

Journal Pre-proof

Chemometric non-targeted analysis for detection of soybean meal adulteration by near infrared spectroscopy

O.Ye Rodionova, J.A. Fernández Pierna, V. Baeten, A.L. Pomerantsev



PII: S0956-7135(20)30375-3

DOI: <https://doi.org/10.1016/j.foodcont.2020.107459>

Reference: JFCO 107459

To appear in: *Food Control*

Received Date: 29 May 2020

Revised Date: 27 June 2020

Accepted Date: 30 June 2020

Please cite this article as: Rodionova O.Y., Fernández Pierna J.A., Baeten V. & Pomerantsev A.L., Chemometric non-targeted analysis for detection of soybean meal adulteration by near infrared spectroscopy, *Food Control* (2020), doi: <https://doi.org/10.1016/j.foodcont.2020.107459>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier Ltd.

CRedit author statement

O.Ye. Rodionova: Methodology, Investigation, Formal analysis, Writing- Original draft preparation

J.A. Fernández Pierna: Data curation, Investigation, Writing- Reviewing and Editing

V. Baeten: Data curation, Investigation, Writing- Reviewing and Editing

A.L. Pomerantsev: Conceptualization, Software, Visualization, Writing- Reviewing and Editing

Chemometric non-targeted analysis for detection of soybean meal adulteration by near infrared spectroscopy

O.Ye. Rodionova¹, J.A. Fernández Pierna², V. Baeten², A.L. Pomerantsev¹

¹N.N. Semenov Federal Research Center for Chemical Physics, RAS, Kosygin 4, 119991 Moscow, Russia

² Walloon Agricultural Research Centre, Knowledge and valorization of agricultural products Department, Quality and authentication of Products Unit, Gembloux, Belgium.

Keywords: melamine, cyanuric acid, non-protein nitrogen contamination, one-class classifier, SIMCA, limit of detection

Abstract

In this study a quick and efficient routine procedure for food fraud detection by multiple adulterants is presented. Non-targeted analysis employs the Near Infrared (NIR) spectroscopy measurements and one-class classification modeling as the chemometric data processing. The approach is illustrated by the analysis of a large collection of NIR spectra of soybean meal. The clean and contaminated samples are studied. The main advantage of the proposed approach is that it is not aimed at identification and quantification of a specific contaminant. The procedure is designed in such a way that it detects any deviations from the clean samples. The non-targeted analysis has its own limit of detection (LoD). In the study we have presented an approach for LoD assessment. This issue is of great importance for practical applications. The proposed approach can be applied for other types of feed and food products.

Keywords

melamine, cyanuric acid, non-protein nitrogen contamination, one-class classifier, SIMCA, limit of detection

1 Introduction

Soybean meal is one of the most important agricultural raw materials and one of the main ingredients in both feed and food because it is an important source of oil and especially due to its complete protein profile. This protein content dictates the price of the soybean meal in the market: the higher the content, the higher the price (Yang, Han, Wang, Li, Fernández Pierna, Dardenne, & Baeten, 2016). This opens the door to multiple cases of fraud and adulteration of soybean meal with illegal substances. The most popular and studied case is probably the

example of melamine (1,3,5-triazine-2,4,6-triamine), a molecule with a high content of Nitrogen and therefore ideal to boost artificially the protein content of soybean meal (Dorne, Doerge, Vandebroek, Fink-Gremmels, Mennes, Knutsen, Vernazza, Castle, Edler, & Benford, 2013; Hilts, & Pelletier, 2008). Since 2007, multiple cases of adulterated soybean meal with melamine have been reported (Andersen, Turnipseed, Karbiwnyk, Clark, Madson, Gieseke, Miller, Rummel, & Reimschuessel, 2008; Adams, 2008) and the European Commission was obliged to take measures to ban the import from China of food and feed containing soya and soya products, after high levels of melamine were found in Chinese soybean meal. To ensure food and feed safety, many countries set the maximum permitted concentration for melamine in food and animal feed products at 2.5 mg/kg (except for powdered infant formula where the limit is fixed to 1 mg/Kg) (Commission Decision, 2008).

Due to the emergency, a large number of targeted analysis methods based on different platforms for detecting melamine in soybean meal and other food/feed matrices were quickly developed. These included chromatography-based methods such as high performance liquid chromatography (HPLC) (Venkatasami, & Sowa, 2010), gas-chromatography mass-spectrometry (GC-MS) (Pan, Wu, Yang, Wang, Shen, & Zhu, 2013) and LC-MS-MS (Filigenzi, Puschner, Aston, & Poppenga, 2008) or enzyme linked immunosorbent assays (ELISA) (Lei, Shen, Song, Yang, Chevallier, Haughey, Wang, Sun, & Elliott, 2010) among others. However, most of those methods are usually time-consuming, expensive, damaging to the sample and need chemical reagent. Therefore, there is a real need for fast, non-destructive and automatically controlled screening methods that will guarantee quality and security. In this context, vibrational spectroscopic methods, such as Near Infrared (NIR) spectroscopy, can play an important role due to their capacity to be a rapid, non-destructive and non-polluting method that requires minimal or no sample preparation (Downey, 1996; Blanco, & Villarroya, 2002; Dardenne, & Baeten, 2002; Baeten, Fernández Pierna, Lecler, Abbas, Vincke, Minet, Vermeulen, & Dardenne, 2016; Sun, 2018) Many studies have investigated the feasibility of using NIR to detect melamine (Ferreira, Galão, Pallonea, & Poppi, 2014; Mauer, Chernyshova, Hiatt, Deering, & Davis, 2009; Balabin, & Smirnov, 2011; Haughey, Graham, Cancouët, & Elliott, 2013; Jawaid, Talpur, Sherazi, Nizamani, & Khaskheli, 2013; Abbas, Lecler, Dardenne, & Baeten, 2013; Chen, Yang, & Han, 2013; Yang, Wang, Han, Li, & Liu, 2014; Fernández Pierna, Abbas, Lecler, Hogrel, Dardenne, & Baeten, 2015; Fernández Pierna, Vincke, Baeten, Grelet, Dehareng, & Dardenne, 2016; Li, Kang, Shi, & Liu, 2016).

