



HAL
open science

Comparison of the efficacy of spectral pre-treatments for wheat and weed discrimination in outdoor conditions

X. Hadoux, N. Gorretta, J.M. Roger, R. Bendoula, Gilles Rabatel

► To cite this version:

X. Hadoux, N. Gorretta, J.M. Roger, R. Bendoula, Gilles Rabatel. Comparison of the efficacy of spectral pre-treatments for wheat and weed discrimination in outdoor conditions. *Computers and Electronics in Agriculture*, 2014, 108, pp.242-249. 10.1016/j.compag.2014.08.010 . hal-01098280

HAL Id: hal-01098280

<https://hal.science/hal-01098280>

Submitted on 23 Dec 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Comparison of the efficacy of spectral pre-treatments for wheat and weed discrimination in outdoor conditions

Xavier Hadoux^a, Nathalie Gorretta^a, Jean-Michel Roger^a, Ryad Bendoula^a,
Gilles Rabatel^a

^a*Irstea UMR ITAP, 361 rue J-F Breton BP 5095, 34196 Montpellier Cedex 5, France*

Abstract

The optimal processing of spectral data often requires specific pre-treatments. In the context of spectral discrimination, results can be greatly improved using the relevant pre-treatment. Most importantly, the pre-treatment must be suited to the nuisance variability that has to be removed. This study focuses on discrimination of weed and wheat using spectra acquired in outdoor conditions with uncontrolled lighting and leaf orientations. Both vegetation spectra are highly similar, and due to the context of acquisition, a lot of spectral variability is present. This nuisance variability is modeled using an additive, multiplicative and noise term, each of which affects the measured spectra. Several pre-treatments were therefore evaluated according to their potential to deal with this variability and their effects were described in the feature space. Finally, results obtained with these pre-treatments combined with two discrimination methods (PLS-LDA and Gaussian SVM) are compared and discussed. Results showed that, thanks to their ability to remove nuisance variability, most pre-treatments are effective in terms of classification accuracy. Gaussian SVM classification results are less influenced by pre-treatments than those of PLS-LDA, since the former compensates the pre-treatment effect by using a different non-linear kernel. For this data-set, the best discrimination result was obtained using the combination logarithm and PLS-LDA. Logarithm actually transforms the multiplicative effect into and additive one, which is then effectively dealt with by PLS-LDA.

Keywords: Pre-treatment, Discrimination, Supervised Classification, Reflectance Spectrum, Chemometrics, Hyperspectral

[☆]full text: <http://dx.doi.org/10.1016/j.compag.2014.08.010>

1. Introduction

In the field of spectral data discrimination, vectors to be classified lie in a high-dimensional space which leads to poor classification results when using standard statistical classifiers. In fact, Hughes (1968) has even demonstrated that as space dimensionality increases there is a corresponding drop in statistical mean accuracy. The problem is even more complex because spectral variables are also highly correlated, which implies that covariance matrices of spectra are rank deficient and are thus not invertible. As a result, usual discrimination methods such as Fisher's Linear Discriminant Analysis (LDA) cannot be applied to spectral data.

To handle these problems, in the chemometrics field, Partial Least Squares (PLS) has been introduced by Herman Wold (Wold, 1966) in order to solve an over-determined regression problem through dimension reduction. In the case of discrimination, the same idea of applying dimension reduction as a first step has been used in Partial Least Squares Linear Discriminant Analysis (PLS-LDA) (Barker and Rayens, 2003). PLS-LDA thus tackles both problems, i.e. high-dimensional vectors and highly correlated vectors, and has led to successful discrimination results in spectroscopy (Liu and Rayens, 2007).

In the pattern recognition field, discrimination problems generally relate to high-dimensional data. For discrimination purposes, Support Vector Machine (SVM) has been introduced in (Vapnik, 1998) and has proven its efficiency in finding robust class separators even with highly correlated spectral variables, e.g., in remote sensing applications (Plaza et al., 2009; Fauvel et al., 2012; Rajan et al., 2008). The success of SVM also comes from its ability to deal with non-linearly separable classes by using a kernel that maps the data into a higher dimensional space (Aizerman et al., 1964; Karimi et al., 2006).

The use of spectral properties to discriminate plant species has been reviewed in (Zwiggelaar, 1998), since then most studies have revealed that spectral indices only lead to an accurate discrimination when weeds and crop have enough spectral differences (Piron et al., 2008; Pedersen, 2001; Thorp and Tian, 2004; Sui et al., 2008). In other cases, finding weeds within crop rows is a much more difficult problem to address, and thus requires a more advanced use of the spectral information (Brown and Noble, 2005; Slaughter

et al., 2008; Hadoux et al., 2012).

The discrimination methods previously described could be directly applied to the spectra. However, these spectra are often affected by nuisance variability due to the environment which has to be dealt with while at the same time keeping the spectral variability that comes from the objects' chemical composition. For example, accurate spectral discrimination has been obtained under controlled leaf orientations in Franz et al. (1991), but authors have shown that a major drop in discrimination accuracy occurs as soon as leaf orientation changes. In Vigneau et al. (2011), authors have explored the effect of leaf orientation in order to build a prediction model of nitrogen content using hyperspectral imaging. They modeled the effect of outdoor lighting and orientation using a linear model and overcame this effect using standard normal variate (SNV) pre-treatment. Recent studies have evaluated the utility of several pre-treatments primarily for regression purposes. For example, Buddenbaum and Steffens (2012a), Buddenbaum and Steffens (2012b) and Stevens et al. (2010) have reviewed different pre-treatments for soil component estimation. Recent studies (Jones et al., 2012; Vidal and Amigo, 2012) have also evaluated the efficacy of several pre-treatments, under controlled lighting conditions, for chemical resolution applications.

