Predictive ability of soil properties to spectral degradation from laboratory Vis-NIR spectroscopy data

Karine R.M. Adeline, Cecile Gomez, N. Gorretta, J.M. Roger

Author-produced version of the article published in Geoderma, 2017, N°288, p.143-153.
The original publication is available at http://www.sciencedirect.com
http://dx.doi.org/10.1016/j.geoderma.2016.11.010

Predictive ability of soil properties to spectral degradation from laboratory Vis-NIR spectroscopy data

K.R.M. Adeline[1], C. Gomez[1], N. Gorretta[2], J.-M. Roger[2]

[1] IRD, UMR LISAH (INRA-IRD-SupAgro), F-34060 Montpellier, France

[2] IRSTEA, UMR ITAP, BP 5095, 34196 Montpellier Cedex 5, France

Corresponding author:

K.R.M. Adeline, karine.adeline@gmail.com

Highlights:

- Clay, free iron, calcium carbonate and pH were predicted from Lab Vis-NIR spectra.
- Eight spectral resolutions were tested from 3 nm to 200 nm.
- PLSR predictive ability is poorly impacted by the spectral degradation.
- All soil properties are predictable until a spectral resolution of 60 nm.
- Spectral features and correlations between soil properties explained the results.

Abstract

Laboratory Visible-Near Infrared (Vis-NIR) spectroscopy is a good alternative to costly physical and chemical soil analysis to estimate a wide range of soil properties. Various statistical methods relate soil Vis-NIR spectra to soil properties including partial least-squares

Author-produced version of the article published in Geoderma, 2017, N°288, p.143-153.
The original publication is available at http://www.sciencedirect.com
http://dx.doi.org/10.1016/j.geoderma.2016.11.010

regression (PLSR), the most common multivariate statistical technique in soil science. Most efforts are generally dedicated to the comparison of methodologies and their optimization for the estimation of soil properties. In this paper, we examined the sensitivity of PLSR model predictions of physico-chemical soil properties to different spectral configurations derived from Vis-NIR spectra and described by three parameters: the number of spectral bands, the spectral resolution and the spectral sampling interval. The initial database is composed of 1961 spectral bands, spectral resolutions of 3 and 10 nm in the 400-1000 nm and 1000-2500 nm ranges, respectively, and a spectral sampling interval of 1 nm. Seven degraded spectral configurations were built from this reference database with a number of spectral bands decreasing from 328 to 10, a spectral resolution decreasing from 3 nm to 200 nm, and a spectral sampling interval equaling the spectral resolution. All of these databases were composed of 148 soil samples collected at a Mediterranean site. Four soil properties were selected for their different spectroscopic behavior: clay, free iron oxides, calcium carbonate ($CaCO_3$) and pH. PLSR predicted these variables, and the results were as follows: (1) the prediction performances of the PLSR models were accurate and globally stable with a spectral resolution between 3 to 60 nm regardless of the soil properties (R^2 decreased from 0.8 to 0.77 for clay, from 0.88 to 0.84 for $CaCO_3$, from 0.66 to 0.58 for pH and remained constant at 0.78 for iron), (2) the prediction performance decreased, but remained acceptable for clay, iron oxides and $CaCO_3$ at spectral resolutions between 60 and 200 nm ($R^2 > 0.7$), (3) the sensitivity of a given soil property to instrumental spectral configurations depended on its spectral features and correlations with other soil properties.

Keywords:

Laboratory Vis-NIR spectroscopy, soil properties, physico-chemical features, partial least squares regression, spectral resolution.

Author-produced version of the article published in Geoderma, 2017, N°288, p.143-153.
The original publication is available at http://www.sciencedirect.com
http://dx.doi.org/10.1016/j.geoderma.2016.11.010

1. Introduction

Soil quality assessment assists decision making for a range of global issues, such as food production and precision agriculture (Tilman et al., 2002), and soil carbon and water stocks related to climate change (Lal, 2004; Seneviratne et al., 2010). The monitoring and determination of soil properties provide a better understanding of the physical and chemical processes in soil environments (Atzberger, 2002). Reflectance spectroscopy can quantitatively estimate these soil properties more cost-effectively and rapidly compared to traditional laboratory analysis. Visible-near infrared (Vis-NIR) spectroscopic data in the 350-2500 nm range has been widely used to analyze soils because it efficiently correlates the chemical components with their specific absorption spectral features (e.g., Abrams and Hook, 1995; Palacios-Orueta and Ustin, 1998; Viscarra et al., 2006; Weidong et al., 2002). Absorption features in the visible spectrum are dominated by electronic molecular excitations, and those in the NIR range contain a combination of overtone molecular vibrations. However, the separation of each soil component contribution from Vis-NIR spectra is a challenging task due to the complex nature of soil matrix with multiple spectral feature overlappings and strong spectral collinearities between soil properties (Gobrecht et al., 2013). Consequently, pre-processing strategies are embedded as a first step before prediction calibration, in order to improve the extraction of useful information from both additive and multiplicative effects superimposed in the reflectance spectra (Peng et al., 2014). Rinnan et al. (2009) and Hadoux et al. (2014) give a good review of them. Various mathematical methods have been employed to predict soil properties from soil Vis-NIR spectra, such as multiple regression analysis (Ben-Dor and Banin, 1995), stepwise multiple linear regression (Shibusawa et al., 2001), multivariate adaptive regression splines (Shepherd and Walsh, 2002), principal component

Author-produced version of the article published in Geoderma, 2017, N°288, p.143-153.
The original publication is available at http://www.sciencedirect.com
http://dx.doi.org/10.1016/j.geoderma.2016.11.010

regression (e.g., Chang et al., 2001), support vector machine regression (SVMR; Stevens et al., 2010) and the continuum removal technique (Lagacherie et al., 2008). The partial least-squares regression method (PLSR; Wold et al., 2001) is the most common multivariate statistical technique for spectral calibration and prediction of soil properties (Viscarra Rossel et al., 2006). Most of the studies using chemometric analysis of spectroscopic data for soil properties primarily focus on optimizing prediction performances, such as comparing methodologies and assessing new methods (Mouazen et al., 2010; Peng et al., 2014) or using spectral feature selection methods (Vohland et al., 2014). And few of them have been carried out to analyze the dependency of these prediction performances based on the quality of the initial spectral database acquired by a given spectroscopic sensor (e.g. Knadel et al., 2013; Mouazen et al., 2005; Peng et al., 2014). Peng et al. (2014) evaluated different spectral sampling intervals from 1 nm to 10 nm with ASD FieldSpec 3 spectra to estimate soil organic carbon content with both SVMR and PLSR methods, and established that 9 nm was the best choice. Knadel et al. (2013) and Mouazen et al. (2005) found poor differences in prediction performance for clay content, soil organic carbon and soil moisture by comparing spectrometers with different spectral specifications (spectral resolutions: 1-10 nm; spectral intervals: 1.377-6 nm; spectral ranges: 300-1700 nm, 350-2500 nm, 1000-2500 nm, 400-498 nm) and also measurement technologies (combination of diode array, scanning monochromator and Fourier Transform).