Adulterations with new, unknown illegal ingredients continue to occur from time to time. It is, then, nearly impossible to make an exhaustive analysis of all potential adulterants in

soybean meal using targeted detection methods with the ever-changing pattern of adulteration. For this reason, methods able to work in an untargeted way are needed. In recent years, untargeted detection methods based on NIR spectroscopy have shown great application potential for detecting contaminants in agro-food products. Non-targeted analysis is an effective tool for combating food fraud (López, Trullols, Callao, & Ruisánchez, 2014; Baeten, Vermeulen, Fernández Pierna, & Dardenne, 2014). Using this approach, the product is analyzed as a whole object. We do not quantify specific adulterants in a product and do not verify whether specific product characteristics match claims. A similar approach is now popular in pharmaceutical chemistry, where counterfeit drug detection is often based on the analysis of the whole remedy profile (e.g. NIR, or Raman spectra), but not on the assessment of an active pharmaceutical ingredient (API) concentration or on the presence of a specific excipient (Rodionova, Titova, Balyklova, & Pomerantsev, 2019).

Various analytical methods, such as non-targeted isotope and trace element analysis, molecular spectroscopy like mid infrared, near infrared and Raman spectroscopy, non-targeted mass-spectrometry, and many others, are used for food fraud detection. The necessity to use these methods is primarily caused by the variety of contaminants and by the evident inability to foresee the type of an adulterant in a future routine testing. Collection of data (spectra or other fingerprints) yielded in the course of non-targeted analysis is usually rather complex and therefore it demands special mathematical/statistical processing for extraction of useful information. A special class of pattern recognition methods, called one-class classifiers (OCC) or class modeling is used (Rodionova, & Pomerantsev, 2020) for these purposes. The key feature in implementation of such methods is a set of the objects which effectively represents a class of the authentic, non-adulterated objects. This class is called the target class.

The goal of the study is the presentation of an approach for food fraud detection that employs rapid NIR measurements and OCC modeling in the case when adulterants are numerous. The approach is illustrated by the chemometric analysis of a large collection of NIR spectra of soybean meals. The clean and contaminated samples are studied. The data driven soft independent modeling by class analogy (DD-SIMCA) method is used for class modeling.

An additional objective is to develop a novel approach for the assessment of the Limit of Detection (LoD) in the non-targeted method. This characteristic is of great importance for any analytical technique since it provides the limits of applicability of the method in practice. To our knowledge, this is the first attempt to evaluate LoD for a qualitative technique.

2 Materials and methods

2.1 DD-SIMCA

Data driven SIMCA (DD-SIMCA) (Pomerantsev, & Rodionova, 2014) is a modification of the well-known SIMCA approach (Wold, 1976). In this technique, the SIMCA model is based on the principal component analysis (PCA) decomposition (Pearson, 1901; Bro, & Smilde, 2014) of training matrix \mathbf{X}

$$\mathbf{X} = \mathbf{T}\mathbf{P}^t + \mathbf{E}, \quad (1)$$

where, \mathbf{X} is the $(I \times J)$ matrix of predictors/fingerprints for I samples described by J variables. $\mathbf{T} = \{t_{ik}\}$ is the $(I \times A)$ *score* matrix, which rows are the coordinates of the projections of data points on the principal component space. \mathbf{P} is the $(J \times A)$ *loading* matrix, which orthonormal columns define the orientation of the principal components (PCs). \mathbf{E} is the $(I \times J)$ *residual* matrix, and A is the *number of principal components*, PCs. The results of decomposition are used for calculation of two relevant statistics. They are the orthogonal distance (OD) and the score distance (SD). OD is the squared Euclidian distance between sample i and the scores subspace. It is calculated as the sum

$$q_i = \sum_{j=1}^J e_{ij}^2 \quad i=1, \dots, I \quad (2)$$

of the squared residuals presented in matrix \mathbf{E} defined in Eq.(1). SD is the squared Mahalanobis distance calculated for each sample i by the formula

$$h_i = \sum_{a=1}^A \frac{t_{ia}^2}{\lambda_a}, \quad i=1, \dots, I. \quad (3)$$

where t_{ia} is an element of matrix \mathbf{T} defined in Eq.(1), and $\lambda_a = \sum_{i=1}^I t_{ia}^2$ is the eigenvalue. In the result, each sample can be plotted in the so-called *Distance plot* in the coordinates q/q_0 vs. h/h_0 .

It has been shown (Pomerantsev, 2008) that both distances are well approximated by the scaled chi-squared distribution, and the full distance (FD) can be calculated as

$$f = N_h \frac{h}{h_0} + N_q \frac{q}{q_0} \propto \chi^2(N_f) \quad (4)$$

where h_0 and q_0 are the scaling factors, N_h and N_q are the numbers of degrees of freedom (DoF), and $N_f = N_h + N_q$. The parameters are considered unknown, and are estimated using the training dataset. For this reason the approach is called data driven.

DD-SIMCA establishes the decision rule that determines whether a new sample belongs to the target class. This is determined by employing a cut-off value

$$f \leq f_{\text{crit}}, \quad \text{where } f_{\text{crit}} = \chi^{-2}(1 - \alpha, N_f), \quad (5)$$

which delineates an acceptance area in the Distance plot. Here χ^{-2} is the quantile of the chi-squared distribution. Samples for which FD satisfies Eq.(5) are attributed to the target class (Rodionova, & Pomerantsev, 2020).