To the best of our knowledge, the effects of pre-treatments have not been studied in a spectral discrimination context, and especially while taking into account the physical meaning of the nuisance variability that occurs in outdoor conditions with uncontrolled lighting and leaf orientations.

In this paper, we aim to answer the following question: based on the acquired spectral data, which combination of pre-treatments and discrimination method leads to the most accurate and reliable results? To do so, in section 2 we model the nuisance variability due to the outdoor environment and optical measurement variations. In this section, we also present several pre-treatments that are adapted to deal with this variability, their scope and their effect in the feature space. In section 3 our experimental setup and experiment design are detailed. Finally, in section 4, our experimental results obtained with PLS-LDA and Gaussian SVM are presented and discussed. Results obtained with PLS-LDA are also detailed in terms of both the class separability and the bias occurring between training and test set.

2. Theoretical aspects

2.1. Notations

In this paper, the following symbolic conventions are used. Boldface capital characters are used for matrices, e.g. \mathbf{X} ; small bold characters are used for column vectors, e.g. \mathbf{x} ; row vectors are denoted by the transpose notation, e.g. \mathbf{x}^T ; non bold characters are used for scalars, e.g. matrix elements x_{ij} , vector element x_i or indices i . The identity matrix of \mathbb{R}^n is denoted by \mathbf{I}_n .

When dealing with spectral data, two representations can be used:

- the individual space (in \mathbb{R}): a spectrum is represented as a curve depending on the wavelength: $r(\lambda)$. This space should be only used for visualization purposes.
- the feature space (in \mathbb{R}^p) in which a spectrum corresponds to a vector \mathbf{r} in a p -dimensional space where p is the number of wavelengths. Since multi-variate discrimination methods act in this space, any modification on the spectrum should be understood in this space.

In our study we therefore describe the pre-treatments effect in the feature space and we represent their effects in the individual space (for practical representation reasons).

2.2. Optical model

Let us consider a reflectance hyperspectral image which was obtained by using a reference surface for lighting measurement and correction (see Vigneau et al. (2011)). Due to the uncontrolled lighting conditions and leaf orientations when acquiring the hyperspectral image, the measured reflectance spectrum for a given leaf can vary a lot. This measured reflectance spectrum $r(\lambda)^{meas}$ has been represented in Vigneau et al. (2011) as a linear function of the leaf reflectance spectrum $\rho(\lambda)$:

$$r^{meas}(\lambda) = m \cdot \rho(\lambda) \quad (1)$$

where m is a multiplicative term that models the effect due to the uncontrolled leaf orientation toward the incident light.

In Vigneau et al. (2011), the additive term only corresponds to a specular reflection from the sun directly toward the sensor. In our study, another additive term that corresponds to spectrometers common defects is also added

to equation 1. It is modeled as a linear trend (Barnes et al., 1989) which is scaled by the wavelength: $a_0 + a_1 \cdot \lambda$. Finally, a noise term $n(\lambda)$ is also added to the model which corresponds to statistical fluctuations in the photoelectron number (Poisson distribution). Note that the noise variance depends on the wavelength and is directly related to the silicon sensitivity of the CCD sensor. The final model of our study is given in equation 2.

$$r^{meas}(\lambda) = m \cdot \rho(\lambda) + a_0 + a_1 \cdot \lambda + n(\lambda) \quad (2)$$

2.3. Spectral pre-treatments

This study is focused on spectral pre-treatments that are able to reduce or remove the nuisance variability sources given in the equation 2, i.e. the multiplicative term m , the additive term a_0 and a_1 and the noise term $n(\lambda)$ from the measured spectrum. Using this notation, the term $\rho(\lambda)$ is referred to as useful spectral information while the others are referred to as a nuisance variability (for discrimination purposes).

The pre-treatments evaluated in this study that can cope with some parts of the nuisance variability sources are listed in Table 1. For clarity, a nomenclature to describe these pre-treatments according to their effects and scopes have been used. The term *global* refers to a pre-treatment that has a unique effect on a given spectrum without differentiating between wavelengths. The term *piecewise* describes a pre-treatment that uses its neighboring wavelengths. The term *local* refers to a pre-treatment that treats each wavelength independently. We also differentiate the actual pre-treatment effect using terms: *reduce*, *transform* or *remove*. For example, smoothing reduces the noise piecewise using a numerical filter of a given width, logarithm transforms the multiplicative effect (into additive) locally and centering removes the additive effect globally.

In the following, the main pre-treatments are detailed and their effect in feature spaces are described.

2.3.1. Dealing with additive effects

The centering pre-treatment (also known as constant detrending or detrending of order 0) and linear detrending (or detrending of order 1) have been chosen to deal with additive term a_0 and the linear trend $a_0 + \lambda \cdot a_1$ respectively.

Their expressions in the feature space can be expressed using the same notation. Let D_k be the matrix $\mathbf{D}_k = [\mathbf{p}^0, \mathbf{p}^1, \dots, \mathbf{p}^k]$, where $\mathbf{p}^i = [1^i, 2^i, \dots, p^i]^T$

Table 1: Pre-treatments evaluated in this study, their effect on spectrum: remove (rm), transform (tr) or decrease (dc) and their scope : global (G), piecewise (P), local (L) as described in Section 2.3.