Accordingly, the purpose of this paper is to study the PLSR model ability of soil properties prediction to spectral degradation from laboratory Vis-NIR spectroscopy using only one spectrometer (ASD). The initial Vis-NIR laboratory spectra were acquired with 1961 spectral bands, spectral resolutions of 3 and 10 nm in the 400-1000 nm and 1000-2500 nm ranges, respectively, and a spectral sampling interval of 1 nm. Seven degraded spectral configurations were built from this initial database with the spectral sampling interval equal to the spectral

Author-produced version of the article published in Geoderma, 2017, N°288, p.143-153.
The original publication is available at http://www.sciencedirect.com
http://dx.doi.org/10.1016/j.geoderma.2016.11.010

resolution. These configurations combine hyperspectral to multispectral scenarios from original spectroscopic data with a number of bands decreasing from 328 to 10 and spectral resolutions decreasing from 3 to 200 nm. Four soil properties were predicted from a dataset of 148 soil samples collected at a Mediterranean site in France: clay, free iron oxides, calcium carbonate ($CaCO_3$) and pH. Clay granulometry is related to bulk density, roughness and permeability in soils. Its association with calcium carbonate content is indicative of soil vulnerability to erosion (Le Bissonnais, 1996). The presence of iron oxides is a pertinent indicator of heavy metal soil contamination (Kemper and Sommer, 2002), weathering (Dematthê and Garcia, 1999) and fertility (Bartholomeus et al., 2007). And variations in pH are related to soil acidity and fertility. These properties were selected for their different spectral spectroscopic behavior and were predicted by PLSR.

2. Materials and methods

2.1 Field sampling

Soil samples were collected from a 24 km² area in the La Peyne catchment (43°29' N and 3°22' E), 60 km west of Montpellier, France (Figure 1a). This area is primarily rural (> 90 %), and the climate is typical of the Mediterranean region characterized by sub-humid to prolonged dry seasons. The annual rainfall median over 20 years is 634 mm, and the average annual evapotranspiration is 1102 mm. In the upstream part of the catchment, little of the land is cultivated because of steep terrain slopes and Mediterranean maquis shrubs, whereas in the downstream part, moderate terrain slopes provide ideal use for vineyard agriculture. At the surface, the soil appears crusted from tillage practices reinforced by strong episodic rainfall. Underneath, the soil substrate is largely heterogeneous Miocene marine sediments, i.e., marl,

Author-produced version of the article published in Geoderma, 2017, N°288, p.143-153.
The original publication is available at http://www.sciencedirect.com
http://dx.doi.org/10.1016/j.geoderma.2016.11.010

sandy loam, and calcareous sandstone with low carbon content (< 2 %). Thin Miocene lacustrine limestone composes the hillslopes in the middle of La Peyne valley, which is representative of cuesta topography. Hill backslopes are partially overlain with successive alluvial deposits, ranging from Pliocene to Holocene, and differ in their initial nature and duration of weathering conditions. The presence of clay minerals is dominated by illite and kaolinite with a weak abundance of a mixture of illite-smectite.

In June of 2009, 148 soil samples were homogeneously collected over this area (Figure 1b). All of them were composed of five sub-samples at 0-5 cm depth within a field plot of 10×10 m^2 at locations recorded on a Garmin GPS instrument with 2 m of accuracy.

[Figure 1]

2.2 Soil analysis

The soil samples were homogenized, air-dried and sieved to a particle size of 2 mm. Four elementary soil properties were determined using classical physico-chemical soil analysis (Baize and Jabiol, 1995): clay content (granulometric fraction < 2 µm, pipette method from NF X 31-107), free iron oxide content (Mehra-Jackson method from NF ISO 11885), calcium carbonate $CaCO_3$ content (volumetric method from NF ISO 10693), and pH H2O (method NF ISO 10390).

2.3 Laboratory Vis-NIR spectroscopy

Spectral reflectance of the 148 soil samples (sieved and dried) were acquired on an ASD pro FR Portable Spectrometer (Analytical Spectral Devices Inc., Boulder, CO, USA) in the

Author-produced version of the article published in Geoderma, 2017, N°288, p.143-153.
The original publication is available at http://www.sciencedirect.com
http://dx.doi.org/10.1016/j.geoderma.2016.11.010

spectral range of 350-2500 nm, which was first calibrated with a white reference plate, i.e., a 12 × 12 cm^2 Spectralon® panel (Labsphere, North Sutton, USA).

Two 90 W tungsten halogen light sources with aluminum reflectors (24 V, ~ 3000° K color temperature, Power DC supply) were placed on each side of the sample, with the light beam at 45° from vertical to create a 50 cm distance between the lamps and the sample. The sensor with a Field of view (FOV) of 8° was positioned from the nadir at 0.6 m from the sample, providing a 0.9 diameter measurement spot. The soil was placed into a Petri dish with dimensions of 1.0 cm (height) × 14 cm (diameter). The reflectance of one sample was computed as the mean of 30 measurements in a spectral range of 350-2500 nm. From 350 to 1000 nm, the spectral sampling interval of the ASD spectrometer is originally 1.4 nm for a spectral resolution of 3 nm. From 1000 to 2500 nm, the spectral sampling interval is 2 nm for a spectral resolution of 10 nm (http://www.asdi.com). However, the reflectance available to the user is pre-processed by the ASD software and is consequently oversampled to 1 nm in both spectral ranges leading to a total number of spectral bands of 2151. This number was then reduced to 1961 by removing the spectral bands within the 350-440 nm and 2400-2500 nm ranges due to their low instrumental signal-to-noise ratios. We finally associated this spectral database to the reference spectral configuration, which was named ASD_1/1.

