Identification of botanical and geographical origin of distillers dried grains with solubles by near infrared microscopy

Noelia Tena, Ana Boix*, Christoph von Holst

European Commission, Joint Research Centre, Institute for Reference Materials and Measurements (EC-JRC-IRMM), Retieseweg 111, 2440 Geel, Belgium

1. Introduction

Complexity of food and feed production systems has increased in the last years. In particular feed materials are obtained from new sources or are produced by new technologies. Distillers dried grains with solubles (DDGS) are one of the feed materials with significant increased production rates in the recent years, mainly due to the growth of biofuel production (Cooper & Weber, 2012). DDGS are a sub-product from the production of ethanol as a fuel additive or from breweries. Like fuel ethanol, DDGS has quickly become a global product, the latter one for use in animal nutrition (Liu, 2011). During fermentation, starch from the grains are converted into fuel and carbon dioxide, which leads to an increased content of key nutrients such as proteins, fibre and fat in the DDGS (Spiehs, Whitney, & Shurson, 2002). DDGS are considered as valuable protein supplement for both ruminants and non-ruminants (Hippenstiel, Südekum, Meyer, & Flachowsky, 2012) and in addition the use of DDGS in animal nutrition also saves costs (Kalscheur, Garcia, Schingoethe, Diaz Royón, & Hippen, 2012). Since DDGS is a recognised feed material it has been included in the European Union feed catalogue (Commission Regulation (EU) No 68/2013) under number 1.12.11. The purpose of this Regulation is to provide specifications on the various feed materials marketed within the EU and according to the definition given in the catalogue, DDGS is a "Product of alcohol distilling obtained by drying solid residues of fermented grains to which pot ale syrup or evaporated spent wash has been added", where pot ale is the product remaining in the still from the first (wash) distillation of a malt distillery (Commission Regulation (EU) No 68/2013). When using DDGS in animal nutrition, precise details about the nutritional composition are required. However the nutrient content can vary due to different type and sources of grains, fermentation conditions or the climate (U.S. Grains Council, 2012a). For instance, biofuel in the USA is mainly produced from corn, whereas the EU utilises wheat and other grains (Kalscheur et al., 2012). Moreover there may be also adverse effects of the use of DDGS in animal nutrition, due to the presence of antibiotics (U.S. Grains Council, 2012b), which may even impair the safety of food of animal origin. In consequence, traceability of DDGS in terms of their botanical and geographical origin is very important for the feed sector. This demand has triggered the development of analytical authentication systems focused on the traceability such as the botanical and geographical origin of food and feed materials. In the field of feed the European project "Quality and Safety of Feeds and Food for Europe, QSAFFE" (www.qsaffe.eu) contributed to the development of such means and the specific spectroscopic method presented in this paper has been developed within this project. Authentication approaches for agricultural products, which also includes the application of

Keywords:
- DDGS
- NIRM
- Authentication
- Traceability
- Multivariate analysis

Article history:
- Received 4 November 2014
- Received in revised form 21 January 2015
- Accepted 22 January 2015
- Available online 31 January 2015

Nowadays there is a strong need for methods capable of differentiating distillers dried grains with solubles (DDGS) according to their geographical and/or botanical origin. In this study, near infrared microscopy (NIRM) has been applied to differentiate corn and wheat DDGS samples and to differentiate corn DDGS from USA, Europe and China. Two multivariate prediction models have been developed for the discriminations. The models have been validated by cross validation and with an external set of samples and have been used to predict the botanical and geographical origin. Furthermore the probability of wrong prediction of the models has been calculated, to assess their performance profile. The prediction showed very good capability of the model to classify samples according to their botanical origin, whereas satisfying results were exclusively obtained when separating samples from China against the pooled set of samples from USA and Europe.

© 2015 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
analytical based product traceability, have been performed using different techniques, e.g. stable isotope analysis (IRMS and SNIF-NMR), multi-element analysis or spectroscopic techniques such as NIR, MIR Raman spectroscopy (Garrigues & de la Guardia, 2013; González, Armenta, & de la Guardia, 2009). In particular isotope ratio mass spectrometry has been used for determination of geographical origin of DDGS (Nietner, Haughey, Ogle, Faulh-Hassek, & Elliott, 2014) and attenuated total reflection FT-IR spectroscopy has been used to authenticate the botanical and geographical origin of DDGS (Nietner, Pfister, Glomb, & Faulh-Hassek, 2013).