Pre-treatment	Abreviation	Additive		Multiplicative	Noise
		Constant	Linear		
None	R				
Center	D_0	rm - G			
Detrend linear	D_1		rm - G		
Normalize std	N_{std}			rm - G	
Normalize Norm 1	N_1			rm - G	
Normalize Norm 2	N_2			rm - G	
Center - Normalize std	$D_0 - N_{std}$	rm -G		rm -G	
Center - Normalize Norm 1	$D_0 - N_1$	rm -G		rm -G	
Center - Normalize Norm 2	$D_0 - N_2$	rm -G		rm -G	
Logarithm	Log			tr - L	
Logarithm - Center	$Log - D_0$			rm - L	
Smoothing	SG_0				dc - P
1st Derivative	SG_1	rm - P			dc - P
2nd Derivative	SG_2	rm - P	rm - P		dc - P
Smoothing -Center	$SG_0 - D_0$	rm - G			dc - P
Smoothing - Detrend linear	$SG_0 - D_1$		rm - G		dc - P
Smoothing - Normalize std	$SG_0 - N_{std}$			rm - G	dc - P
Smoothing - Normalize Norm 1	$SG_0 - N_1$			rm - G	dc - P
Smoothing -Normalize Norm 2	$SG_0 - N_2$			rm - G	dc - P
Smoothing -Center - Normalize std	$SG_0 - D_0 - N_{std}$				dc - P
Smoothing -Center - Normalize Norm 1	$SG_0 - D_0 - N_1$				dc - P
Smoothing -Center - Normalize Norm 2	$SG_0 - D_0 - N_2$				dc - P
Smoothing -Logarithm	$SG_0 - Log$			tr - L	dc - P
Smoothing -Logarithm - Center	$SG_0 - Log - D_0$			rm - L	dc - P

and p corresponds to the number of wavelengths. Detrending of any order k corresponds to projecting the spectrum orthogonal to the matrix \mathbf{D}_k (Boulet and Roger, 2012):

$$\mathbf{r}^{corr} = \left(\mathbf{I}_P - \mathbf{D}_k (\mathbf{D}_k^T \mathbf{D}_k)^{-1} \mathbf{D}_k^T \right) \mathbf{r}^{meas} \quad (3)$$

If $k = 0$, $\mathbf{p}^0 = [1, \dots, 1]^T$, the spectrum is projected on an hyperplane orthogonal to the vector composed only of ones. Similarly, linear detrending corresponds in the feature space to the projection of the spectrum orthogonal to the matrix \mathbf{D}_1 . Note that in practice, as k increases, special care should be given for numerical inversion of the matrix $\mathbf{D}_k^T \mathbf{D}_k$ because of large differences between its eigenvalues (Meyer, 2000).

Expressed in \mathbb{R}^p , it is clear that the spectral dimension is reduced from p to $p - 1$ and from p to $p - 2$ for the detrend 0 and for the detrend 1 respectively. Indeed, the spectrum is projected orthogonal to a matrix of rank

1 and 2 respectively. Removing these dimensions when dealing with data in such a high-dimensional space might not appear problematic, especially since the actual dimension of visible and near infra red (V-NIR) spectra is usually quite low due to the high correlation between wavelengths (Naes et al., 2002). However, when these pre-treatments are applied, the global intensity information is also definitively removed. Class separability therefore decreases if discriminatory information is colinear to these (removed) directions in the feature space.

2.3.2. Dealing with multiplicative effects

In this study, we considered two types of pre-treatment to deal with the multiplicative term, i.e., normalizing and logarithm transformation.

With the normalizing approach, a single factor computed from all wavelengths is used to normalize the spectrum. Several normalizing factors can be used and lead to different results that depend on the context, e.g. number of wavelengths of the spectral data, presence of outliers, intensity of the reflectance values, etc. Usual normalizing factors are norm 1, norm 2 or standard deviation (*std*) of the spectrum:

$$\mathbf{r}^{corr} = \frac{1}{N(\mathbf{r}^{meas})} \mathbf{r}^{meas} \quad (4)$$

where $N(\cdot) = \{\|\cdot\|_1, \|\cdot\|_2, std(\cdot)\}$ corresponds to the normalizing factor. Note that by using variation information around the mean, *std* is less sensitive to centered noise. In the feature space, normalizing using norm 1 corresponds to the projection of the spectrum onto the space orthogonal to the vector $[1, \dots, 1]^T$ in \mathbb{R}^p . When using norm 2, the spectrum is projected onto the unit sphere of dimension p . Finally, normalizing using standard deviation corresponds to the projection of the spectrum onto a $(p - 1)$ -dimensional hypersphere of unit radius (centered on its mean). Spectra with different angles (in \mathbb{R}^p) before the pre-treatment remain different after projection. Therefore, if discriminatory information between spectra is in a space spanned by a difference in intensity (length of the vector in \mathbb{R}^p), the information is lost after projection.

Another approach to deal with multiplicative term is the logarithm transformation which is described as follows:

$$\mathbf{r}^{corr} = \log(\mathbf{r}^{meas}) \quad (5)$$

, where the $\log(\cdot)$ operator is applied to every element of \mathbf{r}^{meas} individually. In the feature space, this transformation corresponds to using a logarithmic scale instead of a linear one. Doing so, no spectral dimension is lost and the transformation is reversible. However, logarithm transformation itself does not remove the multiplicative term, it only transforms the multiplicative term into an additive term.

2.3.3. Dealing with noise: smoothing and derivative

Savitzky-Golay smoothing (Savitzky and Golay, 1964) performs a piecewise polynomial fitting with specified width and order to the spectrum. This smoothing method has been successfully adopted by the spectroscopy community because it is able to reduce noise while keeping the meaningful variation that occurs at these wavelengths.

In the individual space, it increases the correlation between neighboring wavelengths and reduces random noise leading to smoother curves. In the feature space, it increases the correlation between variables, therefore increasing the matrix column correlations and decreasing the real rank of the spectral data.

Savitzky and Golay have also provided a smoothed version of first and second numerical derivatives with a high signal to noise ratio. Some theoretical advantages and practical properties of Savitzky-Golay derivatives have been studied (Luo et al., 2005). In our study, first derivatives are used to remove baselines and second derivatives to remove slopes.

