Received: 10 July 2017 | Revised: 20 October 2017 | Accepted: 1 November 2017

DOI: 10.1002/cem.2982

SPECIAL ISSUE ARTICLE

WILEY Journal of CHEMOMETRICS

# NIR hyperspectral imaging spectroscopy and chemometrics for the discrimination of roots and crop residues extracted from soil samples

Damien Eylenbosch[1] | Bernard Bodson[1] | Vincent Baeten[2] | Juan Antonio Fernández Pierna[2]

[1]Gembloux Agro-Bio Tech, AgroBioChem Department, University of Liège, Crop Science Unit, 2 Passage des Déportés, 5030 Gembloux, Belgium

[2]Valorisation of Agricultural Products Department, Food and Feed Quality Unit, Walloon Agricultural Research Centre, Henseval Building, 24 Chaussée de Namur, 5030 Gembloux, Belgium

**Correspondence**
Juan Antonio Fernández Pierna, Valorisation of Agricultural Products Department, Food and Feed Quality Unit, Walloon Agricultural Research Centre, Henseval Building, 24 Chaussée de Namur, 5030 Gembloux, Belgium.
Email: j.fernandez@cra.wallonie.be

**Abstract**

Roots play a major role in plant development. Their study in field conditions is important to identify suitable soil management practices for sustainable crop productions. Soil coring, which is a common method in root production measurement, is limited in sampling frequency due to the hand-sorting step. This step, needed to sort roots from other elements extracted from soil cores like crop residues, is time consuming, tedious, and vulnerable to operator ability and subjectivity. To get rid of the cumbersome hand-sorting step, avoid confusion between these elements, and reduce the time needed to quantify roots, a new procedure, based on near-infrared hyperspectral imaging spectroscopy and chemometrics, has been proposed. It was tested to discriminate roots of winter wheat (*Triticum aestivum* L.) from crop residues and soil particles. Two algorithms (support vector machine and partial least squares discriminant analysis) have been compared for discrimination analysis. Models constructed with both algorithms allowed the discrimination of roots from other elements, but the best results were reached with models based on support vector machine. The ways to validate models, with selected spectra or with hyperspectral images, provided different kinds of information but were complementary. This new procedure of root discrimination is a first step before root quantification in soil samples with near-infrared hyperspectral imaging. The results indicate that the methodology could be an interesting tool to improve the understanding of the effect of tillage or fertilization, for example, on root system development.

**KEYWORDS**

classification, NIR hyperspectral imaging, PLS-DA, SVM, wheat root

## 1 | INTRODUCTION

Roots play a major role in plant development. They are the link between belowground resources and aboveground growth[1] and therefore have a great influence on plant biomass production.[2] Root morphology and physiology determine the capacity for nutrient uptake and water extraction by plants.[3] Their growth is determined by both the genetic traits of plants and the physical and chemical properties of soil and shows high plasticity under different environmental

  

conditions. The potential of plants to obtain water and mineral nutrients from the soil is primarily attributed to their capacity to develop extensive root systems.[2]

Studies on root development are important for a better understanding of the interaction between crop root systems and the growing environment to identify suitable soil management practices for sustainable crop productions.[4] The time taken by roots to colonize the deep soil profile and the root system biomass production during the whole crop development cycle are therefore good indicators to interpret crop behavior in various management modes and in a given soil-climate context. Nevertheless, roots are rarely directly evaluated in field experiments because soil limits accessibility for their observation.[1,5] Therefore, they need to be extracted from soil before any measurement can be made.

Soil coring is a commonly used method to sample roots in field experiments and measure their production.[6] This technique is not expensive and allows repeated measurements during crop growth at several soil depths. After soil coring, cores need to be washed to extract roots from soil. Others elements, mainly crop residues buried in crop soils, are also extracted during the washing step and need to be separated from roots, commonly by hand-sorting, before any analysis or quantification of roots.[1,4,5,7-10] This sorting step is time-consuming, tedious, and vulnerable to operator ability and subjectivity.[1] It is therefore the most significant limitation of the soil coring method because it limits the frequency of sampling during crop growth.[11]

Near-infrared spectroscopy (NIRS) is used to identify and quantify components in agricultural products and is characterized by the acquisition of a distinctive spectral chemical profile, which can be considered as a fingerprint of the material. This method allows the acquisition of chemical information from samples with numerous advantages like ease of use and the possibility to quickly and simultaneously analyze several components without any reagents and without sample destruction. It is used to classify or even identify any given material without dependency on human subjectivity. Near-infrared spectroscopy being an indirect method of measurement needs to be coupled with chemometrics, a chemical discipline using mathematics and statistics to extract the relevant chemical information out of measured data.[12-17]

In the last decades, NIRS has been applied to study root systems and to detect differences in the chemical composition of roots. It was used to estimate root biomass proportions in different root mixtures,[18,19] to predict the percentage of dead versus living grass roots[7] and to determine the taxa of herbaceous and woody species.[20] Near-infrared spectroscopy was also used as a rapid method to predict pasture and maize root densities in soil cores, directly in the field, without separating soil and roots.[21,22]

Recently, new advances have permitted the coupling of NIR spectrometers with imaging technologies, thus creating new devices (NIR hyperspectral imaging), providing spectral and spatial information simultaneously.[16,23] With this method, thousands of spectra can be obtained for each sample, giving a complete picture of the distribution of chemical compounds at pixel level.[14,24,25]

The aim of this work was to develop a new procedure based on NIR hyperspectral imaging coupled with chemometrics, allowing a faster discrimination of elements extracted from soil cores to decrease the time needed to analyze samples and thereby allowing an increase in the number of soil samples that can be taken in a same time and space. An increase of sampling throughout crop development will allow a better understanding of the root system's development. An increase of sampling in a same field will improve the knowledge of the spatial heterogeneity existing in field conditions and of the root system's development at several depths. Discrimination of roots and crop residues could lead to a better understanding of root system development in field conditions and its adaptation to its growth environment.