The aim of this paper was to study the capability of Near Infrared Microscopy (NIRM) to identify the botanical and geographical origin of DDGS. Since the hyphenation of spectroscopy and microscopy allows for the measurement of many particles of the sample, potential heterogeneity of the test material and corresponding impact on separation between the various groups could be evaluated as well. Moreover, the evaluation of the NIRM spectra obtained on test materials required the application of chemometric tools. Therefore, two partial least squares discriminant analysis (PLS-DA) models have been developed and validated in this study to predict the botanical and geographical origin of DDGS samples, respectively.

2. Material and methods

2.1. DDGS samples

DDGS samples were collected by Bundesinstitut für Risikobewertung (BfR, Berlin, Germany) in the frame of the QSAFFE project, in 2011/2013. In total 160 samples were used in this study. 128 corn DDGS samples and 32 wheat DDGS samples were collected from different countries in Europe (France, Sweden, Czech Republic, Netherlands, Spain, Belgium, Hungary, Poland and Austria), USA (NW Illinois, Wisconsin, SE Indiana), China (Jilin, Heilongjiang) and Canada (Saskatchewan). The overview of analysed samples for model creation is shown in Table 1, specifying the number of DDGS samples used in the study according to their geographical origin, i.e. China, USA and Europe.

The whole sample set used for the development and validation of the chemometric models are shown in Table 1. For the analysis of the botanical origin 160 samples were available, whereas for the analysis of the geographical origin 128 samples were used. All the samples were milled (1 mm) to homogenise the samples as much as possible and were sieved (250 μm). The fraction that was smaller than 250 μm was used to carry out the analysis to obtain the flattest mono-layer of each sample to apply the NIRM measurement.

2.2. Measurement with FTNIR-microscopy

A Fourier-Transform near Infrared spectrometer from Perkin Elmer (Perkin Elmer Spectrum One NTS system) equipped with a Perkin Elmer Spotlight microscope was used. Each sample was analysed by near infrared microscopy, using the mapping technique, in order to assess the inherent heterogeneity of the test material in the measurements. The surface of the Spectralon disk, 1 cm of diameter, was covered with a mono-layer of each sample and the visible image was captured in reflectance mode, with an aperture of 50 × 50 μm, illumination of 80% and acquisition time of 15 min. The mapping area selected per sample was 1 cm² and the numbers of analysed points were 25 × 25. In total 625 spectra per sample were obtained. The acquisition time was 63 min per sample. Each spectrum was scanned in the range of 7800–4000 cm⁻¹ with a resolution of 8 cm⁻¹ and measuring 10 scans per spectrum.

2.3. Data treatment

Prior to statistical assessment all the spectra were visually screened and those spectra showing low quality due to poor focus of the corresponding scanned region were removed. Then, the valid spectra were pre-treated by second derivatize Savitzky-Golay with polynomial order 2 and 7 smoothing points taking symmetrically.

2.3.1. Principle component analysis (PCA) of single spectra
PCA was applied on the 625 spectra of some of the samples to assess the within sample variation and to compare it with the variation observed between samples. For this purpose 20 combinations of pairs of corn and wheat DDGS samples and 20 combinations of pairs of European and Chinese corn DDGS samples were randomly taken and separately subjected to PCA. Since each sample contributed with a maximum of 625 valid spectra, the whole data set for each PCA was comprised of about 1250 spectra. The objective of this comparison was to establish whether the within sample variation was low compared to the between sample variation. If this was the case, the statistical assessment of the whole sample set could be conducted with the mean of the 625 spectra without losing relevant information. Due to the significant reduced size of the data matrix, the required time for the multivariate statistics would be significantly reduced as well. For instance, the whole data set of the mean spectra of the corn DDGS samples contained 128 rows, but when using all 625 spectra the whole data set would be comprised of 80,000 rows.

2.3.2. Partial least squares discriminant analysis (PLS-DA) of mean spectra
PLS-DA were performed to establish two models, namely one that allowed for a separation of corn from wheat DDGS samples and another one that allowed for a separation of corn DDGS samples according to their geographical origin, i.e. China, USA and Europe. Prior to statistical assessment, the mean of the valid spectra (maximum 625 spectra) for each sample was calculated and the obtained mean spectra was then subjected to the previously described pre-treatment. As shown in Table 1, for each model the total number of samples was divided into two different sets, namely (i) the calibration set consisting of 80% of the total samples.