2.4 Degraded spectral configuration

Each spectral configuration is defined by three parameters: the number of spectral bands (*N*), the spectral resolution also called the full half width maximum (*FHWM*) and the spectral sampling interval (*SSI*) (Figure 2). To build each degraded spectral configuration, the initial spectra from ASD_1/1 were resampled with Gaussian filters whose tails were arbitrarily cut to twice their width, following the filter response function *G(x)*:

Author-produced version of the article published in Geoderma, 2017, N°288, p.143-153.
The original publication is available at http://www.sciencedirect.com
http://dx.doi.org/10.1016/j.geoderma.2016.11.010

$$G(x) = \exp\left(\frac{-(x-x_0)^2}{2.\sigma^2}\right) \text{ with } \sigma = \frac{FHWM}{2.\sqrt{2.\log(2)}} \quad (1)$$

where x is the spectral step determined by the *SSI*, x_0 is the mean of the filter and equals the wavelength at which the resampling was performed, and σ is the width of the filter.

[Figure 2]

Seven degraded spectral configurations were derived from the reference ASD_1/1 (Table 1). Except for ASD_1/1, it was assumed that the spectral sampling interval equaled the spectral resolution, such that only two variable parameters remained: the number of spectral bands *N* and the spectral resolution *FHWM*. They were characterized by *N* values decreasing from 328 to 10, and *FHWM* values in the 440-1000 nm and 1000-2400 nm spectral ranges varying from 3 to 200 nm and from 10 to 200 nm, respectively. They were subsequently named Config_3/10, Config_5/10, Config_10/10, Config_40/40, Config_60/60, Config_100/100 and Config_200/200 (Table 1) and covered hyperspectral to multispectral scenarios.

[Table 1]

2.5 Principal component analysis

A principal component analysis (PCA) was performed on the mean-centered reflectance data for each spectral configuration. It provides a set of explanatory orthogonal vectors or principal components in relation to the analyzed variance (Harman, 1976). If the data were composed

Author-produced version of the article published in Geoderma, 2017, N°288, p.143-153.
The original publication is available at http://www.sciencedirect.com
http://dx.doi.org/10.1016/j.geoderma.2016.11.010

of random noise, all of the components would explain approximately the same quantity of variance, and the spectral degradation would regularly and slowly make decreased the variance. If the data were composed of variations around spectral features, the decreasing would be rapid because a small number of components would carry a huge amount of variance. Thus, the amount of useful information carried by the reflectance data to the spectral configurations was assessed by the percentage of variance V_m explained by the m^{th} principal component over a total number of M:

$$V_m = \frac{\lambda_m}{\sum_{m=1}^{M} \lambda_m} \times 100 \quad (2)$$

where λ_m is the m^{th} eigenvalue.

2.6 Partial least squares regression method

The partial least squares (PLS) method (Wold et al., 2001) is a dimensional space reduction technique like PCA. The difference between PLS and PCA is that the former is trained to maximize the covariance between the scores of the spectra and the response variable (i.e. each soil property), whereas PCA scores are not necessarily correlated with the response variable. Following this covariance optimization constraint, the PLS algorithm performs iterative rotated projections to estimate matrices of latent variables and scores associated with both the spectra and the response variable.

Partial least squares regression (PLSR) is the association of a PLS reduction with a classical multivariate linear regression explaining the correlation between the Vis-NIR spectra and the soil property:

Author-produced version of the article published in Geoderma, 2017, N°288, p.143-153.
The original publication is available at http://www.sciencedirect.com
http://dx.doi.org/10.1016/j.geoderma.2016.11.010

$$\hat{\boldsymbol{y}} = \boldsymbol{X} \ . \hat{\boldsymbol{b}} + b_0 \quad (3)$$

where $\boldsymbol{X} \in R^{K,N}$ is a matrix of K spectra with N spectral bands, $\hat{\boldsymbol{y}} \in R^{K,1}$ is the vector of the estimated soil property values of the K soil samples, $\hat{\boldsymbol{b}} \in R^{N,1}$ is the vector of the estimated PLSR regression coefficients (b-coefficients), and b_0 is the intercept (Haaland and Thomas, 1988).

2.6.1 Model development

Prior to statistical analysis, the reflectance spectra (R) were converted into absorbance logarithmic spectra (Log(1/R)) and mean-centered. Then, for each soil property, the database of each spectral configuration was divided into a calibration set of 99 samples (i.e., 2/3 of the total data, named BD_Calib) and a validation set of 49 samples (i.e., 1/3 of the total data, named BD_Valid). The soil property values were first sorted in ascending order. Second, the sample of lowest value was assigned to BD_Valid, and the next two samples were set in BD_Calib. This alternating procedure was continued for all of the samples and ensured that both BD_Calib and BD_Valid had similar distributions for a given soil property. The BD_Calib databases were then inspected to detect the spectral outliers by using the Mahalanobis distance in combination with PCA (the three first principal components are retained). Outliers were removed from BD_Calib when their Mahalanobis distance was greater than 3 (Mark and Tunnell, 1985).

The PLSR model was built with BD_Calib using a leave-one-out cross-validation (LOOCV; Wold, 1978). The LOOCV procedure consists of building a learning calibration model with

Author-produced version of the article published in Geoderma, 2017, N°288, p.143-153.
The original publication is available at http://www.sciencedirect.com
http://dx.doi.org/10.1016/j.geoderma.2016.11.010

K-1 samples from all *K* samples within BD_Calib and then predicting the soil property value of the sample that was not used in that learning model. This process was repeated for the *K* samples. Application of the PLSR model to BD_Valid for final soil property prediction requires the selection of an optimal number of latent variables (p_{opt}), which was based on 2 criteria:

- The value was arbitrarily set below 10 because of the limitation of the number of spectral bands available in the most degraded spectral configuration (i.e., Config_200/200, cf. Table 1),
- The value showing the first local minimum of the root mean square errors of cross-validation (RMSECV) was chosen (Viscarra, 2007). RMSECV was calculated as follows:

$$RMSECV = \sqrt{\frac{\sum_{k=1}^{K}(y_k - \hat{y}_{k,-k})^2}{K}} \quad (4)$$

where *K* is the number of samples for BD_Calib, $\boldsymbol{y}_k$ is the k^{th} measured soil property value and $\hat{\boldsymbol{y}}_{k,-k}$ is the predicted soil property value obtained by removing the k^{th} sample in BD_Calib.