2.3.4. Combination of pre-treatments

The previously described pre-treatments taken individually are designed to correct only for a part of the nuisance variability of the model given in equation 2. We must therefore combine them optimally to attempt to remove more of the nuisance variability. Table 1 lists every combination of pre-treatments evaluated in our study.

Because the order of applying different pre-treatments is important, a consistent notation ordered from left to right is used in this paper. For example, Smoothing - Center - Normalize Norm 1, corresponds to apply first a smoothing, then a centering and finally a normalization (with norm 1).

Among these combinations, the most widely used in spectral data analysis is called Standard Normal Variate (SNV) (e.g. see (Sun et al., 2011)) and it

combines centering and normalizing using std:

$$\mathbf{r}^{SNV} = \frac{1}{std(\mathbf{r}^{meas})} \left(\mathbf{I}_P - \mathbf{p}_0 (\mathbf{p}_0^T \mathbf{p}_0)^{-1} \mathbf{p}_0^T \right) \mathbf{r}^{meas}. \quad (6)$$

In the feature space, a spectrum is first mapped on the hyperplane of dimension $(p - 1)$ orthogonal to $[1, \dots, 1]^T$. The resulting vector is then projected onto a $(p - 1)$ -dimensional hypersphere of unit radius. Geometric details of this transformation are well described in (Fearn et al., 2009).

3. Material and methods

3.1. Experimental setup

Short-range in-field hyperspectral images were acquired using a camera (Hypspec V-NIR 1600, Norks Elektro Optikk, Norway) with a spatial resolution of 0.2 mm/pixel at a distance of 1 meter above the ground. The acquired spectra were composed of 160 spectral bands ranging from 415 nm to 1000 nm. Each image was composed of 1600 pixels per row and the number of rows, that depended on the scanning length, was in average 3000. A typical RGB image reconstructed from the hyperspectral image is represented in Figure 1. In this figure, weeds are also highlighted using the Normalized Difference Vegetation Index (NDVI) (Rouse et al., 1973).

Using a calibrated reference surface in each image (see on the right-hand side of Figure 1), radiance values were transformed into reflectance values. In doing so, the obtained reflectance images were almost independent of the atmospheric conditions and illumination. Image acquisitions were carried out over two consecutive years, at a growth stage that corresponds to the weeding period, and under various lighting conditions, i.e. sunny and cloudy weather, shaded areas, etc. In each image, weeds (dicotyledons), wheat, soil and the reference surface were in the camera field of view as represented in Figure 1. In total, 9 images, in which each weed position was recorded in the field, were used for the study and were randomly separated into two groups: 4 images for the training and 5 images for the test. A ground truth map of weed and wheat was then manually created for each image. Finally, around 1600 spectra for training set and 1600 spectra for the test set were randomly selected and labeled as either wheat or weed according to the ground truth.

We applied every pre-treatment listed in Table 1 to the training and test sets. For Savitzky-Golay smoothing and derivatives, the parameters that

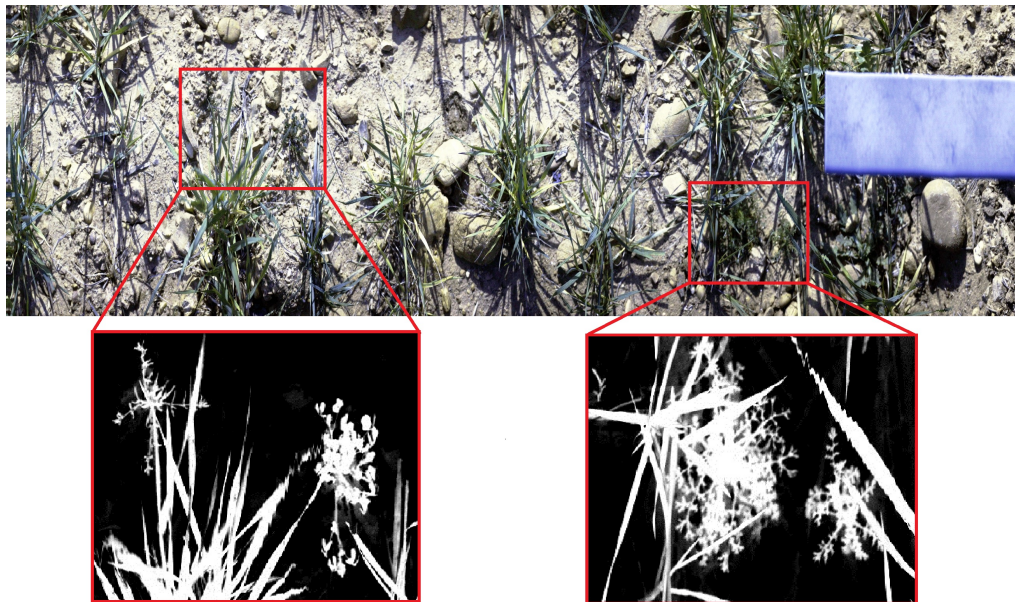


Figure 1: Illustration of a typical image from the data set. (top) RGB image reconstructed from the hyperspectral image. Note the reference surface used for reflectance correction on the right-hand side. (bottom) NDVI images zoomed on the weeds.

gave the best results using a cross-validation (Esbensen and Geladi, 2010) on the training set were a third order polynomial and a kernel width of 17.

Concerning the two-class discrimination problem (wheat and weed), we used Partial Least Square Linear Discriminant Analysis (PLS-LDA) and Support Vector Machine with a Radial Basis Function Kernel (Gaussian SVM). Every model (pre-treatment and classifier) was trained and tuned using a cross-validation on the training set and its performance was then assessed on the independent test set (Esbensen and Geladi, 2010). Both classifiers, i.e. PLS-LDA and Gaussian SVM were compared in terms of the prediction error (on the test set).