Table 1

<table>
<thead>
<tr>
<th>Botanical origin</th>
<th>Number of samples</th>
<th>Total number of samples</th>
<th>Geographical origin</th>
<th>Number of samples</th>
<th>Total number of samples</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Calibration set</td>
<td>Validation set</td>
<td></td>
<td>Calibration set</td>
<td>Validation set</td>
</tr>
<tr>
<td>Corn DDGS</td>
<td>101</td>
<td>27</td>
<td>128</td>
<td>USA</td>
<td>54</td>
</tr>
<tr>
<td>Wheat DDGS</td>
<td>27</td>
<td>5</td>
<td>32</td>
<td>Europe</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>China</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Number of DDGS samples according with the botanical and geographical origin. Distribution of DDGS samples in two different sets used for the calibration and validation of the PLS-DA model. From the initial set of samples, 80% were used for the calibration set and 20% for the validation.
and (ii) the validation set, used for the external validation consisting of 20% of the total samples. The developed models were also cross validated by randomly dividing the calibration data set in 20 segments and leaving out 6 and 4 samples per segment each time for the botanical and geographical origin respectively. Table 1 presents the number of sample used in each set of sample (calibration and validation sets) per each model.

To develop the PLS-DA model for the botanical origin the response variable was set as 1 for corn DDGS samples and 0 for wheat DDGS samples. In the case of the PLS-DA model for the geographical origin, samples from one region were tested against the pooled samples from the remaining two regions. For instance, samples from Europa formed one group which was separated against the second group comprised of the pooled samples from USA and China. In total three different combinations were tested, namely (i) USA versus the group of China and Europe, (ii) Europe versus the group of China and USA and (iii) China versus the group of USA and Europe. In all cases the response variable was set at 1 for the single region group and 0 for the double region group.

Then a specific value between the assigned response values of 1 and 0 was selected as limit, in order to classify new samples. For instance, when classifying samples according to their botanical origin, samples with a predicted response above this limit value were considered as corn samples, whereas samples with a measured response below the limit value were considered as wheat samples. In this study the limit value for the classification of both the botanical and geographical origin was set at 0.5, which is the average value between the response values of the respective groups, i.e. 1 and 0.

In the second step the PLS-DA models were cross validated and validated with an external validation set of samples as specified in Table 1. The validation of the PLS-DA models was based on the principle that the spectra from the validation samples were applied on the developed PLS-DA models and the corresponding mean values and the standard deviation of the predicted response values were calculated. Finally, the probability of wrong classification of the validation samples was assessed by the means of t-statistics and used as indication of the efficiency of the models. The details of this concept have been described in a previous paper (Tena, Fernández Pierna, Boix, Baeten, & von Holst, 2014).

The following Eqs. (1) and (2) were used for the estimation of the probability of wrong classification according to the botanical origin, based on the measurement for cross and external validation.

\[
t - \text{value}_\text{Wheat} = \frac{0.5 - \text{Mean}_\text{Wheat}}{\text{SD}_\text{Wheat}}
\]

(1)

\[
t - \text{value}_\text{Corn} = \frac{\text{Mean}_\text{Corn} - 0.5}{\text{SD}_\text{Corn}}
\]

(2)

where Mean is the mean of the predicted response value of corn and wheat DDGS samples, SD is the corresponding standard deviation and 0.5 is the limit value, used to classify samples according to their botanical origin.

The following equation was used for the estimation of the probability of wrong discrimination according to the geographical origin, based exclusively on external validation. Eq. (3) was used for the estimation of the probability of wrong classification of samples from USA for external validation.

\[
t - \text{value}_\text{USA} = \frac{\text{Mean}_\text{USA} - 0.5}{\text{SD}_\text{USA}}
\]

(3)

where Mean is the mean of the predicted response values of the corn DDGS samples from USA, SD is the corresponding standard deviations and 0.5 is the limit value, used to classify samples according to their geographical origin. Similar equations were also used based on the measurements from corn DDGS samples from Europe and China. The probability corresponding to the calculated t value with a one-tailed distribution gives the rate of false positive results. There are not yet established acceptable maximum values for the % rate of false classifications that can be used as target values in this study. Therefore the target value was set at 5% which is already applied in other fields of food analysis, such as the maximum rate of false negative results in the area of screening methods (European Commission, 2002.)

The pre-treatments of the spectra and the multivariate statistics were carried out using the software package Unscrambler X.2 (Camo-Oslo, Norway) and the t-statistics were calculated with Microsoft Excel.

![Fig. 1. Study of variability within samples. PCA score plot of single spectra of two samples of corn DDGS from Europe and China respectively.](image-url)
3. Results and discussion

By visual inspection a small fraction of the total number of 625 spectra per sample were removed, but in all cases the absolute number of removed spectra was not higher than 30.

