infrared spectra of marine organic matter aggregates. *Microchem. J.* 118, 211–222. <http://dx.doi.org/10.1016/j.microc.2014.10.001>.
- Osborne, B., Fearn, T., Hindle, P.H., 1993. *Practical NIR spectroscopy with applications in food and beverage analysis.* Longman Scientific and Technical, Harlow, UK.

- Piqueras, S., Burger, J., Tauler, R., De Juan, A., 2012. Relevant aspects of quantification and sample heterogeneity in hyperspectral image resolution. *Chemom. Intell. Lab. Syst.* 117, 169–182.
- Qin, J., Chao, K., Kim, M.S., Lu, R., Burks, T.F., 2013. Hyperspectral and multispectral imaging for evaluating food safety and quality. *J. Food Eng.* 118 (2), 157–171. <http://dx.doi.org/10.1016/j.jfoodeng.2013.04.001>.
- Rohart, A., Jouan-Rimbaud Bouveresse, D., Rutledge, D.N., Michon, C., 2015. Spectrophotometric analysis of polysaccharide/milk protein interactions with methylene blue using independent components analysis. *Food Hydrocolloids* 43, 769–776. <http://dx.doi.org/10.1016/j.foodhyd.2014.08.007>.
- Rutledge, D.N., Jouan-Rimbaud Bouveresse, D., 2013. Independent components analysis with the JADE algorithm. *TrAC Trends Anal. Chem.* 50, 22–32. <http://dx.doi.org/10.1016/j.trac.2013.03.013>.
- Santos, P.M., Pereira-Filho, E.R., Rodriguez-Saona, L.E., 2013. Rapid detection and quantification of milk adulteration using infrared microspectroscopy and chemometrics analysis. *Food Chem.* 138 (1), 19–24. <http://dx.doi.org/10.1016/j.foodchem.2012.10.024>.
- Wang, G., Ding, Q., Hou, Z., 2008. Independent component analysis and its applications in signal processing for analytical chemistry. *TrAC Trends Anal. Chem.* 27 (4), 368–376. <http://dx.doi.org/10.1016/j.trac.2008.01.009>.
- Wu, D., Shi, H., He, Y., Yu, X., Bao, Y., 2013. Potential of hyperspectral imaging and multivariate analysis for rapid and non-invasive detection of gelatin adulteration in prawn. *J. Food Eng.* 119 (3), 680–686. <http://dx.doi.org/10.1016/j.jfoodeng.2013.06.039>.
- Wu, D., Sun, D.-W., 2013. Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: a review – Part I: fundamentals. *Innovative Food Sci. Emerg. Technol.* 19, 1–14. <http://dx.doi.org/10.1016/j.ifset.2013.04.014>.
- Yu, J.W., Kagan, R., Verreault, N., Nicolas, N., Joseph, L., St. Pierre, Y., Clarke, A., 2006. Accidental ingestions in children with peanut allergy. *J. Allergy Clin. Immunol.* 118 (2), 466–472. <http://dx.doi.org/10.1016/j.jaci.2006.04.024>.