## 2 | MATERIALS AND METHOD

Root and crop residue samples were collected by soil coring in a long-term trial on tillage and crop residue management. The experimental field was located on the experimental farm of Gembloux Agro-Bio Tech (University of Liège, Gembloux, Belgium). A complete description of the site was given by Degrune et al.[26] The trail was sown with winter wheat (*Triticum aestivum* L.) in 2010, 2011, and 2012. Soil samples were collected during the second and third years of this crop on several dates (November 2011 and April and May 2012), corresponding to different development stages of the crop (first leaf development, tillering, and last leaf deployment) or after the harvest (September 2013). After appropriate weeding of the field, all sampled roots and crop residues were considered as produced by the winter wheat crop. We considered as crop residues the above-ground parts of plants remaining in the field after harvest, mainly stubble, chopped straw, and husks. These crop residues were mixed with the soil by plowing (25-cm depth tillage) or reduced tillage (10-cm depth tillage). Soil samples were taken at a depth of 30 cm by using a 30-mm diameter soil-coring tube

driven into the soil by a tractor-mounted hydraulic push press. They were then divided into 3 soil horizons (0-10, 10-20, and 20-30 cm deep).

As in the common hand-sorting method, soil samples were washed with tap water to extract roots and crop residues. These elements were collected on a sieve (Figure 1). Small stones and soil aggregates were also present. The sieve containing all the elements was dried at 60°C until a constant weight was reached. After drying, extracted elements were sorted by hand into 3 classes to calibrate and validate the discrimination models: roots, crop residues, and soil (small stones and soil aggregates). After model development, this hand-sorting step is no longer necessary. Samples were then kept at ambient temperature and humidity until the acquisition of NIR hyperspectral images.

Near-infrared hyperspectral images were acquired with a moving imager technique combining an NIR hyperspectral line scan, or push-broom imaging system, and a conveyor belt (BurgerMetrics SIA, Riga, Latvia) (Figure 2). The camera was an SWIR XEVA CL 2.5 320 TE4 camera (SPECIM Ltd, Oulu, Finland) using an ImSpector N25E spectrograph that includes a cooled, temperature-stabilized mercury-cadmium-telluride detector (XENICS nv, Leuven, Belgium). Images were acquired in the 1118 to 2424 nm range with a 6.3-nm spectral resolution (ie, 209 wavelengths) with a width of 320 pixels. For each pixel, 32 scans were coadded and the mean absorbance spectrum was saved. Illumination was achieved with 2 halogen lamps of 120 W each. The camera position was set up to give images 10 cm in width. Pixel resolution was of 0.31 mm width. The conveyor belt speed was adapted to ensure that the pixels were always square-shaped. Acquisitions were performed with HyperPro VB software (BurgerMetrics SIA, Riga, Latvia). The system was first calibrated with white ceramic (white reference) and by blocking the entrance of reflected light (dark reference). A dark reference was automatically performed before acquiring each image. A white reference was performed several times during acquisition day. Bad pixels were detected and removed at the same time.

Samples were placed on the conveyor belt, and care was taken so that the overlapping of constituents was reduced. For the calibration and the validation of our models, soil, crop residues, and roots were sorted manually and images were

**FIGURE 1** Picture of the sieve with the different elements extracted from soil cores

**FIGURE 2** Near-infrared (NIR) hyperspectral imaging system installed at the Walloon Agricultural Research Center, Belgium (BurgerMetrics SIA, Riga, Latvia)

acquired separately on each class of elements. Only perfectly washed samples of roots and crop residues were used so that soil was not present on these samples. During manual sorting, only perfectly identified elements were selected for model calibrations and validations. For crop residues, spectra from the inner and outer sides of pieces of straw were acquired. Preliminary tests on spectra showed that spectral profiles of these 2 sides were different. It was therefore important to acquire both sides to take all the spectral variability into account for model calibrations. Images of the sieve were also acquired because, in the case of discrimination of the elements extracted together from soil cores, all elements stay on the sieve during the acquisition of NIR spectra to facilitate sample handling. Near-infrared spectra of the sieve were therefore needed to calibrate discrimination models. To build the sieve library, a piece of sieve has been placed alone on the conveyor belt, in the same way than the other elements used for model calibrations and validations.

Near-infrared spectra acquired on sample surface were extracted with HyperSee software (BurgerMetrics SIA, Riga, Latvia) from 24 NIR hyperspectral images of roots, 56 images of crop residues, 5 images of soil, and 5 images of the sieve. One matrix of spectra was created for each image. These images were chosen to cover the highest possible spectral variability: images of roots and crop residues were acquired on samples collected at different times throughout crop growth and at different soil depths. Moreover, variability due to the system of acquisition was added in the calibration set by acquiring spectra from different samples at different dates and by collecting spectra from the same sample of crop residues used as a reference and acquired on 13 different dates. It was also important to take edge and shading effects into account: spectra were acquired on the whole sample surface, and spectra from shadow areas created on the conveyor belt by roots and crop residues were extracted and saved separately. These shadow spectra did indeed have a specific spectral profile and were extracted from 14 NIR hyperspectral images. They were considered as background together with the sieve spectra.

Spectra with spikes were removed from all images. These spikes are related to dead pixels that are due to a few elements in the detector responding with an unusually high or low value.[27] All spectra with absorbance higher than 1.5 at 1740 nm, corresponding to conveyor belt absorbance, were also removed. Due to noise at the beginning and the end of spectra, only the 150 wavelengths between 1432 and 2368 nm were maintained. All this preprocessing was done on Matlab R2015a software (The MathWorks, Inc, Natick, MA, USA).