3.1. PCA

The results of the PCA for all 40 pairs of samples looked very similar and therefore the results of a randomly selected pair is presented in Fig. 1, showing the projection of the corresponding spectra on the first two factors. The results from the PCA of all pairs revealed that the first two factors explained more than 73% of the total variance. The visual inspection of the scores for these two factors clearly demonstrated that the individual spectra from the European and Chinese samples formed well separated groups. This means that the variability of the scores within the samples was much lower compared to the variability between the samples. Therefore we concluded that it was possible to conduct the PLS-DA statistics of the whole sample set as presented in the following chapters with the mean of the 625 spectra per sample rather than with all the 625 individual spectra.

3.2. PLS-DA model – botanical origin discrimination

The performance of the developed PLS-DA model was separately evaluated by cross validation and validation with an external data set, obtaining the following results.

![Fig. 2. Performance of PLS-DA model to discriminate the botanical origin of DDGS with cross validation. Classification of DDGS samples for calibration (blue) and for cross validation (red). The response value was 1 for corn DDGS and 0 for wheat DDGS. The limit value to separate the two botanical origins is presented as a line set at 0.5.](image)

![Fig. 3. Performance of final PLS-DA model to discriminate the botanical origin of DDGS with external validation. Classification of DDGS samples for calibration (blue) and for external validation (red). The response value assigned for corn DDGS was 1 and for wheat DDGS was 0. The limit value to separate the two botanical origins is presented as a line set at 0.5.](image)

![Fig. 4. Corn DDGS samples grouped according to their geographical origin with the projection of the two first factors of the PLS-DA model to discriminate the geographical origin using cross validation.](image)
3.2.1. Cross validation
The results of the classification of the DDGS samples according to their botanical origin by the means of PLS-DA are shown in Fig. 2. The response values for all wheat samples turned out to be below the limit value of 0.5 and the corresponding values of the corn samples were above 0.5, confirming that perfect separation of both groups was obtained. The model was optimised using the first 7 factors, explaining 98% of the total Y variance and 89% of the total X variance.

3.2.2. Validation with an external data set
The results of the validation of the PLS-DA using the external sample set as specified in Table 1 are shown in Fig. 3. The statistical assessment revealed a perfect separation between corn and wheat DDGS in both the calibration and external validation samples, since the response values of all wheat samples were below 0.5, whereas the corresponding values from the corn samples were above 0.5. The PLS-DA model applied was optimised using the first 6 factors, explaining 92% of the total Y variance and 85% of the total X variance.

Fig. 5. Performance of PLS-DA model to discriminate the geographical origin of corn DDGS with cross validation. Classification of corn DDGS samples for calibration (blue) and for cross validation (red). The response variable was: (A) 1 for corn DDGS from USA and 0 for corn DDGS from Europe and China (B) 1 for corn DDGS from Europe and 0 for corn DDGS from USA and China (C) 1 for corn DDGS from China and 0 for corn DDGS from USA and Europe. Black circles around samples indicate wrong discriminated samples. The limit value to separate the two geographical origins is presented as a line set at 0.5.
The probability of wrong classification was calculated using the previously introduced Eqs. (1) and (2). The probability of wrong classification of the PLS-DA model for the botanical origin has confirmed the high capability of the model to discriminate between corn and wheat DDGS samples. Both for cross validation and validation with an external data set, the probability of wrong classification was below 0.1% (only the probability of wrong classification of wheat as corn in the validation with an external set of samples was somewhat higher, i.e. 0.46%) and therefore considered as fit for purpose given the acceptable target level of 5%.

3.3. PLS-DA model- geographical origin discrimination

The results of the statistical assessment in the previous chapter demonstrated that the botanical origin has a significant impact on NIR spectra measured. In order to avoid any adverse effect of the botanical origin on the statistical evaluation when looking at the geographical origin of the DDGS samples, a PLS-DA model for the geographical origin has been established based on samples from the same species. Furthermore, much more DDGS samples in respect to the different geographical regions (USA, Europe and

Fig. 6. Performance of PLS-DA model to discriminate the geographical origin of corn DDGS with external validation. Classification of corn DDGS samples for calibration (blue) and for external validation (red). The response variable was: (A) 1 for corn DDGS from USA and 0 for corn DDGS from Europe and China (B) 1 for corn DDGS from Europe and 0 for corn DDGS from USA and China (C) 1 for corn DDGS from China and 0 for corn DDGS from USA and Europe. Black circles around samples indicate wrongly discriminated samples. The limit value to separate the two geographical origins is presented as a line set at 0.5.
China) were available for corn compared to wheat as shown in Table 1. Therefore a PLS-DA model to classify samples according to their geographical origin has been exclusively developed for corn. In total 128 samples have been used to develop this model. Table 1 shows the number of samples of each geographical origin in each set of samples.