2.2 Figures of Merit

Performance of an OCC model is first of all characterized by the ability to correctly attribute target class samples. The DD-SIMCA method has a possibility to set a priori a desired risk of the wrong rejection of the target samples using Eq.(5). Parameter α is the type I error, that is a share of misclassified target samples. The complementary characteristic is model *sensitivity* (SEN), the fraction of the correctly classified samples (true positive) from all the samples for the target class,

$$\text{SEN} = \frac{\text{Number of True Positive}}{\text{Number of Target samples}}, \quad (6)$$

Sensitivity is calculated for the target samples separately for the training and test sets. The consistency between the training and test values of SEN helps to evaluate the complexity of the model, i.e. select the number of PCs. It is important to note that α -value is selected before the model development and sensitivity is calculated after that. In case the model complexity is selected using the results of classification of the target samples only, such an approach is called *rigorous* (Rodionova, Oliveri, & Pomerantsev, 2016). In the absence of alien (non-target) classes this is the only way for validation of the results of classification. In the absence of alien classes, it is impossible to estimate the type II error β , which is defined as the error to attribute an alien object to the target class (false positive). The complimentary characteristic is *specificity*, *SPS*, which is preferably to calculate for each alien class separately. SPS is the fraction of samples from an alien class, which is correctly classified as the non-member of the target class (true negative).

$$\text{SPS} = \frac{\text{Number of True Negative}}{\text{Number of Alien samples}}, \quad (7)$$

If the model complexity and the cut-off level are established using both characteristics, SENS and SPS, such an approach is called *compliant* (Rodionova, et al, 2016). Further for the development of a final model we will use the compliant approach.

2.3 Dataset description

Dataset *Soya* comprises 1200 NIR spectra of clean soybean meals at 700 wavelengths (1100 - 2498 nm). The samples present the natural variability in the chemical composition and particle size. These variations are reflected in the NIR spectra of clean soybeans, which also demonstrate a large variability. This is due to the fact that the dataset used in this study comes from a large soybean meal dataset comprising spectra measured with different Foss instruments and coming from all around the world, from different providers and varieties. In the result, the training set can be viewed as a collection of several subsets, as well as individual extreme samples. The *Soya* dataset is divided randomly on the training set of 800 samples and the test set of 400 samples. In case the number of samples is more than 1000, such an approach is appropriate. Thus the entire matrix consists of 1200 rows (samples) and 700 columns (wavelengths).

2.4 Contaminated samples

118 adulterated samples are prepared using various concentration of melamine (Mel), cyanuric acid (CA), and their mixtures (Mix). Together with 10 non-contaminated samples they are used in *Validation* and *Routine* datasets.

Table 1 *Validation* and *Routine* datasets

Contamination	Validation set (V)		Routine set (R)	
	% of contaminant added	Number of samples	% of contaminant added	Number of samples
No adulterant	-----	6	-----	4
Melamine (Mel)	0.53-6.00	29	0.53-6.03	11
Cyanuric acid (CA)	0.50-6.04	31	1.48-5.55	9
Mel+CA (Mix)	0.50-6.05	34	0.54-2.62	6

The *Validation* set consists of 100 NIR spectra that include 6 clean samples and 94 contaminated samples (columns 1 and 2 in Table 1). This set is used in the compliant approach for tuning the SIMCA model.

The *Routine* set consist of thirty NIR spectra of clean and contaminated samples (columns 3 and 4 in Table 1). This set simulates routine testing and its results do not have an influence on the SIMCA model complexity, and on the selection of the cut-off level.

Details on the procedure of the synthetic samples' preparation and chemical analysis are presented in (Abbas, et al, 2013).

2.5 Spectra acquisition

Spectroscopic analyses of the adulterated samples were performed in duplicate using an XDS spectrometer (Foss). Spectra were collected between 400 and 2498 nm with an interval of 2 nm and by co-adding 32 scans. Spectra were collected as log 1/R.

2.6 Software

All calculations are performed using Chemometrics Add-In for the Microsoft Excel (Pomerantsev, 2014) and DD-SIMCA – A MATLAB GUI tool (Zontov, Rodionova, Kucheryavskiy, & Pomerantsev, 2017).

3 Results and discussion

3.1 Model complexity

In case only clean soybean samples are analyzed, the SIMCA model with 4 PCs is an optimal classification. For the selected value of the type I error $\alpha=0.01$, sensitivity equals 97 % for the training set, and 98 % for the test set (Fig. 1, the red dashed line). At the same time, classification of the adulterated samples from the validation set yields rather low specificity, which equals 85 %. Fourteen contaminated samples are wrongly attributed to the target class. This demonstrates that the rigorous approach to the model optimization provides a low specificity. This can be explained by the fact that the model accounts for the main common properties of a large heterogeneous training set.