After the removal of spikes and conveyor belt spectra, data extracted from each image still contained between several hundreds to several thousands of spectra. Such a large amount of spectra was useful to take the large sample spectral variability into account for the calibration of models, but it contained redundancy and it increased the time needed to calibrate models. To keep the spectral variability and reduce redundancy, the duplex algorithm described by Snee was used to select representative spectra for the calibration set.[28] Images were gathered in groups (according to sample nature and acquisition date), and around 1000 spectra from each were selected. The duplex algorithm was applied to each image separately to select homogeneously spectra to build a spectral library. This method starts by selecting the 2 points furthest from each other and puts them both in a first set (training). Then the next 2 points furthest from each other are put in a second set (test), and the procedure is continued by alternatively placing pairs of points in the first or second set.[29] As a result, data are divided into 2 groups, which include the most diverse spectra, cover approximately the same spectral variability, and have similar statistical properties. In total, around 9000 spectra were selected in a first test set that was used for the calibration of the discrimination models. A second selection was done with the duplex algorithm on the remaining spectra to create a second test set, the spectral validation set, which also contained around 9000 spectra.

To normalize spectra, an autoscale transformation was applied on selected spectra. Preliminary tests showed that this preprocess improved discrimination of our spectral classes and gave best results among classical data pretreatments. Autoscale is a common preprocessing method that uses mean-centering followed by the division of each variable by its standard deviation.[30] As a result, all variables of the dataset have a mean of 0 and a standard deviation of 1.

In this work, the 2 discrimination algorithms were the following supervised methods: partial least squares discriminant analysis (PLS-DA) and support vector machine (SVM). Partial least squares discriminant analysis is a dimensionality reduction technique that defines new latent variables such that they explain maximum variation of independent and dependent variables.[31] Support vector machine is an algorithm designed to derive a function based on high-dimensional space that describes the hyperplane that optimally separates 2 classes of data. Support vector machine tries to correctly classify the training data by maximizing the wide of the margin between the groups and by penalizing for errors to get a better generalization performance. This is done in a high dimensional space where the groups are linearly separable via a kernel function.[15,32,33]

Preprocessing and model calibrations were done by using the PLS Toolbox 7.8 software (Eigenvector Research, Inc, Wenatchee, WA, USA). Discrimination models were arranged in a dichotomous classification tree by using Matlab R2015a software (The MathWorks, Inc, Natick, MA, USA). This classification method was chosen for its good performance in a previous study.[15]

## 3 | RESULTS AND DISCUSSION

Spectra of background, soil, crop residues, and roots were discriminated, thanks to a dichotomous classification tree containing 3 nodes: each node was based on individual discrimination models by successively extracting background (sieve and shadows) and soil before discriminating crop residues from roots (Figure 3). The aim of each model was to separate spectra of interest (considered as positive class) from the rest (negative class).

Parameter optimization was performed separately for each model. For PLS-DA models, the number of latent variables was chosen to reduce the cross-validation classification error average. A 10-group cross-validation was performed. For SVM models, the cost (penalizing parameter) and the gamma (radial basis function kernel parameter) were optimized by using the grid search procedure to minimize the misclassification fraction.

Table 1 presents the results obtained with both algorithms during model calibration, cross-validation, and validation. Calibration and cross-validation were obtained with spectra of the calibration set. For the validation, models were tested with spectra of the validation set. These results were expressed in sensitivity and specificity. Sensitivity is the proportion of spectra that are actually positive and classified as positive for one class, whereas specificity is the proportion of spectra that are actually negative and classified as negative for one class.[15] For instance, for the SVM model concerning the discrimination of background versus the rest, a sensitivity of 97.1% and a specificity of 99.8% were obtained, indicating that 97.1% of the spectra from background were correctly classified as background and 99.8% of the other spectra were correctly classified as spectra belonging to the rest (ie, spectra of soil, crop residues, and roots altogether).

All models allowed discrimination of the spectral classes. The best results were obtained with the soil versus rest model: in general, up to 99% of soil spectra were correctly classified. The background versus rest model also provided good discrimination with up to 95% of correctly classified spectra in general. For both models, the specificity was higher with SVM than with PLS-DA. The discrimination between crop residues and roots was more difficult: 14% of root spectra used to validate models were classified as crop residues with SVM. For this third model, crop residues were always better predicted with SVM.

The validation of the 3 discrimination models applied separately on the spectra of the validation set confirmed the results obtained during the calibration of models and the ability of models to discriminate spectra from the different spectral classes. It gave a raw idea of model performance when applied to new spectra and was therefore very useful

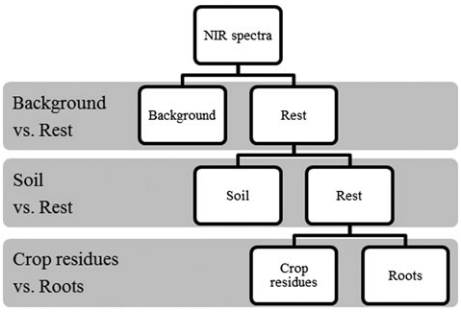


**FIGURE 3** Dichotomous classification tree. Three successive discrimination models allow the separation of near-infrared (NIR) spectra in 4 classes: background, soil, crop residues, and roots

**TABLE 1** Results obtained during calibration, cross-validation, and validation steps of the discrimination models constructed respectively with support vector machine (SVM) or partial least squares discriminant analysis (PLS-DA)

| | | Calibration | | Cross-Validation | | Validation | |
|---|---|---|---|---|---|---|---|
| | | Sensitivity | Specificity | Sensitivity | Specificity | Sensitivity | Specificity |
| Background versus rest | SVM | 97.1% | 99.8% | 95.3% | 99.3% | 94.9% | 95.7% |
| | PLS-DA | 96.2% | 95.9% | 96.2% | 95.8% | 96.2% | 91.4% |
| Soil versus rest | SVM | 99.2% | 99.7% | 99.2% | 99.5% | 99.2% | 99.7% |
| | PLS-DA | 99.3% | 98.2% | 99.3% | 98.2% | 98.8% | 96.6% |
| Crop residues versus root | SVM | 97.1% | 92.5% | 94.0% | 85.2% | 94.1% | 85.9% |
| | PLS-DA | 88.0% | 86.5% | 87.7% | 86.0% | 88.5% | 89.5% |