For Gaussian SVM, the model complexity was also reported as a way to understand how complex the discrimination task as a function of the pre-treatment.

For PLS-LDA, bias and class separability of the scores were also measured. The bias corresponded to a shift occurring between the best separator in the training set and the separator that would be the best in the test set. The class separability was measured by computing the Wilk's lambda

(Λ_{Wilks}) (Everitt and Skrondal, 2002) on the PLS-LDA scores. In Figure 2 the bias and separability are illustrated for our two class discrimination problem.

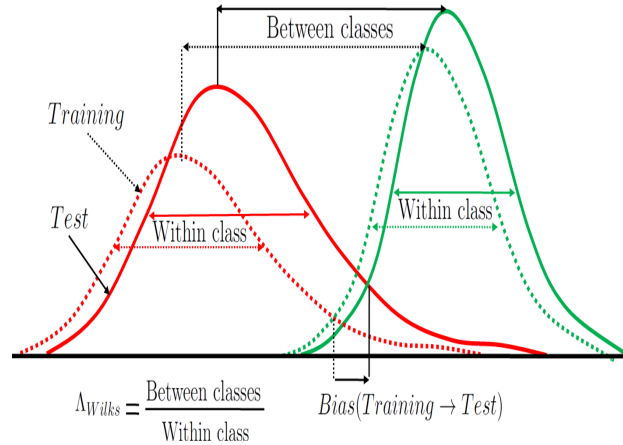


Figure 2: Illustration of the bias and separability for PLS-LDA performance assessment. Green and red represent two different classes. Dotted lines represent a training set and plain lines represent a test set.

4. Experimental results

Pre-treatment effects on the training spectra are displayed in Figure 3 with green for wheat and red for weed. In each class, the solid line corresponds to the average spectrum and the plain color area represents its variability (standard deviation) within the training set. Raw spectra (without any pre-treatment) are shown in Figure 3a. Additive corrections are represented in Figures 3c (detrend of order 0) and Figure 3d (linear detrend). Figures 3b, 3e, 3f and 3g correspond to multiplicative corrections using logarithm, normalizing factors norm 1, norm 2 and std respectively. Smoothing, first and second derivative are shown in Figures 3h, Figures 3i and Figures 3j.

Discrimination results obtained with every combination of pre-treatments (detailed in Table 1) are displayed in Figure 4 using a radar chart. Each edge corresponds to a specific pre-treatment and each color to a discrimination method, i.e. blue for Gaussian SVM and red for PLS-LDA. Gaussian SVM model complexity is defined by the number of support vectors used to create the model divided by the available number of training samples and is displayed (in %) with a dashed blue line. For PLS-LDA, only the rate