2.6.2 Model evaluation

Estimation of the soil properties with respect to the spectral configurations was assessed in terms of (i) the PLS model's learning ability, (ii) the PLSR model's prediction performance in relation to the number of latent variables and (iii) the analysis of key wavelengths used in the PLSR model calibration process.

Author-produced version of the article published in Geoderma, 2017, N°288, p.143-153.
The original publication is available at http://www.sciencedirect.com
http://dx.doi.org/10.1016/j.geoderma.2016.11.010

The learning ability of any regression can be assessed by the ratio of the explained variance var($\hat{\mathbf{y}}$) to the original variance var($\mathbf{y}$), calculated by the coefficient of determination R^2. Since best models concentrate R^2 values close to 1, we used a rescaled R^2, named Rs^2, which is the ratio between R^2 and the residual variance, in order to highlight the differences close to 1, such as:

$$Rs_p^2 = \frac{R_p^2}{1-R_p^2} \quad (5)$$

where p is the number of latent variables in the PLSR model.

The prediction performances of the PLSR models were based on the following figures of merit:

- The coefficient of determination R^2 of the predicted values against the measured values in BD_Calib and BD_Valid, respectively.
- The root mean square errors of cross-validation for BD_Calib (RMSECV, see Equation 4) and prediction for BD_Valid (RMSEP), which was calculated by:

$$RMSEP = \sqrt{\frac{1}{K}.\sum_{k=1}^{K}(\boldsymbol{y}_k - \hat{\boldsymbol{y}}_k)^2} \quad (6)$$

In the PLSR calibration process performed for the optimal number of latent variables (p_{opt}), a wavelength n was considered to have a significant importance if it fulfilled the two following conditions:

- Its b-coefficient was larger than the standard deviation of the b-coefficients for all of the spectral bands (Viscarra-Rossel et al., 2008)

Author-produced version of the article published in Geoderma, 2017, N°288, p.143-153.
The original publication is available at http://www.sciencedirect.com
http://dx.doi.org/10.1016/j.geoderma.2016.11.010

- Its variable importance in the projection (VIP) was higher than 1 (Chong and Jun, 2005; Wold et al., 1993, 2001):

$$VIP_n = N.\sum_{p=1}^{p_{opt}} w_{p,n}^2 . R_p^2 \quad (7)$$

where $w_{p,n}^2$ is the loading weight for the p^{th} latent variable in p_{opt}.

All procedures were performed using the following R software (R Core Team, 2012) and the following specific packages were used: i) ade4 for principal component analysis (Dray and Dufour, 2007) and ii) pls for PLSR (Mevik and Wehrens, 2007).

3. Results

3.1 Soil analysis

Clay and iron contents were normally distributed with means of 232.8 g.kg^{-1} and 1.20 g.100 g^{-1} and standard deviations of 71.6 g.kg^{-1}and 0.59 g.100 g^{-1}, respectively (Figure 3a and 3b). The distribution of $CaCO_3$ concentrations covered a broad range from 0 to 440 g.kg^{-1} and followed a skewed distribution due to numerous soil samples containing very low amounts of $CaCO_3$ (Figure 3c). pH exhibited the smallest variations with an asymmetric normal distribution, a mean value of 8.18 and a standard deviation of 0.67 (Figure 3d). The four soil properties of the 148 soil samples were not correlated with one another (Table 2), except for i) a high positive correlation between pH and $CaCO_3$ (r = 0.68) and ii) a moderate correlation between iron and clay (r = 0.53) and iron and $CaCO_3$ (r = -0.56).

Author-produced version of the article published in Geoderma, 2017, N°288, p.143-153.
The original publication is available at http://www.sciencedirect.com
http://dx.doi.org/10.1016/j.geoderma.2016.11.010

[Figure 3]

[Table 2]

3.2 Reference spectra analysis

The reflectance spectra of the ASD_1/1 configuration contained a chemical absorption peak centered at 2215 nm with a 95 nm width (2150-2245 nm) and a sharp depth (Figure 4b). This result can be associated with a combination of OH stretching and OH-Al bending modes related to the presence of illite, kaolinite and montmorillonite clay materials (e.g., Chabrillat et al., 2006; Lagacherie et al., 2008). A second absorption peak was centered at 2350 nm with a 40 nm width (2330-2370 nm) and a low depth (Figure 4b), which can be associated with CO_3 overtone vibrations related to the presence of $CaCO_3$ (Gaffey, 1987). Different slopes between 400 and 800 nm are related to iron oxides (Bartholomeus and Mulder, 2008; Bayer et al., 2012; Demattê et al., 2004) combining two major constituents: goethite near 550 nm and hematite between 500-600 nm (Figure 4a; Atzberger, 2002). Because pH has no specific spectral signature (e.g., Ben-Dor and Banin 1995; Ben-Dor et al., 2002), this soil property cannot be observed by a specific absorption feature.

[Figure 4]

3.3 PCA analysis for each spectral configuration

Author-produced version of the article published in Geoderma, 2017, N°288, p.143-153.
The original publication is available at http://www.sciencedirect.com
http://dx.doi.org/10.1016/j.geoderma.2016.11.010

Whatever the spectral configuration, the first three PCA principal components (PC1, PC2 and PC3) accounted for more than 98 % of the spectral variance (Table 3).

For the ASD_1/1, the first PCA loading (PC1) exhibited a similar spectral behavior as the mean reflectance of the spectral database (Figure 5), indicating that the highest variance carried by the PCA is primarily dominated by physical multiplicative effects due to light scattering between soil particles. Additionally, computation of the auto-correlation matrix from the spectral database indicated high values of R^2 (> 0.8) for all pairs of wavelengths over the full spectral range (data not shown). This strong spectral correlation might enhance the importance of the variance explained by PC1 (Table 3). Conversely, the second and third PCA principal components loadings (PC2 and PC3) were zero-centered and undertook the chemical spectral features of the soil properties, as highlighted by strong changes for example between 400 and 800 nm and between 2000 and 2400, and close to 1400 nm and 1900 nm (Viscarra and Chen, 2011).

Then, the analysis of the seven degraded spectral configurations showed the same trend for the PC1 loading on one side and the PC2 and PC3 loadings on another side (data not shown). The explained variance of PC1, V_1, was high, above 92 %, regardless of the spectral configuration, whereas V_2 and V_3 represented less than 6 % (Table 3). V_1 decreased from ASD_1/1 to Config_3/10 and then slowly increased until Config_200/200. Oppositely, the variations of V_2 and V_3 increased compared to V_1. Consequently, the first drop in V_1 benefited V_2 and V_3 that carry the useful spectral information (soil property features), but the spectral degradation subsequently favored V_1.