Results are expressed in model sensitivity and specificity. Validation results were obtained on other spectra than those used for calibration and cross-validation.

during model construction. However, it was observed that this validation of the model sometimes gave an inaccurate picture of the model's quality. In some cases, excellent results were obtained during model construction and validation, but bad predictions were obtained when these models were used to predict the spectral nature of pixels on hyperspectral images. Moreover, validation done on each discriminant model separately did not take into account the sequence of models applied with the classification tree. The discrimination models were therefore also validated in a second way. Table 2 gives results obtained when the whole dichotomist classification trees, constructed respectively with SVM and PLS-DA algorithms, were applied to 264 new NIR hyperspectral images to predict the spectral nature of each pixel constituting these images. These images were acquired on samples containing only elements corresponding to 1 spectral class (sieve, soil, crop residues, or roots). This second table presents the number of images used for the 4 spectral classes, the number of pixels predicted for each class (in total and after removal of spectra predicted as background), and the percentage of pixels that were or were not correctly predicted in each group. These results do not take into account pixels having a high absorbance value that was removed during image pretreatment and corresponding to conveyor belt. The results show that background was correctly predicted on images containing only the sieve (no elements extracted from soil cores): less than 1% of pixels were detected as belonging to another spectral class with both algorithms and the SVM tree correctly predicting 99.9% of background pixels. The sieve was not present on the other images, and only shadow areas were thus considered as background. The amount of pixels predicted as shadow reached one to two-thirds of pixels analyzed by models. These pixels were present on the conveyor belt, along the edge of elements. The highest part of pixels predicted as shadow was reached on images of roots. These elements were thin, and the ratio between shadow along the edge of roots and their surface was therefore higher.

Images of soil were also well predicted. The SVM tree correctly predicted 99.3% of pixels, and the PLS-DA tree correctly predicted 97.5% of pixels, with only slight confusion appearing with root spectra.

The weakest results were obtained on images of roots and crop residues with both classification trees. With the SVM tree, 17.4% of pixels of crop residues were predicted as roots and 8.3% of root pixels were predicted as crop residues. With the PLS-DA tree, 16% of crop residue pixels were predicted as roots and part of the root pixels were predicted as crop residues (7.3%) or as soil (5%). These percentages did not take into account the number of pixels predicted as background by the first model of discrimination in the classification tree. Considering the results in absolute values, the number of pixels correctly classified as background, soil, crop residues, or roots on the images of these 4 classes was always higher with the SVM tree than with the PLS-DA tree. The SVM tree correctly classified a higher proportion of pixels as background on images of the sieve and classified a lower proportion of pixels in this class on images of soil, crop residues, and roots. The PLS-DA tree classified a higher proportion of pixels as roots on soil images and a higher proportion as soil on root images. The highest difference between SVM and PLS-DA trees was observed with images of roots: the SVM tree predicted 19% more pixels as roots.

This method of validating models was much more time-consuming in prediction and analysis than the first way of validation, but it gave interesting results: it was possible to validate models on a very large number of spectra, therefore with a higher variability, and to quantify the proportion of spectra wrongly predicted in each spectral class separately. Furthermore, it was possible to identify regions of hyperspectral images where pixels were not well classified. Indeed,

**TABLE 2** Number and percentage of pixels predicted as background (sieve and shadow areas), soil, crop residues, or roots on near-infrared (NIR) hyperspectral images acquired on samples containing only elements corresponding to 1 spectral class

| Classification Tree | SVM | | | | PLS-DA | | | |
|---|---|---|---|---|---|---|---|---|
| **Samples** | **Sieve** | **Soil** | **Residues** | **Roots** | **Sieve** | **Soil** | **Residues** | **Roots** |
| Number of images | 4 | 12 | 131 | 117 | 4 | 12 | 131 | 117 |
| Total number of pixels | 175,583 | 323,407 | 4,880,123 | 2,525,402 | 175,583 | 323,407 | 4,880,123 | 2,525,402 |
| Pixels classified as Background (%) | 99.9% | 35.2% | 49.6% | 56.2% | 99.1% | 40.6% | 51.9% | 63.2% |
| Remaining pixels after background removal | | | | | | | | |
| Pixels classified as Soil, residues or roots | | 209,572 | 2,460,102 | 1,107,303 | | 192,168 | 2,345,244 | 928,890 |
| Soil (%) | | 99.3% | 1.8% | 0.8% | | 97.5% | 2.2% | 5.0% |
| Crop residues (%) | | 0.2% | 80.8% | 8.3% | | 0.0% | 81.7% | 7.3% |
| Roots (%) | | 0.5% | 17.4% | 90.8% | | 2.5% | 16.0% | 87.7% |

Number of pixels and percentages were calculated by cumulating, for each spectral class, the results obtained on all hyperspectral images analyzed for this class.
PLS-DA indicates partial least squares discriminant analysis; SVM, support vector machine.

based on prediction results for each pixel, images were reconstructed by assigning 1 color to each spectral class. The visual analysis of reconstructed images reveals how the different image regions were predicted and allows the identification of areas where spectra were wrongly classified. On images of crop residues, pixels located on the border of crop residues were often classified as roots (Figure 4A). This mistake was lower when using SVM. In contrast, the SVM tree applied to images of the inner side of crop residues classified a higher number of pixels as roots in the central part of crop residues than PLS-DA (Figure 4B). However, the analysis of pixel number predicted for each spectral class showed that, on 92.4% of crop residue images, the number of pixels correctly classified as crop residues was higher with the SVM tree. Dead pixels were sometimes present on images. These pixels were generally better classified with the SVM tree (Figure 4C). On images of roots, pixels classified as crop residues were mostly located in the central area of large roots. On these images, a large part of root pixels were classified as background (Figure 5). Considering all images of roots predicted with both models, a higher pixel number was classified as roots with the SVM tree on 99.1% of root images.

