

Chapter 5

The Use of Laboratory Spectroscopy and Optical Remote Sensing for Estimating Soil Properties

Joachim Hill, Thomas Udelhoven, Michael Vohland, and Antoine Stevens

Abstract The success of precision agriculture requires accurate methods for monitoring the state and health of crops. An additional key issue is the availability of accurate and efficient techniques for in-situ determination of soil properties. Reflectance spectroscopy, a technique which can be applied in the laboratory, in the field and from remote observation systems has attracted the attention of scientists in a variety of disciplines. In soil science, this technology as it relates to precision farming is rapidly developing and has triggered new research initiatives. Although a number of studies are available where soil properties have been derived from reflectance spectra the approach involves substantial scaling problems when transferring methods from laboratory spectroscopy to optical sensor systems onboard satellites and aircrafts. The analysis of reflectance images also requires dealing with data having limited signal-to-noise level, being distorted by atmospheric effects and largely affected by bidirectional effects in reflectance distribution. Starting with a short review of the state-of-the-art we present the potential use of reflectance spectroscopy for retrieving useful soil parameters based on several case studies. These studies serve to illustrate the existing limitations for retrieving soil properties over large heterogeneous areas.

1 Introduction

The success of precision agriculture not only depends on accurate methods for monitoring the state and health of crops but also relies on accurate and efficient techniques for in-situ determination of soil properties. Soil parameters are neither static nor homogenous in space and time, however analytical costs are often a limiting factor when attempting to address spatial soil variability especially in large-scale

J. Hill (✉)

Remote Sensing Department, Faculty of Geography/Geosciences, Trier University, Trier D-54286, Germany

e-mail: hillj@uni-trier.de

applications (Plant 2001, Viscarra-Rossel and McBratney 1998). Some applications, such as precision farming, even require the diagnosis of short or medium term changes in the nutrient content of soils. The traditional way to explore in field soil variation is grid-sampling, which is time consuming, labor intensive and lacks spatial exhaustiveness. Schnug et al. (1998) identified the development of actual physico-chemical soil maps as one of the major bottlenecks for continuous soil monitoring at the farm level. When a new technology saves time and results in greater profitability and reduced environmental risk, it will be rapidly adopted by farmers (Schepers and Francis 1998). Thus, the demand on new techniques for soil monitoring is to find a compromise between analytical speed and precision (Shepherd and Walsh 2002).

The first scientists who systematically investigated the relationship between soil spectral information and soil properties were Condit (1970) and then Stoner and Baumgardner (1981). Their soil spectral library quickly became a classical tool for soil scientists and was further used as a fundamental reference source for future studies. Over the past few years, it has been shown that soil spectra across the Visible (VIS, 0.4–0.7 μm), Near Infrared (NIR, 0.7–1.1 μm) and Short-Wave Infrared (SWIR, 1.1–2.5 μm) spectral regions are characterized by significant spectral features that enable quantitative analysis of several soil properties (e.g. Ben-Dor et al. 1999, 2008, Nanni and Demattê 2006, Schnug et al. 1998, Shepherd and Walsh 2002, Viscarra-Rossel et al. 2006). Pure soil minerals and soil organic matter exhibit distinct spectral fingerprints caused by electronic transitions in the VIS and by overtones and the combination modes of functional groups in the NIR and SWIR, which derive from their respective C-H, N-H and O-H fundamental vibrations bonds in the MIR (2.5–25 μm) region (Salisbury 1993). In general, these overtones and combination modes have reflectance peaks that are less clear than those at the fundamental frequencies. A linkage between the two spectral domains can be established using 2D- correlation analysis (Barton and Himmelsbach 1993). Respective models are useful to interpret broad spectral features in the NIR selected by some multivariate statistical calibration model by means of the related primarily absorption features in the middle infrared. A wide range of soil constituents can be identified from the VIS, NIR and SWIR spectral regions under laboratory conditions if advanced analytical techniques such as artificial neural networks and partial least-squares regression analysis are used (e.g. Ben-Dor and Banin 1995a, b, Udelhoven et al. 2003, Viscarra-Rossel 2007).

It was also suggested that methods which are successful for analyzing spectra recorded in the laboratory or in the field, including traditional quantitative approaches that successfully work for laboratory spectrometry of minerals (Clark and Roush 1984), also may be applicable for analyzing the spatially continuous reflectance data provided by multi- or hyperspectral imaging systems. For the emerging discipline of precision agriculture, optical remote sensing and imaging spectrometry in particular were expected to provide soil parameters before and after the growing season, and thus provide farmers with a spatially explicit quantitative overview of the soil properties and phenomena in question. In this way, farmers may be able to control resources such as irrigation, nutrients and cultivation, as well

as obtain better yields per hectare and gain substantial savings through optimized fertilizer, herbicide and pesticide application.

2 Background

Bi-directional soil reflectance in the wavelength range between 0.4 and 2.5 μm is a cumulative property, which derives from the inherent spectral behaviour of the heterogeneous combination of minerals, water, organic matter and other chromophores (Udelhoven et al. 2003). A chromophore – a part of a molecule that causes it to be colored – is a parameter or substance (physical or chemical) that significantly affects the shape and nature of a soil spectrum. A single soil sample usually features a range of chromophores, which may vary with the environmental conditions and soil forming processes.

Soil chromophores can be divided into chemical and physical categories (Ben-Dor et al. 1999). Chemical chromophores are those materials that absorb incident radiation in discrete energy levels (Fig. 5.1). Usually the absorption process appears on a reflectance spectrum as a feature whose position is attributed to specific chemical groups in various structural configurations, overtone, combination modes, and electronic processes. All features in the VIS- NIR- SWIR spectral regions have a physical basis. In soils, three major optically active chemical chromophores can be roughly categorized as follows: (I) minerals, mainly clay, iron oxide, primary minerals feldspar, salt, and hard to dissolve substances such as carbonates, phosphates; (II) fresh and decomposing organic matter; and (III) water in solid, liquid, and gas phases (Fig. 5.1). Minerals, for example, exhibit distinct spectral fingerprints caused by electronic transitions in the VIS and NIR (0.4–1.1 μm) and by overtones and combination modes of OH-, SO-, and CO-groups in the SWIR (1.1–2.5 μm) (Hunt and Salisbury 1970). Often the spectral signals related to a given chromophore overlap with the signals of other chromophores and thereby render the assessment of a specific chromophore difficult.

Physical chromophores are properties that affect the overall spectral region and a particular waveband position, or in other words, do not relate to the chemical functional group. Examples of these are particle size variation and refraction indexes of a material that changes from one illumination condition to another. A comprehensive review of chemical and physical chromophores in soil and elaborating more generally on minerals, some of which are important in the soil environment is given in Irons et al. (1989), Ben-Dor et al. (1999), Clark (1999) and McBratney et al. (2006).

Soil color is one of the most useful attributes for characterization and identification of soil types that can also be derived from most operational multi- and hyperspectral sensor systems (Torrent and Barron 1993). Its relevance is mainly attributed to the fact that soil color can be correlated to important soil properties (Mulders 1987). Traditionally, soil color is measured using the Munsell soil color chart (Munsell Colour Company 1975), which is a useful system for categorizing soil color, but does not lend itself to statistical analysis (Viscarra-Rossel et al. 2006). Therefore, Melville and Atkinson (1985) recommended the use of the CIE-LAB

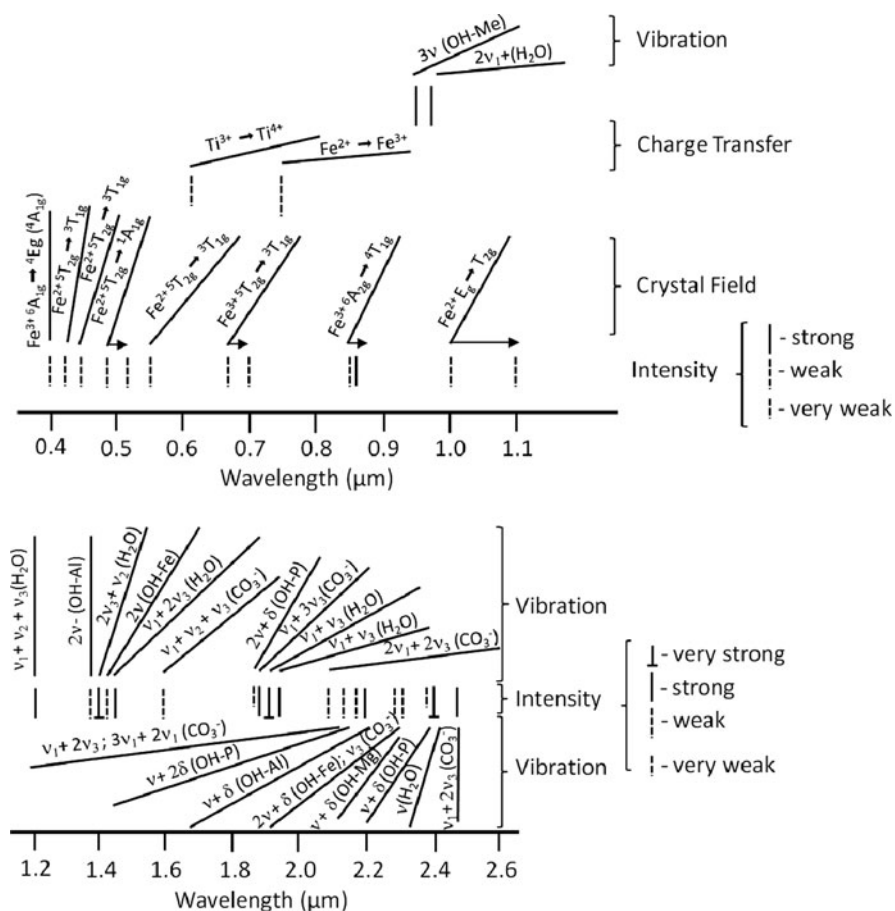


Fig. 5.1 Active groups and mechanisms of chemical soil chromophores. For each possible group, the wavelength range and absorption feature intensity are given (Ben-Dor et al. 1999, modified)

color system instead, and more recently, Jarmer et al. (2009) successfully applied the CIE (Commission Internationale de l'Éclairage) color system from 1931 to assess soil organic carbon concentrations from Landsat TM data on a regional scale level. In the CIE system, color is calculated from reflectance values. In addition the possibility of using statistical analysis of the calculated color values provides another substantial advantage. Satellite data are often characterized by redundant information leading to high correlation among spectral data recorded in various spectral bands, especially those data in the visible domain. Transforming reflectance into the CIE color values leads to a substantial de-correlation of spectral data which is an important advantage for statistical data analysis. Additionally, this transformation additionally allows comparison of soil color derived from different sensors and sensor independent use of developed prediction models.

Major drawbacks for scaling from point (laboratory) to imaging systems are the large size of the pixels that results in a significant mixed pixel problem and the wide spectral band response function and incomplete capture of specific spectral features of the chromophore. While the latter can be efficiently compensated for by using high spectral resolution imaging systems, surface roughness effects cause substantial bidirectional effects which are not straightforwardly controlled. Additional problems arise from the necessity to correct atmospheric distortions in the reflectance signal (Ben-Dor et al. 2009).

3 Retrieval Methods

The general concept for retrieving soil properties requires that a spectrum is processed to provide quantitative information about its chromophores. This can be an index, an equation or a model that is extracted from the spectral information, usually combined with the traditional chemical information. This is usually done by selecting a group of samples, followed by traditional chemical analysis and spectral measurements. Manipulations are done between the two data sets in order to derive a parameter or set of parameters that can describe the property solely from the reflectance readings. Theoretical or empirical models are allowed, whereas validation of each model is essential using external samples (Ben-Dor et al. 2008). This technology is termed Visible and Near-Infrared Spectroscopy (VNIRS) and was adopted from a strategy developed about 40 years ago in food science. In this approach, the reflectance measured from powder or aggregates, across the VIS-NIR- SWIR region, is modeled against constituents determined by wet chemistry methods. After this theoretical chemical model is validated, it can be applied to unknown samples (e.g. Awiti et al. 2008).

In soil science the VNIRS concept has provided promising results for rapid determination of several soil properties. Optically active soil components comprise organic matter (Dalal and Henry 1986, Krishnan et al. 1980, Wilcox et al. 1994, iron oxide minerals, Kosmas et al. 1984, clay and sand content, Al-Abbas et al. 1972, Selige et al. 2006, Waiser et al. 2007), specific surface, hygroscopic moisture, metal and carbonate content (Ben-Dor and Banin 1995a, b). These soil attributes play a decisive role in assessing topsoil characteristics e.g. soil aggregation, aggregate stability and resistance to water and wind erosion (Selige et al. 2006). Recently, He et al. (2007) demonstrated that macronutrients could be predicted via VNIRS for precision farming purposes, although not all are optically active substances. This phenomenon deserves special attention and will be discussed in more detail in one of the next sections. A comprehensive literature review summarizing the VNIRS optical concept and its achievements in soil science can be found in Malley et al. (2004). Another recent review was provided by Nanni and Demattê (2006), who elaborate on the current utilization of this technique for soils, whereas Viscarra-Rossel et al. (2006) provided a detailed list where all soil constituents successfully predictable by VNIRS are presented.

In fact, laboratory VNIRS is accepted as a fast and non-destructive approach (Chang et al. 2001, Shepherd and Walsh 2002), and more recently a number of advanced methods have been suggested to transform reflectance spectra into quantitative estimates of soil constituents. These include multivariate adaptive regression splines (Shepherd and Walsh 2002), radial basis function networks (Fidêncio et al. 2002), and artificial neural networks (Daniel et al. 2003). An important requirement for advanced statistical methods is the ability to handle large sets of collinear predictor variables and to deal with noisy patterns.

3.1 Artificial Neural Networks

The application of artificial feed-forward neural networks (ANNs) is one of the standard methods in spectroscopic applications (Udelhoven and Schütt 2000). A three-layer ANN represents a universal approximator able to fit any continuous function, linear or non-linear, between independent and dependent variables to a pre-defined arbitrary degree of accuracy. A major drawback of ANNs is that they appear to be black boxes due to their high degree of flexibility and the variety of learning parameters and network architectures. ANNs require a learning function to adjust all the weights and biases of a given neural network. There exists a variety of different training algorithms for feed-forward ANNs, including gradient descent methods, conjugate gradient methods, the Levenberg-Marquardt algorithm, to mention only a few. A detailed mathematical description can be found, for example, in Bishop (2005).

Udelhoven and Schütt (2000) tested several variants of feed-forward ANNs for chemical characterization of sediments based on reflectance measurements in a laboratory approach. Ten chemical properties including inorganic carbon, Fe, S, Al, Si, Ca, K, and Mg from 214 samples from various drilling locations all over the central part of the Iberian Peninsula were simultaneously estimated using one ANN model. They concluded that the combined methodology of diffuse reflectance spectroscopy evaluated with a trained and representative neural network can be applied as a rapid and cost-effective screening method to characterize solid samples provided that a representative set of analytical data for the network training is available. A similar conclusion has been drawn by Kemper and Sommer (2002) who used an ANN to predict heavy metals in soils contaminated by mining residuals using reflectance spectroscopy.

3.2 Partial Least Squares Modeling (PLSR, PLSR Combined with a Genetic Algorithm)

Partial least squares regression (PLSR) is an extension of the multiple linear regression and principal component regression models. PLSR projects the data into a low-dimensional space (i.e. a set of orthogonal variables, called latent variables).

It maximizes the covariance between the spectral matrix (X) and chemical concentration matrix (Y) by accomplishing eigendecomposition of both matrices (Otto 1998, Wold et al. 2001). The objective is to model X in such a way that the information in Y can be predicted as precisely as possible. The first latent variable, which is extracted from the matrix X , explains a maximum of the variance of matrix Y . The second latent variable describes a maximum of the residual variance, which has not been described by the first latent variable, and so on. The optimum number d of latent variables to be used in the analysis is determined by comparing the root mean square errors of cross-validation ($RMSE_{cv}$) of the predictions with different values of d . Alternatively, a validation data set can be used to determine an appropriate number of latent variables.

In the following, one example for using PLSR to estimate soil clay content from spectroradiometric measurements is briefly documented. In total 64 soil samples with prevailing loamy sand texture were collected in a floodplain in Central Europe. An ASD FieldSpec II Pro FR instrument (Analytical Spectral Devices, Boulder, USA) was used for the spectral readings of these samples after air drying and grinding. Measured reflectance data were resampled to 10 nm resolution over the 0.4–2.5 μm wavelength range (211 spectral predictor variables). In the PLS approach an optimum number of seven latent variables was found. For this model, the influence of the original spectral variables is reflected by the PLS regression coefficients. The coefficient profile (Fig. 5.2b) exhibits several peaks throughout the complete wavelength range, and thus does not represent the intrinsic spectral features of clay minerals e.g. illite with two strong absorption bands at 2.2 and 2.34 μm . The PLS model provides estimates with an r^2_{cv} value of 0.68 (Fig. 5.2a). $RMSE_{cv}$ amounts to 43.3 g kg^{-1} (relative $RMSE_{cv} = 0.33$), and the RPD (ratio of standard deviation of measured samples to $RMSE_{cv}$) is 1.77. According to the guideline of Malley et al. (2004) these results may be categorized as moderately useful for screening purposes, such as distinguishing low, medium and high values.

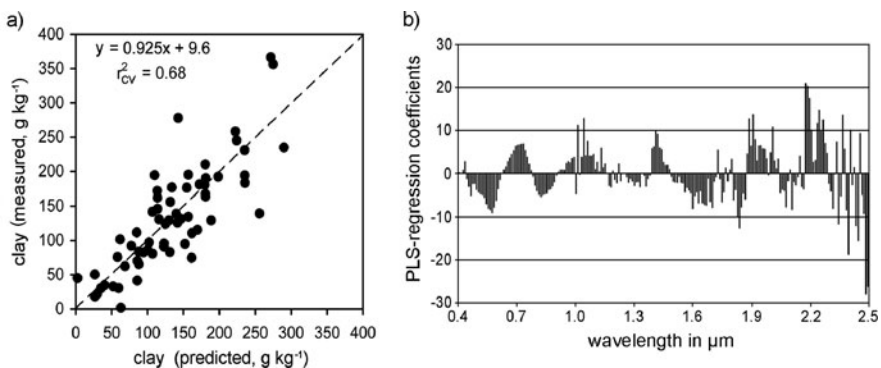


Fig. 5.2 Partial least squares regression (PLSR) for calibrating soil clay contents: Predicted vs. measured values (a) and PLS regression coefficients (b)

An important factor that contributes to the performance and robustness of a statistical prediction model is its parsimony. The parsimony of PLSR is to use a limited number of latent variables for the prediction of a dependent variable. Nevertheless, the number and prediction quality of the relevant latent variables strongly depends on the spectral information that is used to calculate these factors. Therefore improved prediction by PLSR may be achieved by applying some form of variable selection to optimize the quality of the latent variables.

With a large number of features it is not feasible to test all possible subsets for an optimal calibration model. Several techniques can be employed to identify a near optimal solution, among which genetic algorithms allow an efficient search in high-dimensional and complex response surfaces. The overall goal is to identify the most informative variables that allow an improved predictive capacity of the calibration model, or at least a simplified model with less variables and without losing prediction accuracy (Leardi and González 1998, Yoshido et al. 2001).

The genetic algorithm used in our example mainly follows the principles described by Leardi and González (1998). Each chromosome of the initial population is composed of 211 genes (corresponding to the 211 original spectral predictors), each gene being formed by a single bit or binary coding, where each spectral variable can be switched on or off. Chromosomes with an above-average fitness (fitness criterion: cross-validated variance explained by the PLS-regression) are selected as parents. Offsprings are obtained by reproduction (cross-over method) and mutation, and the responses of the new chromosomes are evaluated with the decision to be included in the population or to be discarded. At the end of each run each with 200 evaluations, the selected variables of the fittest chromosome are identified. Selection frequencies at the end of all runs (100 runs per cycle, and then ten repetitions of the complete cycle), decide on the most predictive variables that are accepted for the final PLS calibration (Fig. 5.3).

In this case, only 29 of the 211 original variables are selected. Based on now 6 latent variables, the PLS regression model provides estimates with an r^2_{cv} value

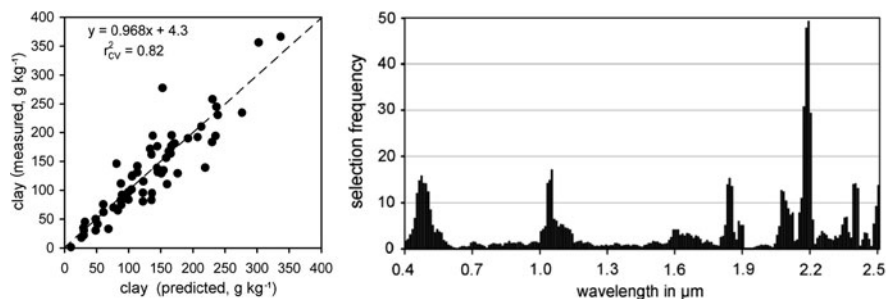


Fig. 5.3 Partial least squares regression with genetic algorithm (GA-PLSR): Predicted versus measured values; based on the averaged frequency of selection per 100 runs (10 repetitions), 29 spectral variables are selected by GA

of 0.82 (Fig. 5.3), and a RMSE_{cv} of 32.0 g kg^{-1} (relative $\text{RMSE}_{\text{cv}} = 0.24$). In comparison to PLS without feature selection the RPD value of 2.40 proves a considerable improvement of GA-PLS up to useful levels for quantification purposes (Malley et al. 2004).

3.3 Support Vector Machine Regression

Support vector machine regression (SVM-R) represents a different model class compared with PLSR techniques since it is based on statistical learning theory (Vapnik 1995). SVM-R has recently been successfully applied for the retrieval of soil organic carbon in Luxembourg based on airborne AHS imaging data (Stevens et al. 2010). The most valuable properties of SVMs are their ability to handle large input spaces efficiently, to deal with noisy patterns and multimodal class distributions, and their restriction to only a subset of training data in order to fit a non-linear function. The SVM-R methodology is described in detail in Schölkopf and Smola (2002). In principle an input vector X is mapped from the input domain into a higher dimensional feature space via a kernel function, where data are spread out in a way that facilitates the finding of an interpolation function (Vapnik 1995). This function is identified by fitting a tube with radius ε to the training data using boundary samples, the so-called support vectors (SV). The optimization problem is solved using Quadratic Programming (QP) techniques (Schölkopf and Smola 2002). This requires fixing a free regularization parameter C beforehand that confines the influence of critical training patterns. As kernel the Gaussian radial basis function (RBF) is often selected due to computational convenience. The RBF kernel requires only selecting one free parameter (σ) beforehand that controls the smoothness properties of the interpolating function.

3.4 Penalized-Spline Signal Regression (PSR)

Penalized-spline Signal Regression (PSR) is a novel technique that has been developed by Marx and Eilers (1999, 2002); PSR is – like PLSR – able to solve a multivariate calibration problem in which the predictors are highly correlated and their number exceeds the number of observations. The main difference between the PLSR and the PSR is that in the former the order of the predictor variables i.e. wavelengths in spectrometry, does not influence the model, whereas PSR forces the coefficient of the regression to vary smoothly across the wavelengths. This is attained by projecting the coefficients onto a set of smooth functions (B-splines). There are several PSR parameters that must be fixed beforehand, including the degree of B-splines and the number of intervals between knots, the point where B-splines join. Stevens et al. (2010) found PSR to be superior to PLSR in the estimation of soil organic carbon from airborne hyperspectral AHS-160 data in an area in Luxembourg.

4 Applications

One soil parameter that is of special relevance in precision farming applications is soil organic carbon (SOC), since it plays an important role with respect to chemical and physical processes in the soil environment (Ben-Dor et al. 1999). Together with clay content and composition, it has a major effect on major soil properties such as the stability of soil aggregates and water retention (Stevens et al. 2010). Furthermore, soil organic carbon strongly influences soil fertility, plant nutrient supply, microbial activity and soil physical properties (Wilcox et al. 1994). The following examples focus on the assessment of organic carbon at different scales.

4.1 Scale Dependencies in the Assessment of Chemical Soil Constituents

Soil monitoring using VNIRS has been applied on different scales, ranging from laboratory measurements, field approaches to airborne and satellite hyperspectral imaging devices. At a field scale VNIRS has been used to estimate chemical soil constituents with portable spectrometers (Kooistra et al. 2003, Odlare et al. 2005, Udelhoven et al. 2003). At the regional scale soil properties were assessable by multispectral satellite systems (Hill and Schütt 2000, Jarmer et al. 2009), airborne imaging spectrometry (Ben-Dor et al. 2002, Stevens et al. 2010, 2006), and recently also from the hyperspectral satellite platform HYPERION (Gomez et al. 2008).

Udelhoven et al. (2003) evaluated soil chemical properties from different locations in the Trier region (Rhineland-Palatinate, Germany) under field and laboratory conditions using a portable spectrometer and PLSR. Generally, laboratory spectrometry using air dried and sieved samples performed better than field spectrometry. This was probably due to strong interferences of soil surface properties such as moisture content, roughness and crusting. In a plot experiment they investigated the accuracy in the retrieval of chemical soil parameters such as Ca, Mg, Fe, Mn, K and SOC. In Fig. 5.4 a spline interpolation is shown for organic and inorganic carbon for both measured and estimated contents from the data set. The prediction accuracy of the SOC-model with the best performance corresponded to a mean square error for cross-validation ($RMSE_{CV}$) of 0.14, and to a coefficient of determination (r^2_{cv}) of 0.6, respectively. Although statistically significant, Fig. 5.4 illustrates that at this level of accuracy measured spatial concentration pattern in the plot are not sustained. In contrast, PLS estimations of inorganic carbon were more accurate ($r^2_{cv} = 0.93$), resulting in a much better representation of the inorganic carbon values in the plot. This demonstrates that a statistically significant relation between dependent and spectral variables does not guarantee that the spatial patterns in concentration can be well reproduced. This is due to the fact that the spatial interpolation also fits the prediction errors of the target variable.

Airborne imaging spectrometry has an even greater potential to overcome the restrictions of ground based or laboratory spectroscopy as spatial interpolation is circumvented and upscaling over large areas is possible. These systems provide high

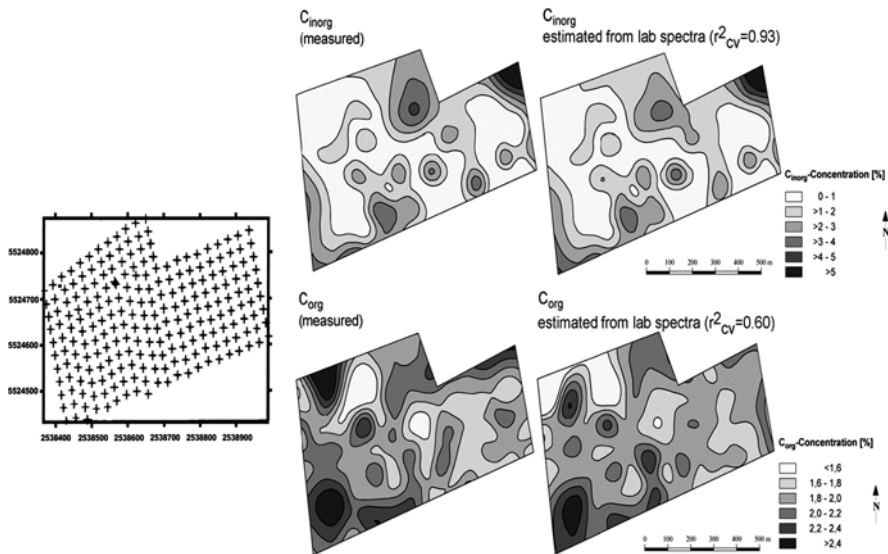


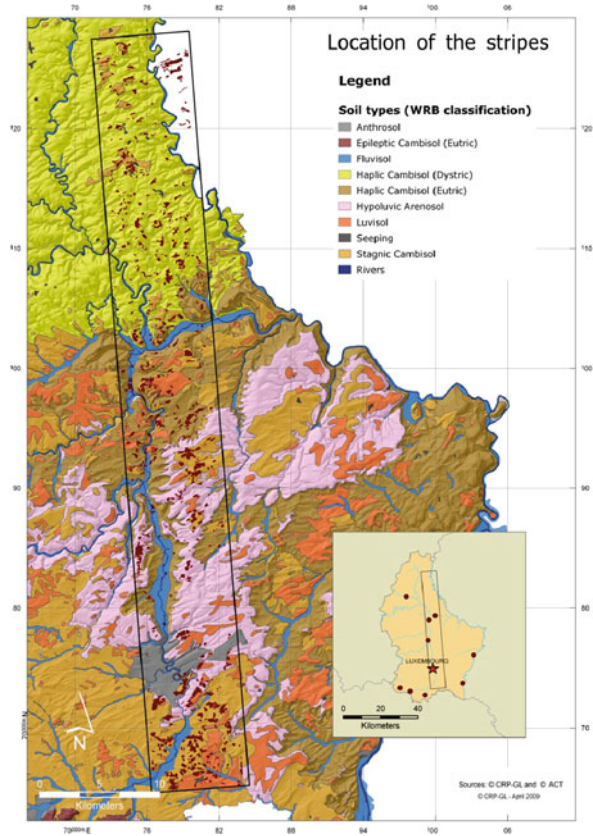
Fig. 5.4 Spline interpolation of measured and estimated (cross-validation) inorganic carbon and organic carbon contents at field scale (plot ‘Dietrichskreuz’, Helenenberg, Rhineland-Palatinate, SW Germany; Udelhoven et al. 2003)

spatial and spectral resolution along with flexible temporal resolution which is ideally suited for soil monitoring in the context of precision agricultural applications. Major constraints include atmospheric absorptions interfering with the spectral measure, spatial variation in surface properties and a lower signal-to-noise ratio, which might result in an incompatibility between field and airborne spectroscopic measurements (Ben-Dor et al. 2008, Chappell et al. 2005, Stevens et al. 2010). Hill and Schütt (2000) could demonstrate that meaningful spatial patterns of soil organic matter which exhibited a positive correlation to crop productivity could be derived from multispectral satellite systems. Gomez et al. (2008) estimated SOC from reflectance data from vertisols in Australia using HYPERION data and a portable field spectrometer and partial least squares regression (PLSR).

Ben-Dor et al. (2002) could explain 83.3% of the variability of SOC using DAIS-7915 airborne data (400–2,500 nm) of clay soils in Israel. Uno et al. (2005) and Stevens et al. (2006) found between 74 and 85% common variability of SOC and spectrometry data from the CASI airborne hyperspectral sensor (400–950 nm). Selige et al. (2006) achieved slightly better results ($r^2 = 0.9$) with the HyMap sensor (420–2,480 nm). Less satisfactory prediction models of SOC were obtained by Bajwa and Tian (2005) with the RDACS/H-3 sensor (471–828 nm; $r^2 = 0.66$) and De Tar et al. (2008) with the AVNIR sensor (429–1,010 nm; $r^2 = 0.48$).

Whereas the majority of these studies addressed comparably small areas or homogeneous soil types, Stevens et al. (2010) analyzed hyperspectral images acquired with the AHS-160 sensor to predict variation in SOC content in Luxembourg (Fig. 5.5), a country which is covered by different soil types and

Fig. 5.5 Locations of bare soils within five parallel flight stripes using the AHS-160 sensor in Luxembourg



a large variation in SOC contents. Reflectance data were related to surface SOC contents of bare croplands ($n = 325$) by means of 3 different multivariate calibration techniques: PLSR, PSR and least square support vector machine (SVM-R). Their performance was tested under different combinations of global and local calibrations stratified according to agro-geological zone. Figure 5.6 illustrates the results for the global calibration and for four separate sub-models obtained for the different agro-geological zones and for each statistical model using an internal validation data set. The measure of accuracy is related to the root mean squared error of prediction (RMSEP).

Nevertheless, a substantial spread in observed versus predicted SOC values above 30 g C kg^{-1} indicates a higher degree of variability in the reflectance data. This is not solely attributed to the organic carbon content, but to other soil chromophores such as soil moisture and ferrous oxides that differ between the agro-geological regions. These chromophores disturb the global correlation with SOC.

Figure 5.6, right, shows the results obtained for the local calibrations over each agro-geological subset separately. This strategy allowed considerable improvement in the accuracy of the models. In addition the problem of non-linearity of PLSR and

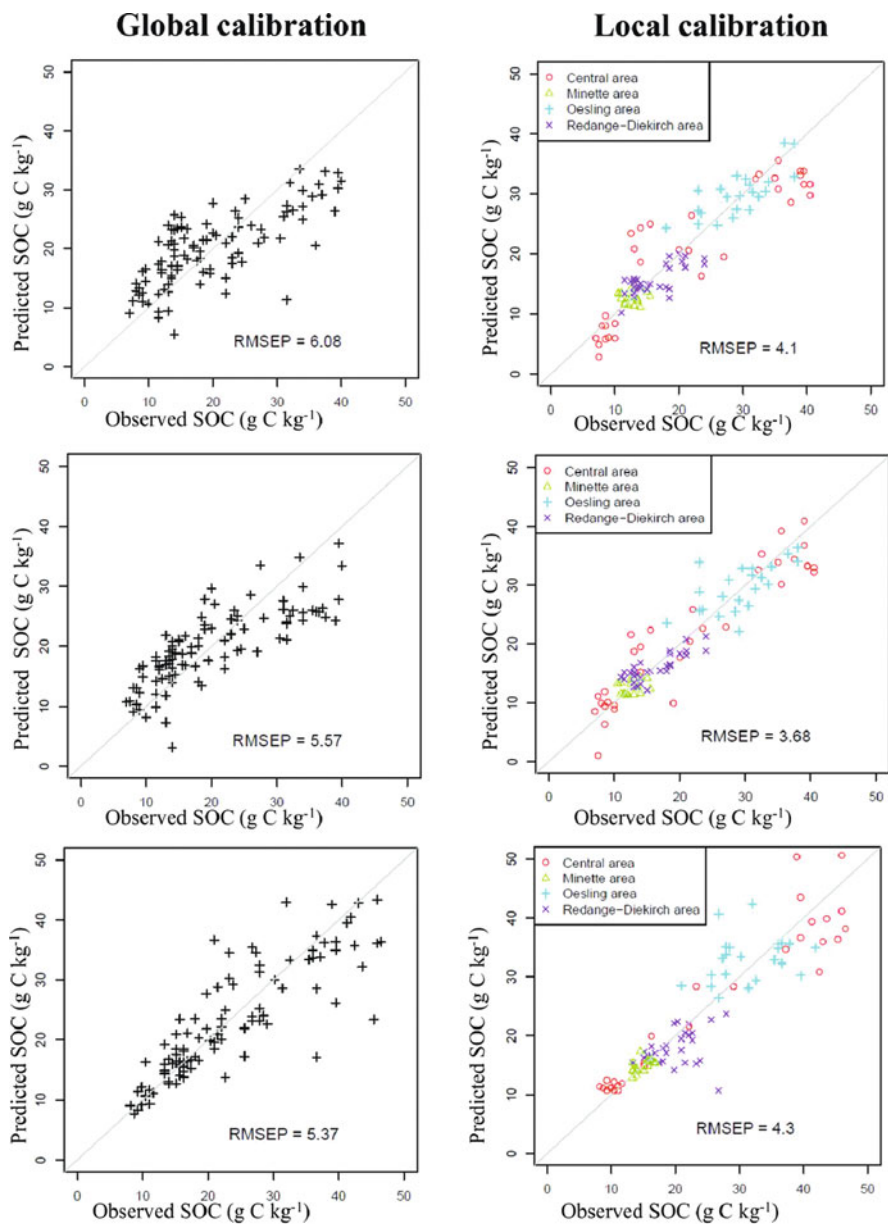


Fig. 5.6 Plots of measured versus predicted SOC as obtained by PLSR (*top*), PSR (*middle*) and SVR (*bottom*) using AHS-160 data validation data set

PSR models at high SOC content no longer occurs when applying local calibrations for each of the agro-geological regions. Other attributes used to stratify the data set e.g. soil type and image number also led to an improvement when compared to the global models (Stevens et al. 2010). For the stratified samples PSR demonstrated the best ability to predict SOC content in the stratified samples. These findings are in line with the conclusions of Marx and Eilers (2002) who showed that PSR offers greater stability in predictions under changing experimental conditions compared to PLSR.

4.2 Estimation of Optically Featureless Soil Components

A number of studies suggest that it is possible to estimate even optically non-active chemical soil properties with featureless spectra. This is possible in that these elements are bonded to active soil components such as Fe oxides, organic matter and clay and these bonds provide a major predictive mechanism (Kemper and Sommer 2002, Kooistra et al. 2003, Vohland et al. 2009, Wu et al. 2007). In statistics this phenomenon is known as spurious correlation. Martínes-Carreras et al. (2010) predicted different chemical properties of suspended sediments from the small catchment of the Wollefsbach, (4.4 km²), a sub-catchment of the Atert River catchment located in the NW of Luxembourg, from spectroradiometer data through PLSR. Apart from major suspended particle components such as organic carbon, calcium and iron oxides they were able to predict trace minerals, such as Li, Sc, Cr, Ni and Cs and even rare earth elements like La, Ce, Pr, Nd, Sm, Eu and Dy. This is due to indirect correlation with spectra caused by optically active background variables, in particular iron oxides, organic matter and clay and clearly reflects the mineralogical nature of the investigated catchments. Spurious correlation can be detected by statistical techniques. Wu et al. (2007) used Principal Component Analysis (PCA) with varimax rotation to clarify the relationships between different chemical tracers. The conception is that highly correlated variables might be estimated from reflectance data if at least one of these properties is optically active.

Another approach is to analyze patterns in the correlation spectra for each soil constituent of interest. The value of the correlation coefficient at a single wavelength describes the univariate importance of a wavelength for the prediction of the given constituent. Another possibility is to analyze factor loadings or regression coefficients in PLSR or related statistical calibration models (Malley and Williams 1997, Vohland et al. 2009). An example is given in Fig. 5.7 which shows the correlation spectra (correlograms) for the sediment properties from the study of Udelhoven and Schütt (2000). The two correlograms are grouped according to the correlation structure of the considered chemical compounds. The first class consists of Fe, K, Al and Si while the second of C, LOI, Ca and calcite. Within both groups the correlation structure is to some extent redundant, however only Fe, C (in carbonates) and calcite have direct optical features. The prediction of the remaining properties is based on spurious correlation. It can be assumed that these patterns cannot be attributed to only one dominant chemical characteristic in these groups, but to the

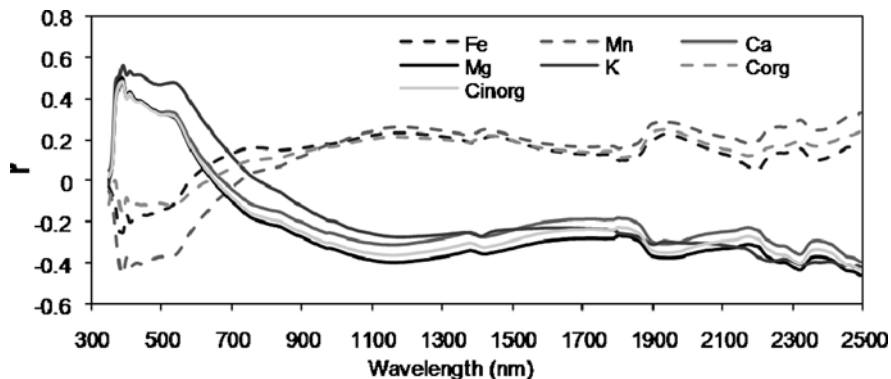


Fig. 5.7 Correlograms of selected soil properties in a plot experiment (Udelhoven and Schütt 2000)

collective occurrence of several optically active minerals in the samples from a specific petrography. The explanatory optically active components in the first groups are iron oxides and clay minerals.

This example demonstrates that spurious correlation of chemical soil constituents with the soil reflectance pattern is often caused by the soil's geological and mineralogical nature. By no means can statistical regression models based on spurious correlation be transferred to other regions beyond the study area with different underlying geology or to soil parameters that are highly variable in space and time. This excludes VNIRS as a diagnosis tool of short or medium term changes of the soil's nutrient status e.g. potassium or phosphorus.

5 Conclusions

Arable soils are important resources which should be preserved for present and future human needs by sustainable agriculture. The monitoring of environmental processes on large-scales requires up-to-date maps of physico-chemical soil properties. Soil reflectance is determined by soil chromophores that are basically determined by soil chemical composition and to soil albedo that is related to soil physical characteristics. Consequently, soil mapping in the context of precision agriculture applications may largely benefit from imaging and non-imaging diffuse reflectance spectrometry, which has the potential to overcome the current problems of high costs, labor and time. Soil mapping should aim to represent the temporal and spatial variability of soil properties at different scales using in-situ and laboratory methodologies. There are still some restrictions that hinder the transfer of visible and infrared spectroscopic methods from the laboratory to the field scale. Two of the most disturbing factors for in-situ spectral measurements are soil roughness and moisture content that must be taken into account to enable models to work accurately under a variety of conditions.

The spectral prediction mechanism may be a direct one that is based on diagnostic spectral fingerprints of the studied parameter, or indirect when variables of interest are without spectral features but correlated with spectrally active soil components. Airborne imaging spectrometry has the advantage that spatial interpolation is circumvented and upscaling over large areas is possible. The spectral range of current and forthcoming airborne (e.g. HyMap, APEX, ARES, AISA, HySpex) and satellite imaging spectrometers (HYPERION, EnMAP) is excellent for detecting electronic transitions in minerals e.g., iron oxides, Fe²⁺ bearing minerals, and vibrational absorptions due to lighter elements e.g. OH, SO₄, CO₃, CH, etc.. Therefore, the spatial distribution of OH-bearing minerals, carbonates, sulfates and organics can be mapped on bare soil surfaces. A substantial restriction in regional soil monitoring is that the related statistical models to retrieve soil parameters from spectroscopic data are applicable only to geological homogeneous areas or 'soilscapes'. Otherwise a stratification of the soil samples according to geological conditions and the calibration of respective sub-models are suggested due to the correlation of the strata with important chromophores like soil moisture or ferrous oxide content.

Despite many encouraging results, the exploitation of chemical soil property maps derived from imaging spectrometry data should still be considered with some caution. In particular, a post-validation over fields not covered by the existing calibration/validation sets would be necessary to assess the actual accuracy of the statistical models. This is important particularly in cases of the prediction of featureless soil properties. Another critical issue is the representativeness of statistical models in case of varying surface and illumination conditions.

Beyond the increasing number of airborne systems, spaceborne hyperspectral imagers provide a viable coverage for large-scale studies of bare soils for future operational applications. HYPERION, a hyperspectral imaging instrument on the EO-1 platform is currently the only operational spaceborne hyperspectral sensor. The sensor measures the radiance with 242 continuous spectral bands, ranging from 356 to 2,577 nm with approximately 10 nm of spectral resolution and 30 m of spatial resolution. HYPERION collects image data for an area of about 7.7 km in the across-track direction and 42 km in the along-track direction.

In 2014 the German Hyperspectral Environmental Mapping and Analysis Program (EnMAP) satellite will be launched. This sensor is expected to provide a higher signal-to-noise ratio than that of HYPERION. However it will not solve the problem of the rather limited spatial resolution of 30 m. Thus, the recorded reflectance is often the mixed result of several surface components. This makes it necessary, especially in heterogeneous landscapes, to apply spectral unmixing techniques to isolate the reflectance signals of soils from disturbing influences especially from green vegetation and crop residues.

Acknowledgements Soil data presented in Section 3.2 were collected and analyzed in a joint project of the University of Applied Sciences of Osnabrück (Department of Agricultural Sciences and Landscape Architecture, Prof. Dr. H.-C. Fründ) and the University of Trier (Remote Sensing Department). This is gratefully acknowledged by the authors.

References

- Al-Abbas AH, Swain PH, Baumgardner MF (1972) Relating organic matter and clay content to the multispectral radiance of soils. *Soil Sci* 6:477–485
- Awiti AO, Walsh MG, Shepherd KD, Kinyamario J (2008) Soil condition classification using infrared spectroscopy: a proposition for assessment of soil condition along a tropical forest-cropland chronosequence. *Geoderma* 143:73–84
- Bajwa SG, Tian LF (2005) Soil fertility characterization in agricultural fields using hyperspectral remote sensing. *Trans Am Soc Agr Eng* 48:2399–2406
- Barton FE, Himmelsbach DS (1993) Two-dimensional vibrational spectroscopy II: correlation of the absorptions of lignins in the mid- and near-infrared. *Appl Spectr* 47:1920–1925
- Ben-Dor E, Banin A (1995a) Near infrared analysis (NIRA) as a simultaneously method to evaluate spectral featureless constituents in soils. *Soil Sci* 59:259–269
- Ben-Dor E, Banin A (1995b) Quantitative analysis of convolved TM spectra of soils in the visible, near infrared and short-wave infrared spectral regions (0.4–2.5 μm). *Int J Rem Sens* 18:3509–3528
- Ben-Dor E, Chabrilat S, Demattê JAM et al (2009) Using imaging spectroscopy to study soil properties. *Rem Sens Environ* 113 (suppl 1):S38–S55
- Ben-Dor E, Irons JA, Epema A (1999). Soil spectroscopy. In: Rencz A (ed) *Manual of remote sensing*. Wiley, New York, pp 111–188
- 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. *Int J Remote Sens* 23:1043–1062
- Ben-Dor E, Taylor RG, Hill J et al (2008) Imaging spectrometry for soil. In: Sparks DL (ed) *Advances in agronomy*. Elsevier/Academic Press, Amsterdam, pp 321–392
- Bishop CM (2005) *Neural networks for pattern recognition*. Oxford University Press, Oxford
- Bowers S, Hanks RJ (1965) Reflectance of radiant energy from soils. *Soil Sci* 100:130–138
- Chang CW, Laird DA, Mausbach MJ, Hurburgh CR (2001) Near-infrared reflectance spectroscopy – principle components regression analyses of soil properties. *Soil Sci Soc Am J* 65:480–490
- Chappell A, Zobeck TM, Brunner G (2005) Using on-nadir spectral reflectance to detect soil surface changes induced by simulated rainfall and wind tunnel abrasion. *Earth Sur Process Landforms* 30:489–511
- Clark RN (1999) Spectroscopy of rocks and minerals, and principles of spectroscopy. In: Rencz AN (ed) *Remote sensing for the earth sciences*. Wiley, New York, pp 3–58
- Clark RN, Roush TL (1984) Reflectance spectroscopy: quantitative analysis techniques for remote sensing applications. *J Geophys Res* 89:6329–6340
- Company MC (1975) Munsell soil color charts. Macbeth Division of Kollmorgen, New Windsor
- Condit HR (1970) The spectral reflectance of American soils. *Photogram Eng Remote Sens* 36:955–966
- Dalal RC, Henry RJ (1986) Simultaneous determination of moisture, organic carbon and total nitrogen by near infrared reflectance spectroscopy. *Soil Sci Soc Am J* 50:120–123
- Daniel K, Tripathi N, Honda K (2003) Artificial neural network analysis of laboratory and in situ spectra for the estimation of macronutrients in soils of Lop Buri (Thailand). *Aus J Soil Res* 41:47–59
- De Tar WR, Chesson JH, Penner JV, Ojala JC (2008) Detection of soil properties with airborne hyperspectral measurements of bare fields. *Trans Am Soc Agric Biol Eng* 51:463–470
- Fidêncio P, Poppi R, De Andrade J (2002) Determination of organic matter in soils using radial basis function networks and near infrared spectroscopy. *Anal Chim Acta* 45:125–134
- Gomez C, Rossel RAV, McBratney AB (2008) Soil organic carbon prediction by hyperspectral remote sensing and field vis-NIR spectroscopy: An Australian case study. *Geoderma* 146:403–411

- He Y, Huang M, Garcia A et al (2007) Prediction of soil macronutrients content using near-infrared spectroscopy. *Comp Electron Agric* 58:144–153
- Hill J, Schütt B (2000) Mapping complex patterns of erosion and stability in dry Mediterranean ecosystems. *Remote Sens Environ* 74:557–569
- Hunt GR, Salisbury JW (1970) Visible and near-infrared spectra of minerals and rocks: I. Silicate minerals. *Mod Geol* 1:283–300
- Irons JR, Weismiller RA, Petersen GW (1989) Soil reflectance. In: Asrar G (ed) *Theory and application of optical remote sensing*. Wiley, New York, pp 66–106
- Jarner T, Hill J, Lavée H, Sarah P (2009) Mapping topsoil organic carbon in non-agricultural semi-arid and arid ecosystems of Israel. *Photogram Eng Remote Sens* 76:85–94
- Kemper T, Sommer S (2002) Estimate of heavy metal contamination in soils after a mining accident using reflectance spectroscopy. *Environ Sci Technol* 36:2742–2747
- Kooistra L, Wanders J, Epema GF et al (2003) The potential of field spectroscopy for the assessment of sediment properties in river floodplains. *Anal Chim Acta* 484:189–200
- Kosmas CS, Curi N, Bryant RB, Franzmeier DP (1984) Free iron oxide determination in Mediterranean soils using diffuse reflectance spectroscopy. *Soil Sci Soc Am J* 48:401–405
- Krishnan P, Alexander DJ, Butler B, Hummel JW (1980) Reflectance technique for predicting soil organic matter. *Soil Sci Soc Am J* 44:1282–1285
- Learidi R, González AL (1998). Genetic algorithms applied to feature selection in PLS regression: how and when to use them. *Chemom Intel Lab Sys* 41:195–207
- Malley DF, Williams PC (1997) Use of near-infrared reflectance spectroscopy in prediction of heavy metals in freshwater sediment by their association with organic matter. *Environ Sci Technol* 31:3461–3467
- Malley DP, Martin D, Ben-Dor E (2004) Application in analysis of soils. In: Craig R, Windham R, Workman R (eds) *Near infrared spectroscopy in agriculture*. Soil Science Society of America Inc. Publishers, Madison, pp 729–784
- Martínez-Carreras N, Krein A, Udelhoven T et al (2010) A rapid spectral reflectance-based fingerprinting approach for documenting suspended sediment sources during storm runoff events. *J Soil Sediment* 10:400–413
- Marx BD, Eilers PHC (1999). Generalized linear regression on sampled signals and curves: a P-spline approach. *Technometrics* 41:1–13
- Marx BD, Eilers PHC (2002) Multivariate calibration stability: a comparison of methods. *J Chemomet* 16:129–140
- McBratney AB, Minasny B, Viscara-Rossel R (2006) Spectral soil analysis and inference system: a powerful combination for solving soil data crisis. *Geoderma* 136:272–278
- Melville MD, Atkinson G (1985) Soil color: its measurement and its designation in models of uniform color space. *J Soil Sci* 36:495–512
- Mulders MA (1987) *Remote sensing in soil science*. Elsevier, Amsterdam
- Nanni MR, Demattê JAM (2006) Spectral reflectance methodology in comparison to traditional soil analysis. *Soil Sci Soc Am J* 70:393–407
- Odlare M, Svensson K, Pell M (2005) Near infrared reflectance spectroscopy for assessment of spatial soil variation in an agricultural field. *Geoderma* 126:193–202
- Otto O (1998) *Statistics and computer application in analytical chemistry*. Wiley, Weinheim
- Plant RE (2001) Site-specific management. The application of information technology to crop production. *Comp Electron Agric* 30:9–29
- Salisbury JW (1993) Mid-infrared spectroscopy: laboratory data. In: Pieters CM, Englert PAJ (eds) *Remote geochemical analysis: elemental and mineralogical composition*. Cambridge University Press, Cambridge, pp 79–98
- Schepers JS, Francis DD (1998) Precision agriculture –What’s in our future. *Comm Soil Sci Plant Anal* 29:1463–1469
- Schnug E, Panten K, Haneklaus S (1998) Sampling and nutrient recommendations – The future. *Comm Soil Sci Plant Anal* 29:1455–1462

- Schölkopf B, Smola A (2002) *Learning with Kernels*. MIT Press, Massachusetts Institute of Technology, Cambridge
- Selige T, Böhner J, Schmidhalter U (2006) High resolution topsoil mapping using hyperspectral image and field data in multivariate regression modeling procedures. *Geoderma* 136: 235–244
- Shepherd KD, Walsh MG (2002) Development of reflectance spectra libraries for characterization of soil properties. *Soil Sci Soc Am J* 66:988–998
- Stevens A, Udelhoven T, Denis A et al (2010) Measuring soil organic carbon in croplands at regional scale using imaging spectroscopy. *Geoderma* in press, doi:10.1016/j.geoderma.2009.11.032
- Stevens A, van Wesemael B, Vandenschrick G et al (2006) Detection of carbon stock change in agricultural soils using spectroscopic techniques. *Soil Sci Soc Am J* 70:844–850
- Stoner ER, Baumgardner MF (1981) Characteristic variations in reflectance of surface soils. *Soil Sci Soc Am J* 45:1161–1165
- Torrent J, Barron V (1993) Laboratory measurement of soil color: theory and practice. In: Bigham JM, Ciolkosz EJ (eds) *Soil colour*. Soil Science Society of America, Madison, pp 21–33
- Udelhoven T, Emmerling C, Jarmer T (2003) Quantitative analysis of soil chemical properties with diffuse reflectance spectrometry and partial least-square regression: a feasibility study. *Plant Soil* 251:319–329
- Udelhoven T, Schütt B (2000) Capability of feed-forward neural networks for a chemical evaluation of sediments with diffuse reflectance spectroscopy. *Chemomet Intel Lab Sys* 51:9–22
- Uno Y, Prasher SO, Patel RM et al (2005) Development of field-scale soil organic matter content estimation models in Eastern Canada using airborne hyperspectral imagery. *Can Biosys Eng* 47:9–14
- Vapnik VN (1995) *The nature of statistical learning theory*. Springer, New York
- Viscarra-Rossel RA (2007) *ParleS*: Software for chemometric analysis of spectroscopy data. *Chemomet Intel Lab Sys* 90:72–83
- Viscarra-Rossel RA, McBratney AB (1998) Soil chemical analytical accuracy and costs. Implications from precision agriculture. *Aus J Exp Agric* 7:765–775
- Viscarra-Rossel RA, Walvoort DJJ, McBratney et al (2006) Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma*, 131:59–75
- Vohland M, Bossung C, Fründ HC (2009) A spectroscopic approach to quantify trace heavy metal contents in contaminated floodplain soils via spectrally active soil components. *J Plant Nutr Soil Sci* 172:201–209
- Waiser TH, Morgan CLS, Brown DJ, Hallmark CT (2007) In situ characterization of soil clay content with visible near-infrared diffuse reflectance spectroscopy. *Soil Sci Soc Am J* 71: 389–396
- Wilcox CH, Frazier BE, Ball ST (1994) Relationship between soil organic carbon and Landsat TM data in Eastern Washington. *Photogram Eng Remote Sens* 6:777–781
- Wold S, Sjöström M, Eriksson L (2001) PLS-regression: a basic tool of chemometrics. *Chemomet Intel Lab Sys* 58:109–130
- Wu Y, Chen J, Gong P et al (2007) A mechanism study of reflectance spectroscopy for investigating heavy metals in soil. *Soil Sci Soc Am J* 71:918–926
- Yoshido H, Leardi R, Funatsu K, Varmuza K (2001) Feature selection by genetic algorithms for mass spectral classifiers. *Anal Chim Acta* 446:485–494