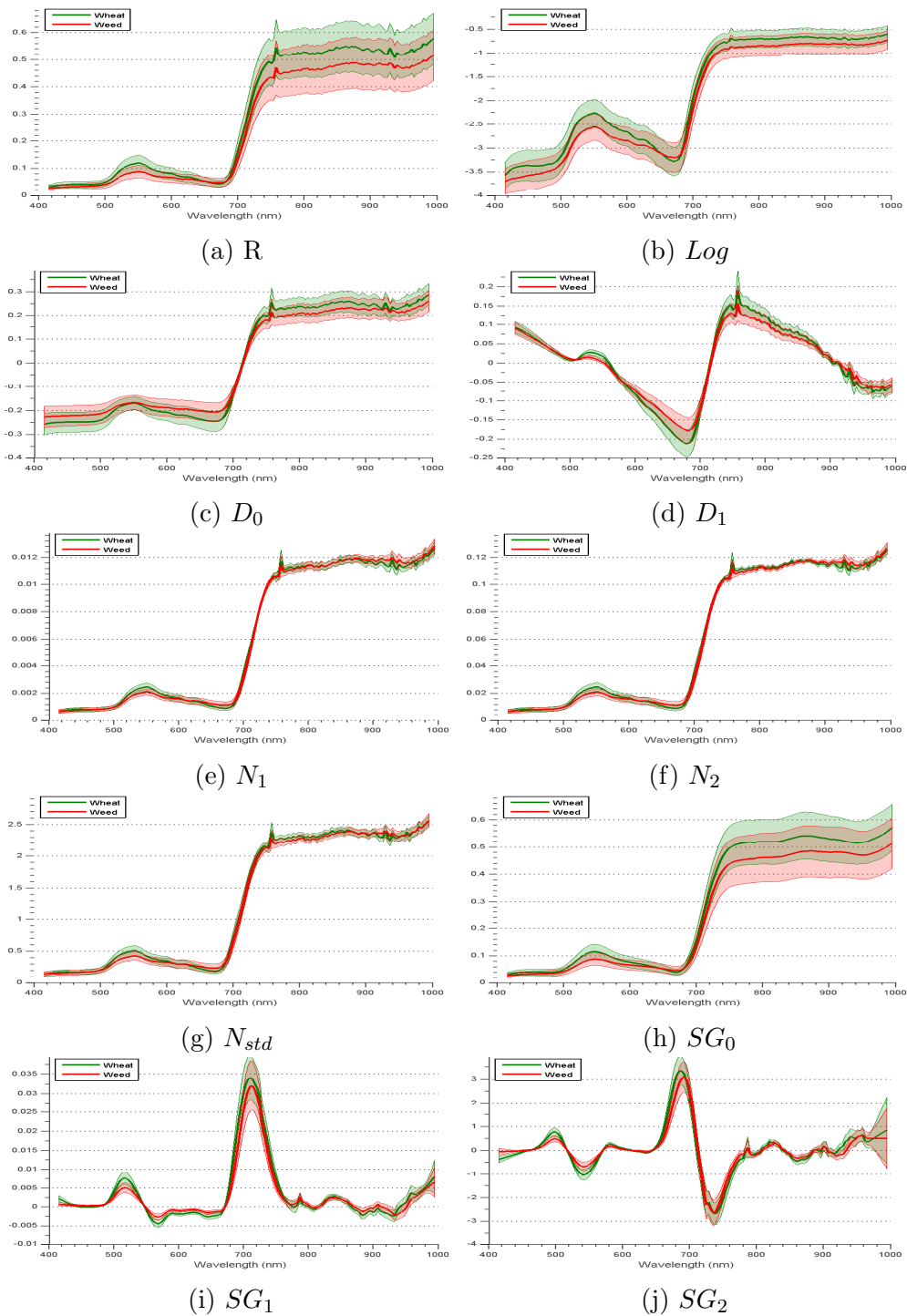


Figure 3: Pre-treatment effects on real reflectance spectra taken from the training set. Wheat spectra are in green and weed spectra in red. Filled areas correspond to spectral variability within each class. Abbreviations are detailed in Table 1.

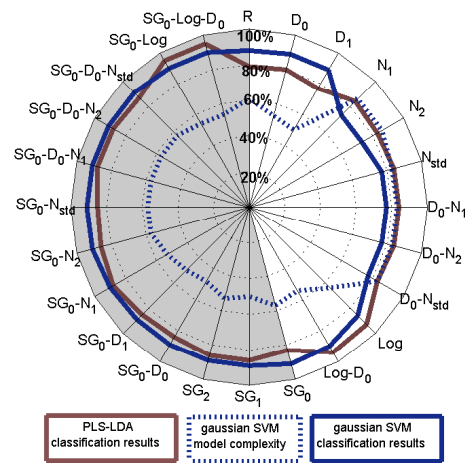


Figure 4: Radar chart displaying the rate of good discrimination in (%) for both discrimination methods : PLS-LDA (red), and Gaussian SVM (blue). SVM model complexity (dotted blue) corresponds to the percentage of support vectors used to create the model (in the available training set). Gray area corresponds to result where at least a smoothing has been applied. Abbreviations are detailed in Table 1.

of good discrimination was considered since the same model complexity (6 latent variables) was obtained with every pre-treatment.

The separability obtained with the training (blue) and test set (red) are represented in Figure 5a. The bias obtained from the training set to the test set is represented in Figure 5b. In both radar charts results are expressed in percentages and relative to the results obtained with raw spectra. Note that on the radar chart, a result lying outside the main area corresponds to an unfavorable pre-treatment and a result close to the center corresponds to a useful pre-treatment.

5. Discussion

5.1. Effects on spectra

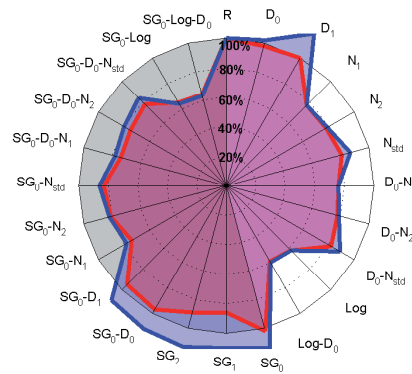
Raw vegetation spectra in Figure 3a clearly indicate a plateau in the near infra red (NIR) region (750 nm - 1000 nm). The increasing slope between 670 nm and 750 nm, called red-edge, is a robust indicator to discriminate vegetation from the background (Brown and Noble, 2005), (Slaughter et al., 2008). The bump between 500 nm and 600 nm corresponds to the green color of vegetation and the strong absorption at around 670 nm indicates the presence of chlorophyll in the leaves. The peak at 760 nm indicates that oxygen absorption was not accurately corrected using the reference surface (due to low signals creating an large amount of quantification noise). Above 900 nm, spectra are noisier because of a lower sensitivity of the CCD sensor in the NIR.

As for the discrimination between these two classes, the NIR plateau is on average higher and the red-edge steeper for wheat spectra than for weed spectra. A slight difference in greenness between both vegetation spectra is also noticeable between 500 nm and 600 nm.

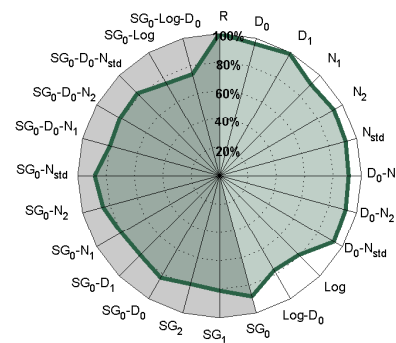
Within each class a large amplitude variation at every wavelength is clearly visible. Comparing Figures 3c and Figures 3d (additive corrections) with Figures 3e, 3f and 3g (multiplicative corrections) indicates that this spectral variation is mostly due to a multiplicative effect.

After smoothing, spectra clearly indicate less random noise in NIR and the 760 nm peak is reduced. However, main spectral features remained, e.g. red edge, plateau and greenness differences between weed and wheat.

With derivatives, visual interpretation of the curves is easier since derivative increases small variations. Boundary effects are also visible at both ends of the spectra due to numerical filtering computational issues.



$$(a) \left(\frac{\Lambda_{Wilks}}{\Lambda_{Wilks}(R)} \right)^{-1}$$



$$(b) \frac{Bias(Training \rightarrow Test)}{Bias(Training \rightarrow Test)(R)}$$

Figure 5: Radar charts obtained with PLS-LDA. (a) the class separability for the training (red) and the test set (blue). The value has been inverted for visualization purpose. (b) the bias between training and test. Results are expressed in percent and relative to the results obtained with raw spectra. Gray area corresponds to results where at least a smoothing has been applied. Abbreviations are detailed in Table 1.

5.2. Discrimination

Curves presented in the previous section demonstrate the difficulty of the discrimination task between wheat and weeds despite any sort of pre-treatment. There is in fact very little difference between weed and wheat spectra as well as the variability within each class being large.