This confusion of roots and crop residues spectra was already observed during the calibration and the first validation of models. The analysis of predicted hyperspectral images allowed to highlight that this confusion appeared mainly on the border of crop residues. A large part of the spectra on the border of elements resulted from the chemical nature of several elements: the conveyor belt and elements extracted from soil cores or, in the case of samples with mixed

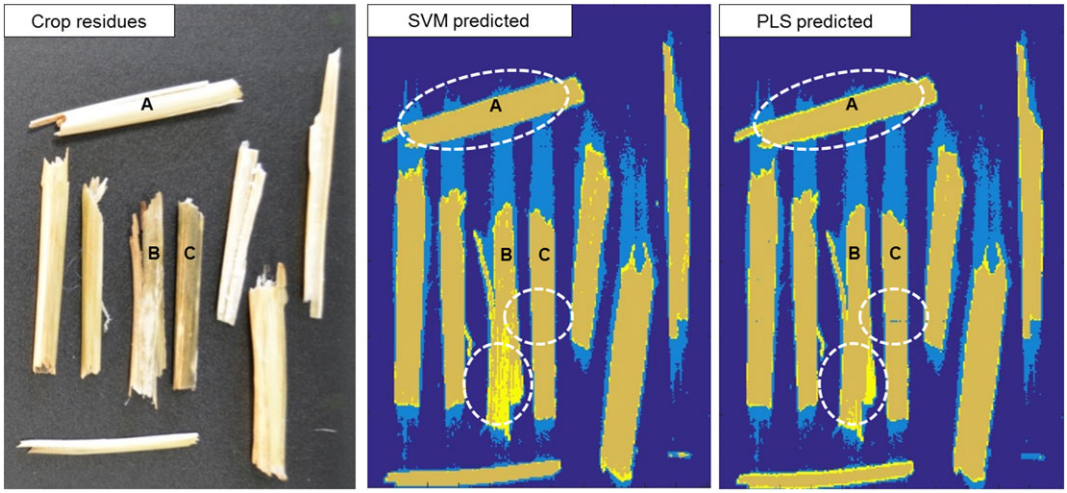


**FIGURE 4** Picture of crop residues and predictions of near-infrared (NIR) hyperspectral images of crop residues with SVM and PLS-DA dichotomous classification trees. Pixels classified as crop residues are in ochre. Pixels classified as roots are in yellow. Differences between predictions with support vector machine (SVM) and partial least squares discriminant analysis (PLS-DA) are illustrated in dashed circles and can be observed (A) on the border of crop residues, (B) on the inner face of straw, and (C) in prediction of dead pixels

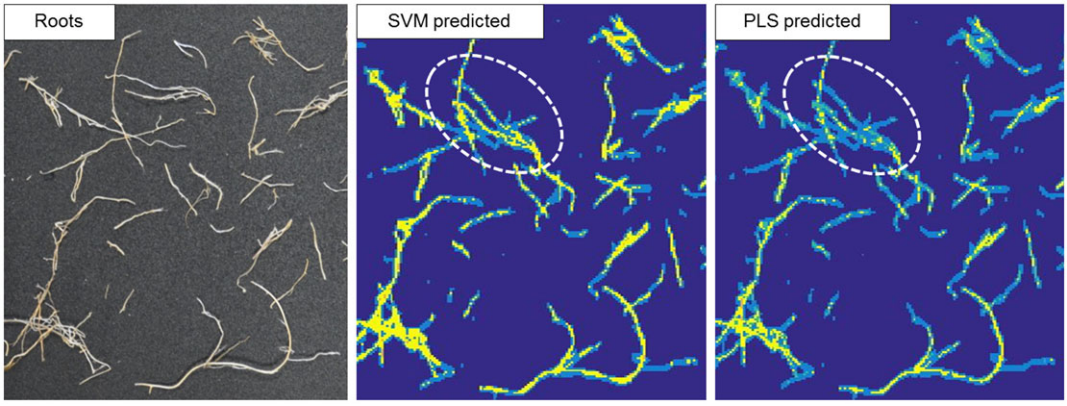


**FIGURE 5** Picture of roots and predictions of near-infrared (NIR) hyperspectral images of roots with support vector machine (SVM) and partial least squares discriminant analysis (PLS-DA) dichotomous classification trees. The main differences between predictions with SVM and PLS-DA were a higher number of pixels classified as roots (in yellow) and a lower number of pixels classified as background (in royal blue) as illustrated in the dashed circle

elements, different elements extracted from soil cores that overlapped on the conveyor belt during image acquisition. Vermeulen et al reported that spectra in a pixel are often a mixture of the pure constituent spectra.[34] Spectra acquired in the central part of crop residues were not affected by this phenomenon because crop residues are quite large elements and it was easy to acquire pure spectra on it. This confusion due to the low spatial resolution of the instrument and the mix of spectra was also well observed on predicted images of roots. The roots used in this study were sampled under a winter wheat crop, and they were very thin: their diameter was mostly smaller than the pixel size. Much of the root spectra were therefore the result of mixed spectra of root and background resulting in confusion between these 2 spectral classes. Preliminary tests showed that model results were improved when these border areas were taken into account to construct models. Other elements of confusion, located in the central part of roots or crop residues, seem to be due to the shape of the elements and their interaction with light. Such edge and shaping effects were also observed and discussed by Manley et al.[35] All spectral confusions were therefore most likely linked to the way spectra were acquired. Figure 6 shows how similar were the spectra of root and crop residues used to calibrate the discriminant models. Furthermore, it seems that these confusions were not due to the chemical composition of roots and crop residues because chemical analysis demonstrated that their C/N content was different. In addition, they were perfectly discriminated when samples were ground (results not shown). Soon and Arshad also measured a higher C/N ratio in the straw of winter wheat than in their roots,[36] and Kamruzzaman et al observed that minced samples were much better classified, probably due to the homogeneous surface.[37]