The results of the statistical assessment are presented in Fig. 4, showing the projection of the two first factors of the PLS-DA model. This figure reveals that factor 1 of the PLS-DA allowed for a vertical discrimination between corn DDGS from China against the group composed of the samples Europe and the USA. However, there was an overlapping between samples from Europe and USA when looking at the horizontal separation of factor 2. The PLS-DA model with cross validation was optimised using the first 5 factors, explaining 80% of the total Y variance and 90% of the total X variance.

The performance of the developed PLS-DA models are now presented, separately when applying cross validation and validation by an external data set.

3.3.1. Cross validation

The classification of corn DDGS samples according to the geographical origin is presented in Fig. 5 for the calibration and cross validation phase of the PLS-DA model, respectively. The results confirmed a perfect separation between samples from China and the pooled samples from USA and Europe (Fig. 5C) but wrong classification for some samples were observed when separating (i) samples of USA against samples of Europe and China (Fig. 5A) or (ii) samples of USA against samples of USA and China (Fig. 5B). Looking more carefully at the wrongly classified samples revealed that they are from Europe, namely five in Fig. 5A and six in Fig. 5B. Moreover, four of these samples are common in both figures. Only two samples from the USA were wrongly classified, one in each of these figures. In contrast, the PLS-DA model was working perfectly when separating the samples from China against the pooled samples from USA and Europe, since all samples were correctly classified.

3.3.2. Validation by an external data set

The samples used for this validation were the samples included in the validation set, in Table 1. The same criteria as explained above, was applied to classify into the three different geographical origins and the limit at 0.5 was used to discriminate between the groups in each case. The model was optimised using the first 2 factors, explaining 63% of the total Y variance and 77% of the total X variance. Fig. 6 presents the classification of corn DDGS samples into the three geographical origins in the calibration and external validation phase of the PLS-DA model. As expected, the discrimination between samples was better when applying cross validation compared to the validation with an external sample set.

In general, better discrimination was observed for the PLS-DA model of the samples from China against the pooled samples from Europe and USA (Fig. 6C) compared to the other two combinations (Fig. 6A and B), thus confirming the results obtained with cross validation. In detail, the results from Fig. 6C showed that the response value from one sample was slightly above 0.5 and would therefore be not classified as belonging to the joint group of Europe and USA. However, the response value of this sample was also far away from corresponding response values of the other samples, which were close to 1. Therefore the sample would be classified neither from Europe/USA nor from China and therefore classification of this specific sample would be impossible with the developed PLS-DA model.

In order to study the capability of the model to classify new samples, the probability of wrong discrimination of the model was studied with the response values obtained from the external validation samples, thus applying the strictest conditions possible. For the calculation of the probability of wrong classification of the external samples, Eq. (3) was applied. Whilst an unacceptable high value of 16.8% for the probability of wrong classification was observed for the samples from Europe against the pooled samples from China and USA, significant better results were obtained for the discrimination of samples from the USA against the pooled samples from Europe and China. However, the obtained value for the probability of wrong classification was 6.6% and therefore still above the target level of 5%. Only for the discrimination of the samples from China against Europe and the USA a satisfying value below 1% was obtained.

4. Conclusions

Two PLS-DA models have been developed and validated, namely one to discriminate between corn and wheat DDGS samples and another to discriminate corn DDGS from USA, Europe and China. The evaluation of the results of external validation demonstrated that the model for the botanical origin was perfectly working and very robust compared with the model to classify samples against their geographical origin. This is in line with the results obtained by Nieter et al. 2013 applying attenuated total reflection FT-IR spectroscopy.

Amongst the various classifications in respect to the geographical origin, excellent results were only obtained for the discrimination of samples from China against the pooled samples from Europe and USA. In contrast acceptable classification performance has not been obtained for the discrimination of (i) samples from Europe against the samples from USA and China and (ii) of samples USA against samples from China and Europe. Given the fact that near infrared microscopy is a non-destructive analytical method combined with the short time required to conduct the analysis, the present study confirmed its applicability in the area of traceability of DDGS samples. This especially applies to classification against the botanical origin.

Acknowledgements

The research presented in this paper has received funding from the European Commission’s Seventh Framework Programme (FP7/2007–2013) under grant agreement n° 265702 (QSAFE-project).

The authors wish to acknowledge the help of Nanta S.A. Madrid, and Servizo de Seguridade Alimentaria nas Productions Ganaderas (Xunta de Galicia) for providing samples with different botanical origins.

References