However, most pre-treatments evaluated in this study were effective in dealing with the discrimination problem. Both PLS-LDA and Gaussian SVM

obtained a discrimination accuracy of more than 80% which demonstrates their potential in spectral analysis. With raw data (R), high discrimination rates are observed for both methods, i.e. 87% for SVM and 80% for PLS-LDA. The SVM model used 60% of the available training spectra to create its model. Around a thousand points in a 160-dimensional vector space is still low and confirms the high correlation of the spectral data (the space is almost empty). PLS-LDA appears more sensitive to the pre-treatment than Gaussian SVM, the latter compensates for the loss of information by creating a more complex non-linear model.

The best pre-treatment is clearly the logarithm transformation, especially when used with PLS-LDA (96%). Logarithm does not suppress multiplicative effect, it transforms this effect into an additive one which is then entirely handled by PLS-LDA.

On the other hand, normalization forced spectra to lie in a lower dimensional subspace by forcing low and high intensity values to be on the same scale. Therefore, some discriminatory features such as NIR plateau value and greenness were diminished. Gaussian SVM model complexity indicates that problems occurred when using normalization of any sort. The obtained model was therefore more complex and only based on subtle features that were still present in the corrected spectra. For PLS-LDA, however, normalization gave better results than with raw spectra. Since PLS-LDA cannot deal with a multiplicative effect, it was better to lose some information than to be perturbed by the multiplicative effect.

For both classifiers, smoothing appears to be an important first pre-treatment step by reinforcing the correlation between neighboring wavelengths and by reducing noise. It thus helps the discrimination method to build simpler models that are less sensitive to an error when using a specific wavelength.

Removing additive effects (by centering or removing linear trend) did not improve PLS-LDA discrimination results since it can handle additive effects on its own if they are present within the training set. On the contrary, Gaussian SVM cannot deal with additive effects even if they are present in the training set, i.e. it learns the discrimination boundaries knowing only some examples of translated vectors. By removing additive effects, SVM model complexity decreased therefore helping the model to find a better class separation.

Using the first and second derivative does not significantly improve discrimination results for either classifier. It provides slightly lower Gaussian

SVM model complexity while keeping the discrimination result constant. PLS-LDA is not greatly affected by this pre-treatment step since it can handle the same effects on its own.

5.3. PLS-LDA separability and bias

In this study, spectra came from images acquired at different times and parts of the field, creating between-image variability. Since spectra from the training set were taken from images different from those of the test set this variability could not be learned by the model. PLS-LDA first step is to reduce the dimension of the space. The class separability and the bias were thus computed in this lower dimensional space avoiding statistical estimation problems that occur in high-dimensional spaces (Hughes, 1968).

In the radar chart of Figure 5a, similar patterns are observed with the training set and the test set. The separability was always worse with the test set than with the training set (except when using the logarithm). This difference indicates a slight over-fitting of the model to the training set even using a proper validation method. None of the pre-treatments was designed to handle the bias occurring between training set and test set, however, from the chart in Figures 5b, we observe they all had a positive influence on the bias.

Generally, removing multiplicative effects gave more separability than with raw data. Logarithm transformation provided the best results in terms of both separability and bias even without smoothing and/or centering. Logarithm kept all the data information by only transforming the multiplicative effect. The discrimination model was therefore based on the main spectral differences and was more robust to small changes occurring between the training set and the test set. With normalizing, results of each factor were nearly equivalent with only a slightly poorer result for standard deviation. Standard deviation, as explained in section 2.3.2, makes more assumptions about the data distributions.

Removing additive effects by detrending or derivative considerably reduced the class separability and gave worse results than with raw data. Indeed, when using PLS-LDA, if the additive effect is present in the training set, there is no need to remove it through any pre-treatment. In terms of bias, almost no improvement is noticeable by removing additive effect.

Smoothing had an large positive effect in terms of bias but had a negative effect in terms of separability. The bad effect on separability was likely caused by the filter being too strong, thus resulting in the removal of some useful

discrimination information. On the other hand, smoothing removed small variations that might have occurred between images thus reducing the bias.

6. Conclusion

This study examines the influence of pre-treatments on the discrimination of wheat and weed spectra in outdoor conditions. We demonstrated that to deal with the modeled nuisance variability the most effective combination of pre-treatment and discrimination methods was the logarithm transformation with PLS-LDA. Logarithm was the only pre-treatment that dealt with multiplicative effect without explicitly projecting the data on a lower dimensional subspace in which useful discriminatory information was lost.

Experimental results also validate the model of measured reflectance for outdoor application with uncontrolled light, and leaf orientations. This approach will however require further studies to increase the understanding of the bias occurring between images. This would help to create a more detailed model and hopefully some treatments that can handle this bias.

7. References

- Aizerman, M.A., Braverman, E.A., Rozonoer, L., 1964. Theoretical Foundations of the Potential Function Method in Pattern Recognition Learning. *Automation and Remote Control* 25, 821–837.
- Barker, M., Rayens, W., 2003. Partial least squares for discrimination. *Journal of Chemometrics* 17, 166–173.
- Barnes, R., Dhanoa, M., Lister, S.J., 1989. Standard normal variate transformation and detrending of near-infrared diffuse reflectance spectra. *Applied spectroscopy* 43, 772–777.
- Boulet, J.C., Roger, J.M., 2012. Pretreatments by means of orthogonal projections. *Chemometrics and Intelligent Laboratory Systems* 117, 61–69.
- Brown, R., Noble, S., 2005. Site-specific weed management: sensing requirements - What do we need to see? *Weed Science* 53, 252–258.
- Buddenbaum, H., Steffens, M., 2012a. Mapping the distribution of chemical properties in soil profiles using laboratory imaging spectroscopy, SVM and PLS regression. *EARSeL eProceedings* 11, 25–32.