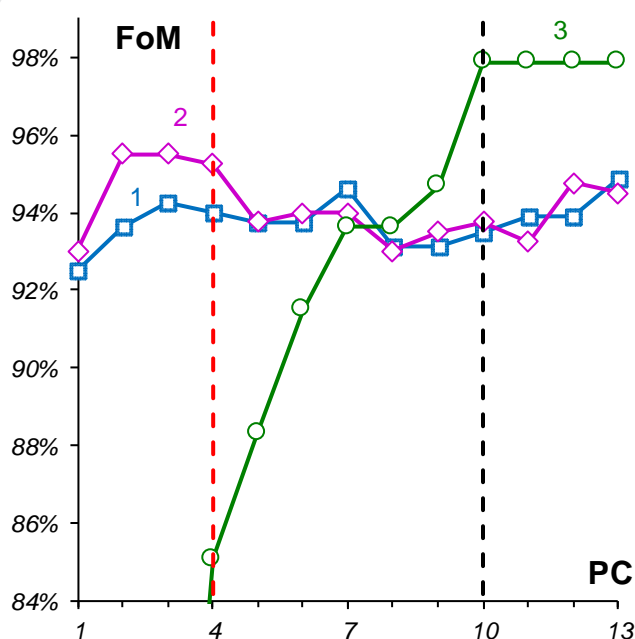


Fig. 1. Figures of merit (FoM) for the SIMCA model vs number of PC: 1 - training SEN, 2 - test SEN, 3-

195

196 To reach the highest possible model specificity we increase the model complexity till 10
 197 PCs (Fig. 1, the black dashed line). For this model only two adulterated samples (V9 and V10)
 198 are wrongly attributed to the target class. This corresponds to specificity about 98 %. Six clean
 199 samples from the validation set are classified properly.

200 The selected model is applied to the *Routine set* to verify how the constructed model
 201 classifies samples in routine testing. Again two samples with a low concentration of
 202 contaminants, R17 and R19, are wrongly attributed to the target class. Fig. 2a represents a
 203 general overview of the classification results for the *Validation* and *Routine* sets. Samples are
 204 located in accordance with their SD and OD values (Eqs.(2), (3)). The critical FD is $f_{crit}=13.28$ in
 205 accordance with Eq.(5). The greatest FD value 3566.6 is achieved for a sample from the
 206 *Validation* set; the sample is contaminated with 1.08% of melamine and 3.42 % of cyanuric acid.
 207 Its spectrum is depicted in Fig. 3 (curve 4). The axes in Fig. 2a are logarithmically transformed
 208 to present all the adulterated samples. The general trend reflects the dependence between
 209 contamination and FD –the higher the contamination, the greater FD is, and, respectively, the
 210 farther is the location in the distance plot. Of course, the greatest interest is for the samples
 211 located near the acceptance border. These are the samples with a low level of contamination that
 212 are shown in the inset of Fig. 2a.

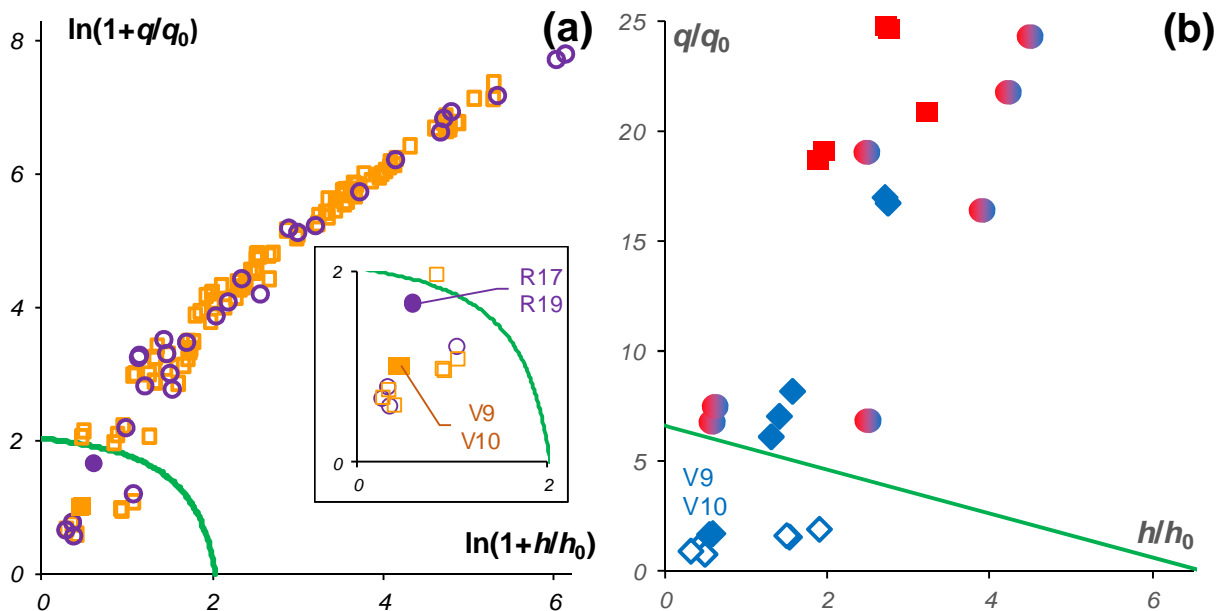


Fig. 2 . The SIMCA results of with PCs=10, $\alpha=0.01$, presented in Distance plots.

(a) *Validation* set (the open yellow squares), *Routine* set (the open violet circles); the closed violet circle stand for misclassified samples; (b) *Validation* set: Melamine (closed red squares), CA (the open blue diamonds), mixtures (the closed bi-colored circle); closed blue diamonds stand for misclassified samples

The correspondent part of the Distance plot is shown in Fig. 2b. It presents the *Validation* set. The non-contaminated samples (the open diamonds) and the contaminated samples wrongly attributed to the target set (two closed blue diamonds, V9 and V10) are located inside the acceptance area bounded by the threshold (the green solid line). All other samples are located beyond the threshold. As expected, the samples with low concentration of adulterants are harder to reveal. In Fig. 2b the samples marked as the closed blue diamonds are contaminated by CA only, the red closed squares stand for the samples adulterated with melamine only, and the closed bi-colored circles represent the samples with both adulterants mixed. As can be seen the 'pure' melamine contamination can be easily discriminated from the clean samples. This cannot be said about the low CA concentrations (0.5 % CA), nor about the low mixture contaminants (0.37 % Mel, 0.13 % CA).