The comparison of models has indicated that models built with SVM better discriminated root and crop residue spectra both during model calibration and during prediction on hyperspectral images. These results confirmed the good performances of SVM previously observed and its superiority to solve complex problems.[15,16,32] Spectral classes were better separated, and areas of confusion like crop residue borders, thin roots, or bad pixels were better predicted. It was observed on predicted images, and it could explain the difference observed in Table 1 between the selectivity and the sensitivity of the SVM model discriminating crop residues and roots. Applied to the sample

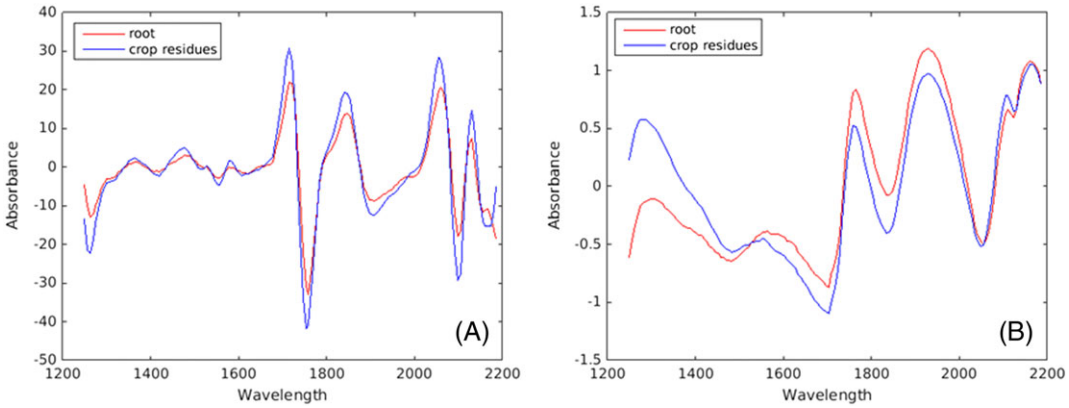


**FIGURE 6** Mean spectra of roots and crop residues of winter wheat after A, second derivative or B, Standard Normal Variate (SNV) transformation. These mean spectra were calculated on the spectra used to calibrate the models. They were very similar, and the discrimination of root and crop residues was therefore challenging

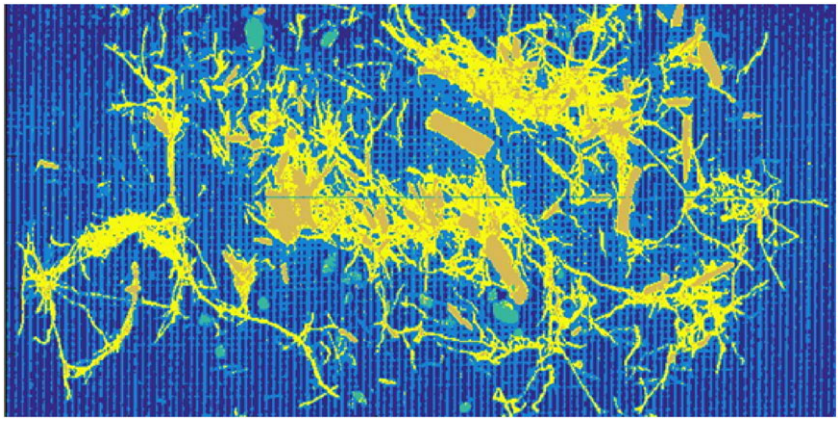

**FIGURE 7** Support vector machine (SVM) prediction on an near-infrared (NIR) hyperspectral image acquired on a sample containing wheat roots (in yellow), crop residues (in ochre), small stones and clots (in turquoise), and the sieve (in royal blue). The conveyor belt is in dark blue

presented in Figure 1, SVM gave a good prediction of all the elements extracted from a soil core (Figure 7). The only limitation identified with SVM was the time needed to construct models, which was longer than with PLS-DA. Due to the large quantity of data, several hours were necessary to calibrate the SVM models, whereas only a few seconds were needed for PLS-DA. However, once the models are built, the performance of SVM is larger in sensitivity and robustness.

## 4 | CONCLUSION

This work demonstrated the ability of NIR hyperspectral imaging spectroscopy and chemometrics to discriminate winter wheat roots, crop residues, and soil particles extracted from soil cores. To reach the best discrimination, it was important to identify all sources of spectral variability on hyperspectral images. This variability can be due to the nature of analyzed elements, presence of shadow areas on images, or variability induced by the acquisition system itself. In routine use, predictions would be improved if discrimination models were regularly updated with new spectra to take into account as much spectral variability as possible. The interest of validating discrimination models on NIR hyperspectral images was demonstrated. It allows the validation of models when they are combined in a dichotomous classification tree and the identification of areas of spectral confusion. This method of validation proved to be very useful to estimate a model's ability to discriminate elements extracted from soil cores.

The comparison of chemometric tools confirmed the good performances of SVM. Models built with this algorithm classified a higher proportion of spectra in the correct spectral class and were less sensitive in areas of spectral confusion than models built with PLS-DA.

Although discrimination between the spectra of roots and crop residues was not perfect, this new procedure, which is rapid and does not depend on operator subjectivity, offers new possibilities in studies on root systems. Near-infrared hyperspectral imaging and chemometrics could be used to quantify roots extracted from soil cores sampled in field conditions to replace the time-consuming and tedious manual step currently required with the soil coring method after washing soil samples. Furthermore, this new procedure could also be of use to quantify crop residues in soils and study their degradation dynamics, an important step in soil carbon fixation. This root or crop residue quantification could be based on the link between the weight of roots or crop residues in a sample analyzed with the hyperspectral camera and the number of pixels predicted as belonging to these spectral classes by the discriminant models. Coupled with the soil coring method, this new procedure could be used to study tillage and nitrogen fertilization effects on winter wheat root system development in field conditions.