- Buddenbaum, H., Steffens, M., 2012b. The Effects of Spectral Pretreatments on Chemometric Analyses of Soil Profiles Using Laboratory Imaging Spectroscopy. *Applied and Environmental Soil Science* 2012, 1–12.
- Esbensen, K.H., Geladi, P., 2010. Principles of Proper Validation: use and abuse of re-sampling for validation. *Journal of Chemometrics* 24, 168–187.
- Everitt, B., Skrondal, A., 2002. *The Cambridge dictionary of statistics*. Cambridge University Press Cambridge.
- Fauvel, M., Chanussot, J., Benediktsson, J.J., 2012. A spatial - spectral kernel-based approach for the classification of remote-sensing images. *Pattern Recognition* 45, 381–392.
- Fearn, T., Riccioli, C., Garrido-Varo, A., Guerrero-Ginel, J.E., 2009. On the geometry of SNV and MSC. *Chemometrics and Intelligent Laboratory Systems* 96, 22–26.
- Franz, E., Gebhardt, M., Unklesbay, K., 1991. The use of local spectral properties of leaves as an aid for identifying weed seedlings in digital images. *Transactions of the American Society of Agricultural Engineers* .
- Hadoux, X., Gorretta, N., Rabatel, G., 2012. Weeds-wheat discrimination using hyperspectral imagery, in: *CIGR-Ageng 2012, International Conference on Agricultural Engineering*.
- Hughes, G., 1968. On the mean accuracy of statistical pattern recognizers. *Information Theory, IEEE Transactions on* 14, 55–63.
- Jones, H.D., Haaland, D.M., Sinclair, M.B., Melgaard, D.K., Collins, A.M., Timlin, J.a., 2012. Preprocessing strategies to improve MCR analyses of hyperspectral images. *Chemometrics and Intelligent Laboratory Systems* 117, 149–158.
- Karimi, Y., Prasher, S., Patel, R., Kim, S., 2006. Application of support vector machine technology for weed and nitrogen stress detection in corn. *Computers and Electronics in Agriculture* 51, 99–109.
- Liu, Y., Rayens, W., 2007. PLS and dimension reduction for classification. *Computational Statistics* 22, 189–208.

- Luo, J., Ying, K., He, P., Bai, J., 2005. Properties of SavitzkyGolay digital differentiators. *Digital Signal Processing* 15, 122–136.
- Meyer, C., 2000. *Matrix analysis and applied linear algebra*.
- Naes, T., Isaksson, T., Fearn, T., Davies, T., 2002. *A user friendly guide to multivariate calibration \ classification*. NIR Publications.
- Pedersen, B., 2001. *Weed density estimation from digital images in spring barley*. Ph.D. thesis. Royal Veterinary and Agricultural University.
- Piron, a., Leemans, V., Kleynen, O., Lebeau, F., Destain, M.F., 2008. Selection of the most efficient wavelength bands for discriminating weeds from crop. *Computers and Electronics in Agriculture* 62, 141–148.
- Plaza, A., Benediktsson, J.A., Boardman, J.W., Brazile, J., Bruzzone, L., Camps-Valls, G., Chanussot, J., Fauvel, M., Gamba, P., Gualtieri, A., Marconcini, M., Tilton, J.C., Trianni, G., 2009. Recent advances in techniques for hyperspectral image processing. *Remote Sensing of Environment* 113, S110–S122.
- Rajan, S., Ghosh, J., Crawford, M., 2008. An Active Learning Approach to Hyperspectral Data Classification. *IEEE Transactions on Geoscience and Remote Sensing* 46, 1231–1242.
- Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W., 1973. Monitoring vegetation systems in the Great Plains with ERTS, in: *Third ERTS Symposium*, NASA SP-351, pp. 309–317. doi:citeulike-article-id:12009708.
- Savitzky, A., Golay, M., 1964. Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry* 36, 1627–1639.
- Slaughter, D., Giles, D., Downey, D., 2008. Autonomous robotic weed control systems: A review. *Computers and Electronics in Agriculture* 61, 63–78.
- Stevens, A., Udelhoven, T., Denis, A., Tychon, B., Liroy, R., Hoffmann, L., van Wesemael, B., 2010. Measuring soil organic carbon in croplands at regional scale using airborne imaging spectroscopy. *Geoderma* 158, 32–45.
- Sui, R., Thomasson, J.A., Hanks, J., Wooten, J., 2008. Ground-based sensing system for weed mapping in cotton. *Computers and Electronics in Agriculture* 60, 31–38.

- Sun, H., Li, M., Li, D., 2011. The vegetation classification in coal mine overburden dump using canopy spectral reflectance. *Computers and Electronics in Agriculture* 75, 176–180.
- Thorp, K., Tian, L., 2004. A Review on Remote Sensing of Weeds in Agriculture. *Precision Agriculture* 5, 477–508.
- Vapnik, V., 1998. *Statistical Learning Theory*. Wiley.
- Vidal, M., Amigo, J.M., 2012. Pre-processing of hyperspectral images. Essential steps before image analysis. *Chemometrics and Intelligent Laboratory Systems* 117, 138–148.
- Vigneau, N., Ecarnot, M., Rabatel, G., Roumet, P., 2011. Potential of field hyperspectral imaging as a non destructive method to assess leaf nitrogen content in Wheat. *Field Crops Research* 122, 25–31.
- Wold, H., 1966. Estimation of Principal Components and Related Models by Iterative Least squares. Academic Press, New York. pp. 391–420.
- Zwiggelaar, R., 1998. A review of spectral properties of plants and their potential use for crop/weed discrimination in row-crops. *Crop Protection* 17, 189–206.