The reason why two contaminated samples (V10 and V9 with 0.53 % of CA) are located inside the acceptance area can be explained as follows. In classifying samples we, in some sense, assess the main similarity between the spectral characteristics. In our case, the difference between the clean soybean samples – the blue spectrum, 1, and the green spectrum, 2, in Fig. 3 – is higher than that for the contaminated sample (the red spectrum 3).

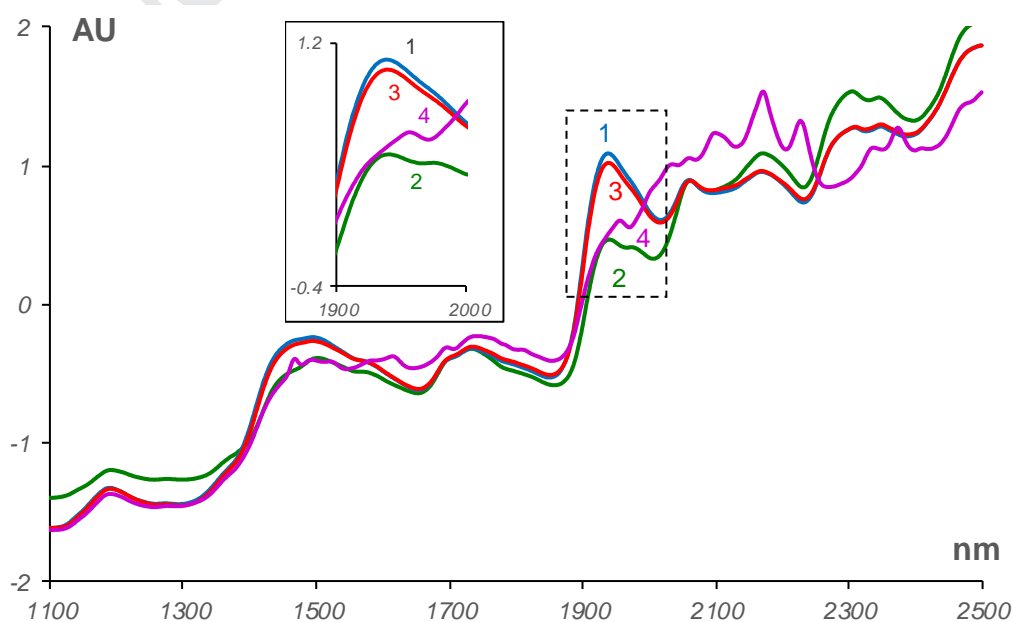


Fig. 3. SNV corrected NIR spectra (Absorbance (Log 1/R) versus Wavelength (nm)). Spectra 1 (blue) and 2 (green) stand for clean samples, spectra 3 (red) and 4 (pink) stand for contaminates samples

More systematic analysis of limits of detection is presented in the next section.

3.2 Limit of detection

The presented results demonstrate a relation between the adulterant concentrations and the FD value, though this dependence is not so straightforward. We model this relationship using the popular response surface technique (Bezerra, Santelli, Oliveira, Villar, & Escaleira, 2008) utilizing the quadratic response function

$$F(\mathbf{z}) = z_0 + \mathbf{bz}^t + \mathbf{zAz}^t \quad (8)$$

where vector \mathbf{z} contains the adulterants concentrations: $z_1=[\text{Mel}]$, $z_2=[\text{CA}]$. Variable z_0 , vector \mathbf{b} , and matrix \mathbf{A} comprise six unknown parameters, which are estimated to obtain the best agreement between the known FD values, f , and the response function, F ,

$$\|F(\mathbf{z}) - f\|^2 \rightarrow \min \quad (9)$$

Function $F(\mathbf{z})$ can be used to predict the FD value for the given concentrations of contaminants.

Considering the whole range of z 's and FDs, we come to the conclusion that the quadratic function, given in Eq. (8), is inappropriate since the modeled response is highly nonlinear, exponential, or Gaussian. However, within the range we are mostly interested in, namely, at a small level of contaminants ($[\text{Mel}] < 1.5$, $[\text{CA}] < 1.5$), the quadratic model works rather well.

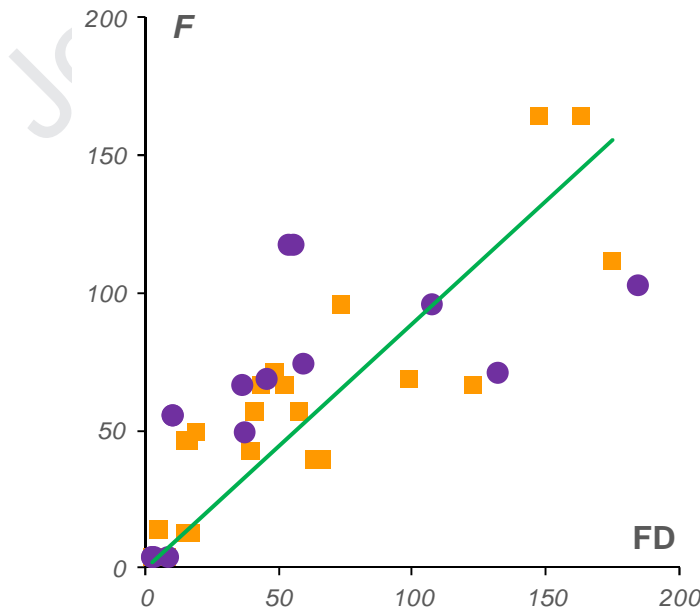


Fig. 4. Response function F modeling the FD values. The yellow squares are for the *Validation* set, the violet circles stand for the *Routine* set.