## ACKNOWLEDGEMENTS

The authors would like to thank Françoise Vancutsem and Jérôme Pierreux from the Crop Science Unit (GxABT) for supplying samples of roots and crop residues, as well as Remi Delcroix and Loïc Dewaele for their help in collecting these samples. They would also like to thank Guillaume Fraipont from the Crop Science Unit (GxABT) and Nicaise Kayoka Mukendi from the Valorisation of Agricultural Products Department (CRA-W) who helped with image acquisitions, as well as Frederic Collinet and the CAMI technical platform at the AgroBioChem Department (GxABT) for access to the computational server.

## ORCID

*Damien Eylenbosch* http://orcid.org/0000-0002-7294-2748
*Bernard Bodson* http://orcid.org/0000-0002-2968-8086
*Vincent Baeten* http://orcid.org/0000-0003-4342-333X
*Juan Antonio Fernández Pierna* http://orcid.org/0000-0003-4953-8402

## REFERENCES

1. Plaza-Bonilla D, Álvaro-Fuentes J, Hansen NC, Lampurlanés J, Cantero-Martínez C. Winter cereal root growth and aboveground-belowground biomass ratios as affected by site and tillage system in dryland Mediterranean conditions. *Plant and Soil*. 2014;374(12):925-939. https://doi.org/10.1007/s11104-013-1926-3

2. Guan D, Al-Kaisi MM, Zhang Y, et al. Tillage practices affect biomass and grain yield through regulating root growth, root-bleeding sap and nutrients uptake in summer maize. *Field Crop Res*. 2014;157:89-97. https://doi.org/10.1016/j.fcr.2013.12.015

3. Fan J, McConkey B, Wang H, Janzen H. Root distribution by depth for temperate agricultural crops. *Field Crop Res*. 2016;189:68-74. https://doi.org/10.1016/j.fcr.2016.02.013
4. Huang G, Chai Q, Feng F, Yu A. Effects of different tillage systems on soil properties, root growth, grain yield, and water use efficiency of winter wheat (*Triticum aestivum* L.) in arid northwest China. *J Integr Agric*. 2012;11(8):1286-1296. https://doi.org/10.1016/S2095-3119(12)60125-7
5. Cheng W, Coleman DC, Box JE. Root dynamics, production and distribution in agrosystems on the Georgia Piedmont using minirhizotrons. *J Appl Ecol*. 1990;27(2):592-604. https://doi.org/10.2307/2404304
6. Jose S, Gillespie AR, Seifert JR, Pope PE. Comparison of minirhizotron and soil core methods for quantifying root biomass in a temperate alley cropping system. *Agr Syst*. 2001;52(2):161-168. https://doi.org/10.1023/A:1010667921970
7. Picon-Cochard C, Pilon R, Revaillot S, Jestin M, Dawson LA. Use of near-infrared reflectance spectroscopy to predict the percentage of dead versus living grass roots. *Plant and Soil*. 2009;317(1-2):309-320. https://doi.org/10.1007/s11104-008-9810-2
8. Izumi Y, Uchida K, Iijima M. Crop production in successive wheat-soybean rotation with no-tillage practice in relation to the root system development. *Plant Prod Sci*. 2004;7(3):329-336. https://doi.org/10.1626/pps.7.329
9. Qin R, Stamp P, Richner W. Impact of tillage on root systems of winter wheat. *Agron J*. 2004;96(6):1523-1530. https://doi.org/10.2134/agronj2004.1523
10. Kätterer T, Hansson AC, Andrén O. Wheat root biomass and nitrogen dynamics-effects of daily irrigation and fertilization. *Plant and Soil*. 1993;151(1-2):21-30. https://doi.org/10.1016/j.agee.2011.02.029
11. Majdi H. Root sampling methods—application and limitations of the minirhizotron technique. *Plant and Soil*. 1996;185(2):255-258. https://doi.org/10.1007/BF02257530
12. Baeten V, Rogez H, Fernández Pierna JA, Vermeulen P, Dardenne P. Vibrational spectroscopy methods for the rapid control of agro-food products. In: Nollet LML, Toldra F, eds. *Handbook of Food Analysis. CRC Press*. 3rd ed. 2015;2:591-614.
13. Abbas O, Dardenne P, Baeten V. Near-infrared, mid-infrared, and Raman spectroscopy. In: Picó Y, ed. *Chemical Analysis of Food: Techniques and Applications*. Waltham, MA: Elsevier Inc; 2012:59-91. https://doi.org/10.1016/B978-0-12-384862-8.00003-0
14. Dale LM, Thewis A, Boudry C, et al. Hyperspectral imaging applications in agriculture and agro-food product quality and safety control: a review. *Appl Spectrosc Rev*. 2013;48(2):142-159. https://doi.org/10.1080/05704928.2012.705800
15. Fernández Pierna JA, Vermeulen P, Amand O. NIR hyperspectral imaging spectroscopy and chemometrics for the detection of undesirable substances in food and feed. *Chemometr Intell Lab*. 2012;117:233-239. https://doi.org/10.1016/j.chemolab.2012.02.004
16. Fernández Pierna JA, Baeten V, Michotte Renier A, Cogdill RP, Dardenne P. Combination of support vector machines (SVM) and near-infrared (NIR) imaging spectroscopy for the detection of meat and bone meal (MBM) in compound feeds. *J Chemometr*. 2004;18(7-8):341-349. https://doi.org/10.1002/cem.877
17. Massart DL, Vandeginste BGM, Buydens LMC, De Jong S, Lewi PJ. In: Smeyers-Verbeke J, ed. *Handbook of Chemometrics and Qualimetrics: Part A*. Amsterdam: Elsevier; 1997.
18. Roumet C, Picon-Cochard C, Dawson LA, et al. Quantifying species composition in root mixtures using two methods: near-infrared reflectance spectroscopy and plant wax markers. *New Phytol*. 2006;170(3):631-638. https://doi.org/10.1111/j.1469-8137.2006.01698.x
19. Rumbaugh MD, Clarck DH, Pendery BM. Determination of root mass ratios in alfalfa-grass mixtures using near infrared reflectance spectroscopy. *J Range Manage*. 1988;41(6):488-490. https://doi.org/10.2307/3899523
20. Rewald B, Meinen C. Plant roots and spectroscopic methods—analyzing species, biomass and vitality. *Front Plant Sci*. 2013;4:393. https://doi.org/10.3389/fpls.2013.00393
21. Kusumo BH, Hedley MJ, Tuohy MP, Hedley CB, Arnold G. Predicting soil carbon and nitrogen concentrations and pasture root densities from proximally sensed soil spectral reflectance. In: Viscarra Rossel R, McBratney A, Minasny B, eds. *Proximal Soil Sensing. Progress in Soil Science*. Dordrecht: Springer; 2010:177-190.
22. Kusumo BH, Hedley MJ, Hedley CB, Tuohy MP. Measuring carbon dynamics in field soils using soil spectral reflectance: prediction of maize root density, soil organic carbon and nitrogen content. *Plant and Soil*. 2011;338(1-2):233-245. https://doi.org/10.1007/s11104-010-0501-4
23. Dale LM, Thewis A, Rotar I, et al. Chemometric tools for NIRS and NIR hyperspectral imaging. *Bull UASVM Agric* 2012;69(1);70-76. http://orbi.ulg.ac.be/handle/2268/132855. Accessed July 06, 2017.
24. Shahin MA, Symons SJ, Hatcher DW. Quantification of mildew damage in soft red winter wheat based on spectral characteristics of bulk samples: a comparison of visible-near-infrared imaging and near-infrared spectroscopy. *Food Bioproc Tech*. 2014;7(1):224-234. https://doi.org/10.1007/s11947-012-1046-8
25. Fernández Pierna JA, Baeten V, Dubois J, Burger J, Lewis EN, Dardenne P. NIR imaging—theory and applications. In: Brown S, Tauler R, Walczak B, eds. *Comprehensive Chemometrics*. Oxford: Elsevier; 2009;4:173-196.
26. Degrune F, Theodorakopoulos N, Dufrêne M, et al. No favorable effect of reduced tillage on microbial community diversity in a silty loam soil (Belgium). *Agr Ecosyst Environ*. 2016;224:12-21. https://doi.org/10.1016/j.agee.2016.03.017