[Table 3]

[Figure 5]

Author-produced version of the article published in Geoderma, 2017, N°288, p.143-153.
The original publication is available at http://www.sciencedirect.com
http://dx.doi.org/10.1016/j.geoderma.2016.11.010

3.4 Learning ability of the PLS model

For the determination of clay, $CaCO_3$ and pH, the learning ability of their PLS models was sensitive to the spectral configurations from a number of latent variables (p) of six, five and six, respectively (Figures 6a, 6c and 6d). However, for iron, it was poorly sensitive to the spectral configurations, regardless of p (Figure 6b). Variations in the rescaled R^2, i.e. Rs^2 (see section 2.6.2), only represent 1.7 for pH compared to 7.0 for $CaCO_3$ and 5.0 for clay. Then, different behaviors in model learning ability were observed with increasing values of p. PLS models for clay, $CaCO_3$ and iron contents continued to learn until Config_100/100 with a gain in Rs^2 values, but stopped learning for Config_200/200 with stable Rs^2 values from p equals to 5. A significant change in model learning ability was noticeable for $CaCO_3$, whose Rs^2 curvature first followed a convex line for fine spectral configurations (e.g., ASD_1/1) and gradually became concave for coarse spectral configurations (e.g., Config_200/200; Figure 6c). At last, PLS models for pH had a low learning ability with almost constant Rs^2 values whatever p or the spectral configuration.

[Figure 6]

3.5 PLSR model prediction performance

The prediction performances of clay, $CaCO_3$ and pH contents were sensitive to the spectral configurations when p was superior to 6 (Figure 7a, 7c and 7d), whereas prediction performances of iron were insensitive to the spectral configurations, whatever the number of latent variables p (Figure 7b). The maximum difference in RMSEP among the spectral

Author-produced version of the article published in Geoderma, 2017, N°288, p.143-153.
The original publication is available at http://www.sciencedirect.com
http://dx.doi.org/10.1016/j.geoderma.2016.11.010

configurations was 12 $g.kg^{-1}$ for clay content, 15 $g.kg^{-1}$ for $CaCO_3$ content and 0.17 for pH. The worst RMSEP was obtained for Config_200/200, but the best RMSEP was not always derived from ASD_1/1, as expected. Variations in RMSEP as a function of *p* were globally similar for all spectral configurations. The first local minimum in RMSEP remained the same except for clay content at Config_3/10 (Figure 7a), and for pH, which did not reach a minimum of convergence (Figure 7d). This result is in line with the low learning ability of the PLS models for pH at increasing values of *p* shown in the previous section. As a result, the optimal number of latent variables p_{opt} for clay content was set to 4 for Config_3/10 and 5 for the other spectral configurations (Figure 7a). Furthermore, p_{opt} was fixed at 8 for $CaCO_3$, 7 for iron and 9 for pH, in all spectral configurations (Figure 7b, 7c and 7d).

The ability to predict soil properties with p_{opt} decreased with spectral degradation, which was highlighted by a decline in R^2_{val} values and an increase in RMSEP values, excepted for iron content (Table 4). In spite of these decreased performances, all soil properties were predictable until Config_60/60 ($R^2_{val} > 0.5$), and in addition, clay, $CaCO_3$ and iron contents still were until Config_200/200 ($R^2_{val} > 0.7$). From ASD_1/1 to Config_200/200, the decrease in R^2_{val} accounted for 0.1 in clay, 0.01 in iron, 0.1 in CaCO3 and 0.3 in pH. On the one hand, the impact of the spectral degradation began slightly after Config_40/40 and strengthened after Config_100/100 for clay content. It started after Config_10/10 for $CaCO_3$ content, and for pH, it began after Config_3/10 and strengthened after Config_60/60 (Table 4). This behavior was confirmed by the analysis of normalized b-coefficients. The variation of the normalized b-coefficients showed similar spectral patterns close to absorption features from ASD_1/1 to Config_40/40 (Figure 8). For the Config_60/60, the spectral patterns at 2200 nm disappeared (Figure 8). And finally, from Config_100/100, rough trends in the b-coefficients variation may only reflected the differences in physical baselines due to light scattering

Author-produced version of the article published in Geoderma, 2017, N°288, p.143-153.
The original publication is available at http://www.sciencedirect.com
http://dx.doi.org/10.1016/j.geoderma.2016.11.010

changes (Figure 8). Same observations were found for the two other soil properties (data not shown). On the other hand, the prediction performance for iron content were constant with R^2_{val} close to 0.78 and low values of RMSEP. This behavior was confirmed by the analysis of normalized b-coefficients (data not shown).

[Figure 7]

[Table 4]

[Figure 8]

3.6 Key wavelengths in the PLSR calibration process

For the reference ASD_1/1, the number of key wavelengths ranged between 100 and 300, which only represents between 5 % and 15 % of the initial number of available spectral bands (i.e., N = 1961; Table 5). Because the pH has no spectral feature, its PLSR model required a number of key wavelengths higher than the other soil properties, whereas $CaCO_3$ having a specific spectral absorption, needs the least. From ASD_1/1 to Config_3/10, this number decreased by a factor of approximately 4 for all soil properties except $CaCO_3$ (Table 5). From Config_3/10 to Config_200/200, this number decreased until a value of 2 for clay content, 4 for iron and $CaCO_3$ contents, and 3 for pH.

Key wavelengths were primarily located close to the spectral absorption features of each soil property (Figure 9). Overall, regardless of the spectral configuration or soil property, these spectral bands were grouped into four spectral ranges. One group was located from 440 to 900 nm and resulted from the presence of iron oxides. The second included spectral bands from

Author-produced version of the article published in Geoderma, 2017, N°288, p.143-153.
The original publication is available at http://www.sciencedirect.com
http://dx.doi.org/10.1016/j.geoderma.2016.11.010

1850 to 2000 nm and was correlated to water content. The third included spectral bands close to 2200 nm related to the presence of illite, kaolinite and smectite clay minerals, and the fourth included spectral bands from 2300 to 2400 nm correlated to $CaCO_3$ and clay minerals.

[Table 5]

[Figure 9]

4. Discussion

The prediction performances of the PLSR models for the four studied soil properties (clay, iron, $CaCO_3$ and pH) were accurate until a spectral resolution of 60 nm and 33 spectral bands (i.e. Config_60/60). So our results were in agreement with Knadel et al (2013) which found no impact of spectral resolution until 10 nm, for clay content prediction. The impact of spectral degradation was weak until Config_200/200, except for pH, whose prediction performance drastically fell after Config_60/60. Two categories of soil properties could be distinguished according to their sensitivity to the spectral degradation. The first category was composed of clay, iron and $CaCO_3$, which are driven by chemical absorption features, whereas the second was composed of pH, which has no spectral feature.

In the first soil property category represented by clay, iron and $CaCO_3$, the prediction performances were affected by the spectral degradation following two criteria: (i) the characteristics of their absorption features (slope or spectral peak width and depth) and (ii) their correlation with other soil properties with good prediction performance. As an illustration, the estimation of the iron content was not impacted by spectral degradation, as shown by the independence of its PLSR models both in terms of learning ability (Figure 6b)

Author-produced version of the article published in Geoderma, 2017, N°288, p.143-153.
The original publication is available at http://www.sciencedirect.com
http://dx.doi.org/10.1016/j.geoderma.2016.11.010

and prediction results (Figure 7b) to the spectral configurations. One reason may be that PLSR models take advantage of the fact that iron content is primarily calibrated on soil colorimetry represented by slopes in a broad spectral region over 400-800 nm and correlated to clay and $CaCO_3$ (Table 2), which also had spectral features. Moreover, the PLSR calibration process can rely on the selection of at least one key wavelength in the visible range, even until the most degraded spectral configuration (Config_200/200, Figure 9b). The estimation of clay and $CaCO_3$ contents would be expected to suffer more from spectral degradation because these soil properties are both dominated by chemical absorptions with narrower or more sparse spectral features than iron content. Nevertheless, if the spectral sensitivity of their PLSR model learning abilities was demonstrated (Figure 6a and 6c), it was relatively low for their PLSR prediction results (Figure 7a and 7c). In addition, the progressive extinction of their spectral absorption features with widths of 95 nm for clay and 40 nm for $CaCO_3$ does not exactly match the initial decrease in their prediction performance. Indeed, the $CaCO_3$ predictions were expected to be more sensitive to the smoothing of the spectra because of its narrower width and lower depth absorption peak than that of the clay content. The correlation of both clay and $CaCO_3$ with iron likely played an important role in maintaining acceptable prediction results until roughly Config_40/40, and only afterwards their prediction performance slowly decreases. Moreover, clay content is governed by both chemical (clay mineralogical abundance) and physical (clay granulometry) effects, which can aid in finding more key wavelengths in PLSR calibration (Figure 9a). As such, they can conserve good prediction results with spectral degradation.

In the second soil property category, represented only by pH and containing no spectral feature, the prediction performances essentially relied on correlation with other properties having spectral features or good prediction performance (e.g., Ben Dor and Banin, 1995). For example, pH estimation was moderately sensitive to the spectral configurations because the

Author-produced version of the article published in Geoderma, 2017, N°288, p.143-153.
The original publication is available at http://www.sciencedirect.com
http://dx.doi.org/10.1016/j.geoderma.2016.11.010

learning ability of its PLS model was very low (Figure 6d) and its PLSR model prediction performance tended to diverge (Figure 7d). PLSR models seemed to rely on the high $CaCO_3$-pH correlation (Table 2) to provide acceptable prediction results until Config_60/60. This is depicted by the sharing of common key wavelengths with $CaCO_3$ (close to 2430 nm; Figure 9d). Additionally, its coefficient of determination in validation (R^2_{val}) approximately followed the same decreasing variations as $CaCO_3$. However, after Config_60/60, pH was no longer predictable ($R^2_{val} < 0.5$).

PLSR models were used to study their dependence on changes in spectral configurations following different numbers of latent variables (p). This analysis was performed on both their learning ability and prediction performance. Until Config_60/60 with the first category of soil properties (clay, iron, $CaCO_3$), the models still learned with a high p value, especially for $CaCO_3$. However, from Config_60/60 to Config_200/200, the ability of the PLSR models to build useful information from increased p values steadily decreased due to less accurate information from the spectral smoothing. pH had a moderate to negligible impact on spectral configurations following the increase in p value. Selecting the optimal number of latent variables (p_{opt}) is typically a critical step in the construction of PLSR models to avoid under- and over-fitting (e.g., Viscarra Rossel, 2007), but the sensitivity of our PLSR models to the spectral configurations was independent of the choice of p_{opt} above a value of 5 for all soil properties.

Analysis of the spectral degradation led us to consider many parameters, such as the number of spectral bands (N), the spectral resolution ($FHWM$) and the spectral sampling interval (SSI). PCA applied to the spectra from the spectral configurations revealed that most soil information was obtained from Config_3/10 (with lowest PC1 variance) and not from ASD_1/1 as one would expect (Table 3). This may result from spectral oversampling by the ASD software, which adds artificial noisy spectral information to the raw spectral acquisition

Author-produced version of the article published in Geoderma, 2017, N°288, p.143-153.
The original publication is available at http://www.sciencedirect.com
http://dx.doi.org/10.1016/j.geoderma.2016.11.010

as also illustrated by comparing ASD_1/1 and Config_3/10 b-coefficients (Figure 8). Therefore, the best spectral configuration tend to be the original ASD acquisitions (prior to oversampling) that is close to the spectra defined by Config_3/10, unless the spectral filters from the ASD software are not ideally Gaussian. Here, the undesirable added noise is specific to the ASD instrument, but this specificity can vary among other spectrometers. Nevertheless, this type of noise is generally reduced by the use of the PLSR and prior spectral transformations such that the results may not fundamentally change. Furthermore, in our study, the impact of the signal-to-noise ratio (*SNR*) was assumed to be negligible (Knadel et al., 2013; Lagacherie et al., 2008) and the spectra acquired by the ASD had a good *SNR* over the 350-2500 nm range except at extreme wavelengths. Finally both sources of noise might have an impact on the initiation of the decrease in prediction ability with the spectral degradation for soil properties.

The applicability and reproducibility of our results need to be evaluated for other soil properties such as organic carbon or sand contents, which could be classified in the first and second soil property categories, respectively. As well these results need to be evaluated with different soil properties correlations and distributions and from spectral databases acquired by other spectroscopic sensors. Nevertheless, by identifying two main drivers of the sensitivity due to spectral degradation, namely the inter-correlation between soil properties and the soil spectral features, this study provides an a priori assessment of the prediction of soil properties from Vis-NIR spectra.

5. Conclusions

The ability to predict soil properties from initial to degraded spectral configurations was assessed for clay, free iron oxides, $CaCO_3$ and pH, with PLSR models and spectra acquired on

Author-produced version of the article published in Geoderma, 2017, N°288, p.143-153.
The original publication is available at http://www.sciencedirect.com
http://dx.doi.org/10.1016/j.geoderma.2016.11.010

an ASD portable field spectrometer. The estimation of soil properties involving specific spectral features (i.e., clay, iron and $CaCO_3$) were qualitatively sensitive to spectral degradation, as expected. However, in our dataset, this decrease in performance was low, especially when the spectral feature was large and pronounced (e.g., iron) or the correlation between soil properties was strong (e.g., clay-iron and $CaCO_3$-iron). On the other hand, soil properties with no spectral features such as pH only relied on the beneficial effect of correlations (e.g., $CaCO_3$-pH) and their prediction was more sensitive to spectral degradation because less soil information was progressively contained in the spectra.

For all of our four soil properties, the prediction performance of the PLSR models were accurate with respect to the spectral configurations at a resolution from 3 to 60 nm with a respective number of spectral bands decreasing from 1961 to 33. Subsequently, only clay, iron and $CaCO_3$ were predictable up to a spectral resolution of 200 nm when only 10 spectral bands were available. Further analyses should use spectral feature selection methods with irregular spectral samplings and spectral resolutions similar to Volhand et al. (2014), who studied soil properties independently. In addition, spectral band selection methods such as CovSel (Roger et al., 2011) can be used to study the prediction performance of all soil properties simultaneously. Finally, because the sensitivity to spectral degradation in hyperspectral scenarios was relatively low, high-resolution spectrometers may not be necessary for soil investigations. As initiated by Knadel et al., 2013 et Mouazen et al., 2005 and followed by this study, this assumption should be further examined by testing different spectral ranges, which may also provide new perspective for the design of new low-cost spectrometers for soil Vis-NIR spectroscopy.

Acknowledgments

Author-produced version of the article published in Geoderma, 2017, N°288, p.143-153.
The original publication is available at http://www.sciencedirect.com
http://dx.doi.org/10.1016/j.geoderma.2016.11.010

This research was granted by the TOSCA-CNES project « MiHySpecSol - **Mi**ssion **HY**PXIM : Impact de la résolution **Spec**trale pour la cartographie des propriétés pérennes des **Sol**s en milieu Méditerranéen » (2014-2015). We are indebted to Yves Blanca (IRD-UMR LISAH Montpellier) for the soil sampling in 2009 over the La Peyne catchment.

References:

Abrams, M., Hook, S.J., 1995. Simulated Aster data for geologic studies. IEEE Transactions on Geoscience and Remote Sensing 33, 692-699.

Atzberger, C., 2002. Soil optical properties – A review. University of Trier, Remote sensing department, Tech. Rep. D-54286, Trier, Germany.

Bartholomeus, H., Mulder, V.L., 2008. The influence of slope on the spectroscopic quantification of soil iron content. in R. Viscarra-Rossel (Ed.), High resolution digital soil sensing and mapping, Sydney, Australia, p. 9.

Bartholomeus, H., Epemab, G., Schaepman, M., 2007. Determining iron content in Mediterranean soils in partly vegetated areas, using spectral reflectance and imaging spectroscopy. International Journal of Applied Earth Observation and Geoinformation 9, 194-203.

Ben-Dor, E., Patkin, K., Banin, A., Karnieli, A., 2002. Mapping of several soil properties using DAIS-7915 hyperspectral scanner data—A case study over clayey soils in Israel. International Journal of Remote Sensing 23, 1043-1062.

Ben-Dor, E., Banin, A., 1995. Near-infrared analysis (NIRA) as a rapid method to simultaneously evaluate several soil properties. Soil Science Society of American Journal 59, 364-372.

Bayer, A., Bachmann, M., Müller, A., Kaufmann, H., 2012. A Comparison of Feature-Based MLR and PLS Regression Techniques for the Prediction of Three Soil Constituents in a Degraded South African Ecosystem. Applied and Environmental Soil Science 2012, Article ID 971252, 20 pages.

Chabrillat, S., Goetz, A.F.H., 2006. Remote sensing of expansive soils - Use of hyperspectral methodology for clay mapping and hazard assessment, in: Al-Rawas, A.A., Goosen, M.F.A., Taylor & Francis (Eds.), Expansive Soils: Recent Advances in Characterization and Treatment, pp. 187-209.

Chang, C.-W., Laird, D.A., Mausbach, M.J., Hurburgh, C.R., 2001. Near-Infrared Reflectance Spectroscopy–Principal Components Regression Analyses of Soil Properties. Soil Science Society of American Journal 65, 480-490.

Chiang, L.H., Pell, R.J., Seasholtz, M. B., 2003. Exploring process data with the use of robust outlier detection algorithms. Journal of Process Control 13, 437-449.

Chong, I. G., Jun, C. H., 2005. Performance of some variable selection methods when multicollinearity is present. Chemometrics and Intelligent Laboratory Systems 78, 103-112.

Clark, R. N., 1999. Spectroscopy of rocks and minerals and principles of spectroscopy. Remote Sensing for the Earth Sciences. Manual of Remote Sensing. Ed. by A.N. Rencz. John Wiley & Sons Ltd, Chichester, UK, pp. 3-58.

Demattê, J.A.M., Campos, R.C., Alves, M.C., Fiorio, P.R., Nanni, M.R., 2004. Visible–NIR reflectance: a new approach on soil evaluation. Geoderma 121, 95-112.

Demattê, J.A.M., Garcia, G.J., 1999. Alteration of soil properties through a weathering sequence as evaluated by spectral reflectance. Soil Science Society of America Journal 63, 327-342.

Dray, S., Dufour, A.B., 2007. The ade4 package: implementing the duality diagram for ecologists. Journal of Statistical Software 22, 1-20.

Gaffey, S. J., 1987. Spectral reflectance of carbonate minerals in the visible and near infrared (0.35–2.55 um): Anhydrous carbonate minerals. Journal of Geophysical Research, 92, 1429-1440.

Gobrecht, A., Roger, J. M., Bellon-Maurel, V., 2013. Major issues of diffuse reflectance NIR spectroscopy in the specific context of soil carbon content estimation: a review. Advances in Agronomy, 123, 145-175.

Haaland, D.M., Thomas, E.V., 1988. Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information. Analytical Chemistry 60, 1193-1202.

Hadoux, X., Gorretta, N., Roger, J.-M., Bendoula, R., Rabatel, G., 2014. Comparison of the efficacy of spectral pre-treatments for wheat and weed discrimination in outdoor conditions. Journal of Computers and Electronics in Agriculture, 108, 242-249.

Harman, H. H., 1976. Modern factor analysis. Chicago University Press, Chicago.

Kemper, T., Sommer, S., 2002. Estimate of heavy metal contamination in soils after a mining accident using reflectance spectroscopy. Environmental Science & Technology, 36(12), 2742-2747.

Knadel, M., Stenberg, B., Deng, F., Thomsen, A. G., Greve, M. H., 2013. Comparing predictive abilities of three visible-near infrared spectrophotometers for soil organic carbon and clay determination. Journal of Near Infrared Spectroscopy, 21(1), 67-80.

Lagacherie, P., Baret, F., Feret, J. -B., MadeiraNetto, J., Robbez-Masson, J.-M., 2008. Estimation of soil clay and calcium carbonate using laboratory, field and airborne hyperspectral measurements. Remote Sensing of Environment 112, 825-835.

Lal, R., 2004. Soil carbon sequestration to mitigate climate change. Geoderma 123(1-2), 1-22.

Le Bissonnais, Y., 1996. Aggregate stability and assessment of soil crustability and erodibility I. Theory and methodology. European Journal of Soil Science 47, 425-437.

Mark, H.L., Tunnell, D., 1985. Qualitative near-infrared reflectance analysis using Mahalanobis distances. Analytical Chemistry 57, 1449-1456.

Mevik, B.-H., Wehrens, R., 2007. The pls Package: Principal Component and Partial Least Squares Regression in R. Journal of Statistical Software 18, 1-24.

Mouazen, A.M., Kuang, B., De Baerdemaeker, J., Ramon, H., 2010. Comparison among principal component, partial least squares and back propagation neural network analyses for accuracy of measurement of selected soil properties with visible and near infrared spectroscopy. Geoderma 158, 23-31.

Mouazen, A., Saeys, W., Xing, J., De Baerdemaeker, J. and Ramon, H., 2005. Near infrared spectroscopy for agricultural materials: an instrument comparison. Journal of Near Infrared Spectroscopy, 13(2), 87-98.

Palacios-Orueta, A., Ustin, S.L., 1998. Remote Sensing of Soil Properties in the Santa Monica Mountains I. Spectral Analysis. Remote Sensing of Environment 65, 170-183.

Pearson, R. K., 2002. Outliers in process modeling and identification. IEEE Transactions on Control Systems Technology 10, 55-63.

Peng, X., Shi, T., Song, A., Gao, W., 2014. Estimating soil organic carbon contents from Visible/Near-infrared spectroscopy using the combination of support vector machine regression and successive projection algorithm. Remote Sensing 6, 2699-2717.

Rinnan, A., Van den Berg, F., Engelsen S. B., 2009. Review of the most common pre-processing techniques for near-infrared spectra. TrAC Trends in Analytical Chemistry, 28(10), 1201-1222.

Roger, J.M., Palagos, B., Bertrand, D., Fernandez-Ahumada, E., 2011. CovSel: variable selection for highly multivariate and multi-response calibration: application to IR spectroscopy. Chemometrics and Intelligent Laboratory Systems 106, 216-223.

Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., Orlowsky, B., Teuling, A. J., 2010. Investigating soil moisture–climate interactions in a changing climate: A review. Earth-Science Reviews 99(3-4), 125-161.

Shepherd, K.D., Walsh, M.G., 2002. Development of Reflectance Spectral Libraries for Characterization of Soil Properties. Soil Science Society of America Journal 66, 988-998.

Shibusawa, S., Imade Anom, S.W., Sato, S., Sasao, A., Hirako, S., 2001. Soil mapping using the real-time soil spectrophotometer. In G. Grenier & S. Blackmore (Eds.), ECPA 2001, Agro Montpellier, Third European Conference on Precision Agriculture 1, 497–508.

Stevens, A., Udelhoven, T., Denis, A., Tychon, B., Lioy, R., Hoffmann, L., van Wesemael, B., 2010. Measuring soil organic carbon in croplands at regional scale using airborne imaging spectroscopy. Geoderma 158, 32-45.

Tilman, D., Cassman, K. G., Matson, P. A., Naylor, R., Polasky, S., 2002. Agricultural sustainability and intensive production practices. Nature 418, 671-677.

Viscarra Rossel, R.A., Chen, C., 2011. Digitally mapping the information content of visible–near infrared spectra of surficial Australian soils. Remote Sensing of Environment 115, 1443-1455.

Viscarra Rossel, R. A., Jeon, Y. S., Odeh, I. O. A., McBratney, A. B., 2008. Using a legacy soil sample to develop a mid-IR spectral library. Soil Research, 46, 1-16.

Viscarra Rossel, R.A., 2007. Robust modelling of soil diffuse reflectance spectra by bagging-partial least squares regression. Journal of Near Infrared Spectroscopy 15, 39-47.

Viscarra Rossel, R.A., Walvoort, D.J.J., McBratney, A.B., Janik, L.J., Skjemstad, J.O., 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma 131, 59-75.

Vohland, M., Ludwig, M., Thiele-Bruhn, S., Ludwig, B., 2014. Determination of soil properties with visible to near- and mid-infrared spectroscopy: Effects of spectral variable selection. Geoderma 223-225, 88-96.

Weidong, L., Baret, F., Xingfa, G., Qingxi, T., Lanfen, Z., Bing, Z., 2002. Relating soil surface moisture to reflectance. Remote Sensing of Environment 81, 238-246.

Wold, S., Sjöström, M., Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems 58, 109-130.

Wold, S., Johansson, E., Cocchi, M., 1993. PLS - partial least squares projections to latent structures. in H. Kubinyi Eds., 3D-QSAR in Drug Design, Theory, Methods, and Applications, ESCOM Science Publishers, Leiden, 523-550.

Wold, S., 1978. Cross-Validatory Estimation of the Number of Components in Factor and Principal Components Models. Technometrics 20, 397-405.