Fig. 4 shows the results of the response modeling. The yellow squares represent the *Validation* set that is used for the model training; the violet circles stand for the *Routine* set that is utilized for the response model validation; the green line is the trend between F and f . In the plot we can see the large deviations from the trend. Certainly, they are explained by the high variations observed in the clean spectra of soybean that have been discussed above.

The obtained function $F(\mathbf{z})$ can be used to calculate the limits of detection (LoD), which are the Mel and CA concentrations that cannot be revealed by the method presented. The LoD border is obtained from the following equation

$$F(\text{Mel}, \text{CA}) = f_{\text{crit}} \quad (10)$$

where f_{crit} is defined in Eq. (5). The curve that delineates the LoD area is shown in Fig. 5. The Mel and CA concentrations located below the border cannot be detected.

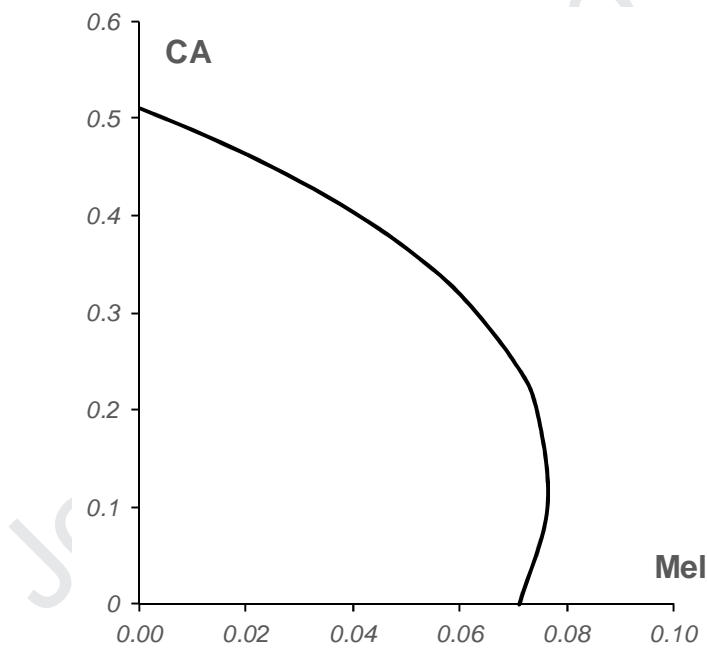


Fig. 5. The LoD border

From this plot, we can conclude that adulterations with Mel and CA are not equivalent. The cut-off concentration of CA is about 0.5 %, while for Mel it is much lower, about 0.07%. This means that the proposed method, combination of NIR spectroscopy and SIMCA, has higher specificity related to melamine than to cyanuric acid. Certainly, the LoD border given in Fig. 5 should not be considered exactly due to the mentioned variability in the spectra of clean soybean samples. Unfortunately we do not have samples with very low concentrations of contaminants for testing the proposed LoD.

3.3 Comments on the Soya dataset

In trying to develop a universal classification model we compiled a Soya dataset with very different soybean samples. From the statistical point of view the Soya dataset is not an entire set, but a collection of several, at least two, different subsets, and maybe several outliers. Applying the distance plot and the recently proposed sequential outlier detection procedure (Rodionova, & Pomerantsev, 2020), it is possible to extract the core group consisting of 1076 samples. The remaining 124 samples are considered as *outliers*. In Fig. 3a sample #1 comes from the main group and sample #2 belongs to the outliers. At the same time, it should be underlined that model based on the core samples wrongly classifies half of the clean samples in the Validation set as the extraneous ones, but the gain in the model specificity is rather small. We always have a trade-off between higher sensitivity and high specificity. The appropriate decision is to collect a representative set of samples, which are similar to the outlier subset and develop two models for revealing the contaminated samples.

4 Conclusions

A quick and efficient routine procedure for revealing adulterated soybean meal is presented. It is based on NIR spectra acquisition and chemometric data processing using the DD-SIMCA method. The procedure is based on the collection of a large dataset of clean soybean meal spectra and on a small set of synthesized samples. The procedure provides a possibility to reveal melamine, cyanuric acid and mixed contaminants simultaneously.

The main advantage of the proposed approach is that it is not aimed at identification and quantification of a specific contaminant, as is usually done for revealing melamine in various products. The procedure is designed in such a way that it detects any deviations from the clean samples. This allows us to foresee that the procedure will be valid in cases where some new and unexpected contaminant is used. Of course, this is provided a signal from this contaminant presents in the acquired spectra. The same approach can be applied for other types of feed and food products.

A non-targeted analysis has its own limit of detection. In the study we have presented the approach for LoD assessment. This issue is of great importance for practical applications.

Declaration of competing interest

None.

Acknowledgments

This work was initiated in the frame of the IAEA/FAO projects D5240 and G42007. OR and AP acknowledge the support within the Russian state assignment No AAAA-A18-118020690203-8. The authors want to thank also Bernard Lecler from the CRA-W for his NIR support.