27. Burger J, Geladi P. Hyperspectral NIR image regression part I: calibration and correction. *J Chemometr*. 2005;19(3-4):355-363. https://doi.org/10.1002/cem.986

28. Snee RD. Validation of regression models: methods and examples. *Technometrics*. 1977;19(4):415-428. https://doi.org/10.2307/1267881

29. Fernández Pierna JA, Lecler B, Conzen JP, Niemoeller A, Baeten V, Dardenne P. Comparison of various chemometric approaches for large near infrared spectroscopic data of feed and feed products. *Anal Chim Acta*. 2011;705(1-2):30-34. https://doi.org/10.1016/j.aca.2011.03.023

30. Wise BM, Gallagher NB, Bro R, Shaver JM, Windig W, Koch RS. PLS_Toolbox version 4.0 for use with MATLAB TM. Wenatchee, WA: Eigenvector Research Inc; 2006. http://mitr.p.lodz.pl/raman/jsurmacki/pliki/zajecia/LMDiT/cw4i5/LMDiT_PLS_Manual_4.pdf. Accessed July 06, 2017.

31. Konda Naganathan G, Grimes LM, Subbiah J, Calkins CR, Samal A, Meyer GE. Partial least squares analysis of near-infrared hyperspectral images for beef tenderness prediction. *Sens Instrumen Food Qual*. 2008;2(3):178-188. https://doi.org/10.1007/s11694-008-9051-3

32. Zhang X. Support vector machines. In: Sammut C, Webb G, eds. *Encyclopedia of Machine Learning*. New York: Springer; 2010:941-946. https://doi.org/10.1007/978-0-387-77242-4

33. Fernández Pierna JA, Baeten V, Dardenne P. Screening of compound feeds using NIR hyperspectral data. *Chemometr Intell Lab*. 2006;84(1-2):114-118. https://doi.org/10.1016/j.chemolab.2006.03.012

34. Vermeulen P, Fernández Pierna JA, van Egmond HP, Dardenne P, Baeten V. Online detection and quantification of ergot bodies in cereals using near infrared hyperspectral imaging. *Food Addit Contam A*. 2012;29(2):232-240. https://doi.org/10.1080/19440049.2011.627573

35. Manley M, du Toit G, Geladi P. Tracking diffusion of conditioning water in single wheat kernels of different hardnesses by near infrared hyperspectral imaging. *Anal Chim Acta*. 2011;686(1):64-75. https://doi.org/10.1016/j.aca.2010.11.042

36. Soon YK, Arshad MA. Comparison of the decomposition and N and P mineralization of canola, pea and wheat residues. *Biol Fertil Soils*. 2002;36(1):10-17. https://doi.org/10.1007/s00374-002-0518-9

37. Kamruzzaman M, Barbin D, ElMasry G, Sun DW, Allen P. Potential of hyperspectral imaging and pattern recognition for categorization and authentication of red meat. *Innovative Food Sci Emerg Technol*. 2012;16:316-325. https://doi.org/10.1016/j.ifset.2012.07.007

**How to cite this article:** Eylenbosch D, Bodson B, Baeten V, Fernández Pierna JA. NIR hyperspectral imaging spectroscopy and chemometrics for the discrimination of roots and crop residues extracted from soil samples. *Journal of Chemometrics*. 2017;e2982. https://doi.org/10.1002/cem.2982