References

- Abbas, O., Lecler, B. , Dardenne P., Baeten, V. (2013) Detection of melamine and cyanuric acid in feed ingredients by near infrared spectroscopy and chemometrics. *J. Near Infrared Spectroscopy*, 21(3) 183-194.
<https://doi.org/10.1255/jnirs.1047>
- Adams, M. (2008) Melamine Found Contaminating Soy Meal Fed to Organic Chickens (2008) Natural News December 2.
https://www.naturalnews.com/news_000571_melamine_organic_chickens_china.html
- Andersen, W. C., Turnipseed, S. B., Karbiwnyk, C. M., Clark, S. B., Madson, M. R., Giesecker, C. M., Miller, R. A., Rummel, N. G. Reimschuessel, R. (2008). 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, 4340-4347.
<https://doi.org/10.1021/jf800295z>
- Baeten V., Vermeulen Ph., Fernández Pierna, J.A. & Dardenne, P. (2014). From targeted to untargeted detection of contaminants and foreign bodies in food and feed using NIR spectroscopy. *New Food*, 17(3), 16-23.
- Baeten, V., Fernández Pierna, J.A., Lecler, B., Abbas, O., Vincke, D., Minet, O., Vermeulen, P., Dardenne, P. (2016) Near Infrared spectroscopy for food and feed: a mature technique. *NIR news*, 27 (1), 4-6.
<https://doi.org/10.1255%2Ffnirn.1573>
- Balabin, R.M., Smirnov, S.V. (2011) 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) 562-568.
<https://doi.org/10.1016/j.talanta.2011.04.026>
- Bezerra, M.A., Santelli, R. E., Oliveira, E.P., Villar, L.S., Escaleira, L.A. (2008) Response surface methodology (RSM) as a tool for optimization in analytical chemistry,

- 327 *Talanta*, 76(5), 965-977,
 328 <https://doi.org/10.1016/j.talanta.2008.05.019>.
- 329 Blanco, M., Villarroya, I. (2002) NIR spectroscopy: a rapid-response analytical tool, *Trends in*
 330 *Analytical Chemistry*, 21 (4), 240 – 250.
 331 [https://doi.org/10.1016/S0165-9936\(02\)00404-1](https://doi.org/10.1016/S0165-9936(02)00404-1)
- 332 Bro, R., Smilde, A.K. (2014) Principal component analysis, *Anal. Methods*, 2812–2831.
 333 doi:10.1039/c3ay41907j
- 334 Chen, L., Yang, Z., Han, L. (2013) A Review on the Use of Near-Infrared Spectroscopy for
 335 Analyzing Feed Protein Materials. *Applied Spectroscopy Reviews*, 48(7) 509-522,
 336 <https://doi.org/10.1080/05704928.2012.756403>.
- 337 Commission Decision (2008). 2008/757/EC imposing special conditions governing the import of
 338 products containing milk or milk products originating in or consigned from China (2008a).
 339 *Official Journal of the European Union*, No. L 259/10-11.: Commission Decision.
 340 <https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX:32008D0798>
- 341 Dardenne, P., Baeten, V. (2002) Spectroscopy: developments in instrumentation and analysis,
 342 *Grasas y Aceites*, 53 (1), 45-63.
 343 <https://doi.org/10.3989/gya.2002.v53.i1.289>
- 344 Dorne, J.L., Doerge, D.R., Vandenbroeck, M., Fink-Gremmels, J., Mennes, W., Knutsen, H.K.,
 345 Vernazza, F., Castle, L., Edler, L., Benford, D. (2013) Recent advances in the risk
 346 assessment of melamine and cyanuric acid in animal feed. *Toxicology and Applied*
 347 *Pharmacology*, 270, 218-229, 2013.
 348 <https://doi.org/10.1016/j.taap.2012.01.012>
- 349 Downey, G. (1996) Authentication of Food and Food Ingredients by near Infrared Spectroscopy,
 350 *J. Near-infrared Spectroscopy*, 4 47–61.
 351 <https://doi.org/10.1255/jnirs.75>
- 352 Fernández Pierna, J. A., Abbas, O., Lecler, B. , Hogrel, P. , Dardenne, P., Baeten, V. (2015)
 353 NIR fingerprint screening for early detection of non-conformity at feed mills. *Food*
 354 *Chemistry*, 189, 2-12.
 355 <https://doi.org/10.1016/j.foodchem.2014.09.105>
- 356 Fernández Pierna, J. A., Vincke, D., Baeten, V. , Grelet, C. , Dehareng, F. , Dardenne, P. (2016)
 357 Use of a multivariate moving window PCA for the untargeted detection of contaminants in
 358 agro-food products, as exemplified by the detection of melamine levels in milk using
 359 vibrational spectroscopy, *Chemom. Intell. Lab. Syst.*, 152, 157-162
 360 <https://doi.org/10.1016/j.chemolab.2015.10.016>

- Ferreira, D.S., Galão, O.F., Pallonea, J.A.L., Poppi, R.J. (2014) Comparison and application of near-infrared (NIR) and mid-infrared (MIR) spectroscopy for determination of quality parameters in soybean samples, *Food Control*, 35, 227-232, <https://doi.org/10.1016/j.foodcont.2013.07.010>.
- Filigenzi, M. S., Puschner, B., Aston, L.S., Poppenga, R.H. (2008) Diagnostic Determination of Melamine and Related Compounds in Kidney Tissue by Liquid Chromatography/Tandem Mass Spectrometry. *J. Agr. Food Chem.*, 56, 7593-7599. <https://doi.org/10.1021/jf801008s>
- Haughey, S.A., Graham, S.F., Cancouët E., Elliott, C.T., (2013) The application of Near-Infrared Reflectance Spectroscopy (NIRS) to detect melamine adulteration of soya bean meal. *Food Chemistry*, 136(3-4), 1557-1561 <https://doi.org/10.1016/j.foodchem.2012.01.068>.
- Hilts, C., Pelletier, L. (2008) Background Paper on Occurrence of Melamine in Foods and Feed. Report for the WHO Expert Meeting on Toxicological and Health Aspects of Melamine and Cyanuric Acid: https://www.who.int/foodsafety/fs_management/Melamine_3.pdf (last access April 2020).
- Jawaid, S., Talpur, F.N., Sherazi, S.T.H., Nizamani, S.M., Khaskheli, A.A. (2013) Rapid detection of melamine adulteration in dairy milk by SB-ATR–Fourier transform infrared spectroscopy. *Food Chemistry*, 141(3), 3066-3071. <https://doi.org/10.1016/j.foodchem.2013.05.106>
- Lei, H., Shen, Y., Song, L., Yang, J., Chevallier, O. P., Haughey, S. A., Wang, H., Sun, Y., Elliott, C.T., (2010), Hapten synthesis and antibody production for the development of a melamine immunoassay. *Anal. Chim. Acta*, 665, 84-90. <https://doi.org/10.1016/j.aca.2010.03.007>
- Li, Q., Kang, X., Shi, D., Liu, Q. (2016) Determination of Melamine in Soybean Meal by Near-Infrared Imaging and Chemometrics. *Analytical Letters* 49(10), 1564-1577 <https://doi.org/10.1080/00032719.2015.1118482>
- López, M.I., Trullols, E., Callao, M.P., Ruisánchez, I. (2014). Multivariate screening in food adulteration: Untargeted versus targeted modelling. *Food Chemistry*, 147, 177–181. <https://doi.org/10.1016/j.foodchem.2013.09.13>
- Mauer, L.J. , Chernyshova, A.A. , Hiatt, A., Deering, A., Davis, R. (2009) Melamine Detection in Infant Formula Powder Using Near- and Mid-Infrared Spectroscopy. *J. Agricultural and Food Chemistry*, 57(10), 3974-3980. <https://doi.org/10.1021/jf9024163>

- Pan, X.D., Wu, P.G., Yang, D.J., Wang, L.Y., Shen, X.H., Zhu, C. Y., (2013) Simultaneous determination of melamine and cyanuric acid in dairy products by mixed-mode solid phase extraction and GC–MS, *Food control*, 30, 545-548
<https://doi.org/10.1016/j.foodcont.2012.06.045>.
- Pearson, K., (1901) On Lines and Planes of Closest Fit to Systems of Points in Space, *Philosophical Magazine*, 2 (11), 559–572.
[doi:10.1080/14786440109462720](https://doi.org/10.1080/14786440109462720).
- Pomerantsev, A. (2008) Acceptance areas for multivariate classification derived by projection methods, *J. Chemom.*, 22, 601-609.
<https://doi.org/10.1002/cem.1147>
- Pomerantsev, A.L. (2014) Chemometrics in Excel, Hoboken: John Wiley & Sons, Inc.,
<https://doi.org/10.1002/9781118873212>
- Pomerantsev, A.L., Rodionova, O.Ye. (2014) Concept and role of extreme objects in PCA/SIMCA, *J. Chemom.*, 2014, 28, 429–438.
<https://doi.org/10.1002/cem.2506>
- Rodionova, O.Ye., Oliveri, P., Pomerantsev, A.L. (2016). Rigorous and compliant approaches to one-class classification. *Chemom. Intell. Lab. Syst.* 159, 89-96.
<https://doi.org/10.1016/j.chemolab.2016.10.002>
- Rodionova, O.Ye., Pomerantsev, A.L. (2020) Detection of Outliers in Projection-Based Modeling", *Anal. Chem.*, 92, 2656–2664
<https://doi.org/10.1021/acs.analchem.9b04611>
- Rodionova, O.Ye., Pomerantsev, A.L. (2020), Chemometric tools for food fraud detection: The role of target class in nontargeted analysis, *Food Chemistry*, 317, 126448
<https://doi.org/10.1016/j.foodchem.2020.126448>
- Rodionova, O.Ye., Titova, A.V., Balyklova, K.S., Pomerantsev, A.L. (2019). Detection of counterfeit and substandard tablets using non-invasive NIR and chemometrics - a conceptual framework for a big screening system; *Talanta*, 120150.
<https://doi.org/10.1016/j.talanta.2019.12015>
- Sun, D.W. (2018) Modern Techniques for Food Authentication (2nd ed.), Amsterdam: Academic Press
- Venkatasami, G., Sowa, J. R. (2010) A rapid, acetonitrile-free, HPLC method for determination of melamine in infant formula. *Anal. Chim. Acta*, 665, 227-230.
<https://doi.org/10.1016/j.aca.2010.03.037>

- 428 Wold S. (1976) Pattern recognition by means of disjoint principal components models, *Pattern*
429 *Recognition*, 8: 127-139.
- 430 Yang, Z., Han, L., Wang, C., Li, J., Fernández Pierna, J.A., Dardenne, P., Baeten, V. (2016).
431 Detection of Melamine in Soybean Meal Using Near-Infrared Microscopy Imaging with
432 Pure Component Spectra as the Evaluation Criteria. *J.Spectroscopy*, Volume 2016, Article
433 ID 5868170. <https://doi.org/10.1155/2016/5868170>
- 434 Yang, Z., Wang, C., Han, L., Li, J., Liu, X . (2014) Rapid screening and visual tracing of
435 melamine in soybean meal by NIR microscopy imaging. *Journal of Innovative Optical*
436 *Health Sciences*, 07(04): 1350072
437 <https://doi.org/10.1142/S1793545813500727>
- 438 Zontov, Y.V., Rodionova, O.Ye., Kucheryavskiy, S.V., Pomerantsev, A.L. (2017). DD-SIMCA
439 – A MATLAB GUI tool for data driven SIMCA approach", *Chemom. Intell. Lab. Syst.*
440 167, 23-28
441 <https://doi.org/10.1016/j.chemolab.2017.05.010>

Highlights

- Non-targeted analysis for revealing multiple adulterants is proposed
- NIR spectroscopy and data driven SIMCA used for detection fraud of soybean meal
- Contaminants of Melamine and Cyanuric acid are analyzed
- Risk-based modeling is applied
- The assessment of the limit of detection for multiple adulterants is presented

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: