POSSIBILITIES OF IMAGING SPECTROSCOPY FOR THE CLASSIFICATION OF CONTAMINATED AREAS IN RIVER FLOODPLAINS

Lammert Kooistra[1,2], Ron Wehrens[1], Lutgarde M.C. Buydens[1], Rob S.E.W. Leuven[2], and Piet H. Nienhuis[2]
[1]Laboratory of Analytical Chemistry
[2]Department of Environmental Studies
University of Nijmegen
Nijmegen, The Netherlands
Kooistra@sci.kun.nl

KEY WORDS: contaminated sediment, floodplains, soil spectroscopy, multivariate regression

ABSTRACT

During the past decades, large amounts of diffuse contaminated soil material have been deposited in the floodplains of the river Rhine. The dynamic character of these river causes a large spatial variability in the contamination level of its floodplain soils. As spatial information is required at different scale levels, characterisation exclusively based on soil sampling and analysis is often insufficient or too expensive. Hyper-spectral images can provide additional spatial information for a proper characterisation of the contamination situation of river floodplains. Earlier studies have investigated the application of imaging spectroscopy for the monitoring of vegetation stress as indicator for the contamination situation of a field. An alternative approach would be to determine the contamination level directly from soil reflectance characteristics. Imaging spectroscopy has already been used for the quantification of organic matter and clay content in the soil. In river floodplains, these characteristics are closely related to the elevated metal concentrations in the soil.
This paper describes the possible application of imaging spectroscopy to classify contaminated areas in river floodplains. Soil reflectance spectra in the visible-near infrared region (VNIR) were measured in the laboratory for soil samples taken from two river floodplains along the river Waal, the main tributary of the river Rhine in the Netherlands. Multivariate calibration techniques, in particular Partial Least Squares (PLS) regression, were used to test the performance of reflectance spectra in analysing both soil properties like clay and organic matter content and the level of metal contamination. Results for the analysis of two river floodplains are summarised and the influence of scale-level and sub soil material on the prediction capability are discussed.

1 INTRODUCTION

The spatial variability of diffuse soil contamination (e.g., heavy metals and polyaromatic hydrocarbons) in river floodplains along the Rhine is very high. Low pollutant concentrations can be found only a short distance from sites with relatively high contamination. The coming years, large-scale changes will be made to the floodplain landscape of the rivers in the Netherlands (e.g., lowering of riverbed and floodplains, displacement of river dikes and creation of side channels) to prevent future damage from flooding. Within these projects, spatial information on the soil quality along the rivers is required for applications like soil sanitation, ecological risk assessment, and volume estimations for clay and sand extraction. Depending on the application, this information has to be available at different scale-levels, ranging from river tributary to the individual floodplain. Under the present Dutch soil-protection

laws, sediment in large parts of the floodplains exceeds the intervention level of the Dutch soil qualification system, resulting in high removal costs and stagnation of projects. The recently formulated policy of dynamic soil management for river floodplains (Schouten et al. 2000) offers new possibilities, as pollution levels only have to be reduced to river-specific background levels instead of to the national reference values. For this approach, floodplains are divided into homogeneous pollutant units, *i.e.*, areas with a minimal variation in soil composition and soil quality (Kooistra et al. 2000). Remote sensing is a promising tool for providing spatial information to map the soil quality of river floodplains.

An important requirement for a successful application of remote sensing in mapping contaminated land is a detailed spatial and spectral resolution of the images (Folkard 1998). As the dimensions of the river floodplains are long and narrow, a minimal spatial resolution of five meter will be required. Furthermore, to detect subtle reflectance variations that can be attributed to elevated contamination levels, almost continuous spectral ranges are necessary. Imaging spectroscopy sensors fulfil these requirements. Within a sensor's field of view two main components of a river floodplain can be identified: vegetation and soil. Several studies have already investigated the application of imaging spectroscopy for the characterisation of vegetation stress as an indicator for soils contaminated with polyaromatic hydrocarbons (e.g. Jago et al. 1999, Llewellyn et al. 2000) or metals (e.g. Sommer et al. 1998). These results show that under a homogeneous vegetation cover, e.g. grassland, contamination patterns can be classified. Since large parts of the river floodplains are regularly flooded, resulting in large areas of bare soil material, an alternative approach would be to determine the contamination level directly from soil reflectance characteristics. Until now mapping soil characteristics through imaging spectroscopy has mainly focused on mineral deposits (e.g. Crosta et al. 1998) and organic matter content (e.g. Palacios-Orueta et al. 1999).

Laboratory studies have already shown successful application of soil reflectance in the visible-near infrared (VNIR) region for predicting soil characteristics like organic matter content (e.g. Ben-Dor et al. 1997, Morra et al. 1991, Henderson et al. 1989) and clay content (Ben-Dor and Banin 1995). In river floodplains, these soil characteristics are directly correlated to elevated metal concentrations in the soil (Middelkoop 1997). Although metals at low concentration levels do not have spectral features within the VNIR region, via intercorrelation with soil properties that are spectrally active in this region, predictions for the featureless soil constituents can be made (Ben-Dor and Banin 1995). The objective of this study is to investigate the possibilities of predicting metal contamination levels in river floodplains from high resolution reflectance spectra based on laboratory measurements of soil samples and multivariate regression.

2 METHODS

2.1 Soil Sampling

Soil reflectance characteristics were investigated in two river floodplains: the Afferdensche and Deestsche Waarden (ADW) and the Millingerwaard (MILW) (Figure 1). Both floodplains are located along the river Waal, the main tributary of the river Rhine in the Netherlands. A total of 69 soil samples were taken in the two

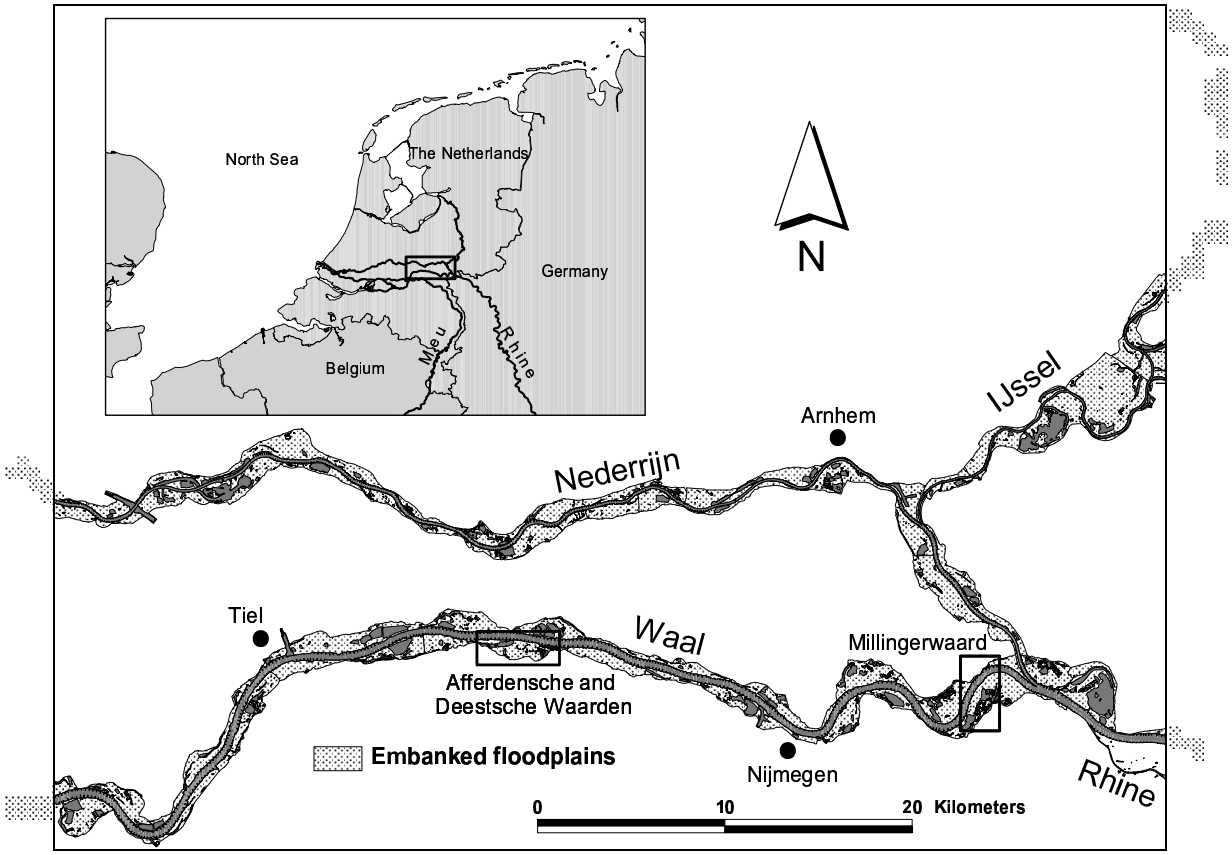

Figure 1. Location of the two floodplains along the river Waal

floodplains. In the ADW, a total of 31 sampling points distributed evenly over the floodplain were visited, and 44 soil samples were taken. For all locations the top soil (0-10 cm) was sampled. For a selection of 13 sampling points also soil material from the lower horizons and/or recently deposited soil material was sampled. Subsoil horizons were taken into account, because after reconstruction of the floodplains, which includes removing the first meter of the soil, also this soil material will be at the surface. For the MILW no information from earlier studies was available. For this floodplain, 25 soil samples were taken at a regular grid of 18 sampling points. Also for this floodplain, all top horizons were sampled with 8 additional samples from lower or recently deposited soil horizons. For the samples of the top horizons, three scoops within 1 square meter were taken. For the lower horizons, 10 cm cores were taken from the auger and used as sample material. Each sample was split into two subsamples. One was used for spectroscopic measurements, the other analysed to characterise soil chemical properties.

2.2 Chemical Analysis

The organic matter content for the soil samples was obtained by a loss-on-ignition method (Houba et al. 1989) and the clay content was determined with the laser diffraction method (Buurman et al. 1997). For the determination of the total metal concentration, approximately 20 g of field moist soil material was dried at 105 °C for 24 hours. Soils were grounded with a mortar and particles larger than 2 mm were removed by sieving. Dried soil samples (1 g) were treated with a HNO_3/H_2O_2 solution in Teflon-lined bombs using the microwave digestion method; after mineralisation, total cadmium (Cd) was measured by means of graphite furnace atomic absorption spectrometry (AAS) and total zinc (Zn) by ICP-AES spectrometry. All individual soil samples were digested in triplicate.

2.3 Spectral measurements

The sample preparation followed the procedure from Henderson et al. (1992). Soil samples were dried at 40 °C for 24 h to standardise the moisture level. In order to minimise anisotropic scattering of light by soil aggregates of variable size, soils were grounded with a mortar and passed through a 2 mm sieve. The absorbance spectra of the soil samples were obtained by means of a NIRS 6500 spectrophotometer (NIRSystems, Silver Spring, MD), configured with a reflection module and PbS detectors. The spectra were measured between 400 and 2500 nm, at 2 nm increments (1050 variables), using a resolution of about 10 nm bandwidth. The spectra were corrected for background absorption by division of a reference spectrum of Spectralon® 99% (Labsphere inc.), and the final spectrum is an average based on 32 scans.

2.4 Modelling

The partial least squares (PLS) regression method (Geladi and Kowalski 1986, Massart et al. 1997) was used to establish relations between the spectra and measured soil variables. It is based on latent variable decomposition of two blocks of variables, matrices **X** and **Y**, which contain spectral data and soil characteristics respectively. The purpose of the method is to find a small number of latent factors that are predictive for **Y** and use **X** efficiently. For the calculation of the regression models, the spectral response was expressed in absorbance values. All calculations were written in Matlab™. Some standard techniques were taken from the PLS_Toolbox for Matlab™.

Because a limited number of samples was available, a leave-one-out validation procedure was adopted to verify the prediction capability of the method. At each time, from all *n* samples within the dataset, *n* – 1 were utilised to build the regression model. Based on this model, the value for the soil property of the sample not used in developing the model was predicted. This procedure was repeated for all *n* samples, resulting in predictions for all samples. The parameter used to evaluate the quality of the results was the root mean square error of prediction (RMSEP) given by

$$\text{RMSEP} = \sqrt{\frac{\sum (y_m - y_p)^2}{n}} \tag{1}$$

where y_m is the measured value for a soil parameter and y_p is the value predicted by the PLS model.

3 RESULTS AND DISCUSSION

The results for the chemical analysis of the soil samples from the two floodplains are given in the box plots of Figure 2. It is apparent that the samples from the ADW floodplain have both a higher clay content and organic matter content compared to the samples of the MILW floodplain. Both the median value and the ranges of the distribution are higher for the samples of the ADW floodplain. In contrast, the box plots for the Cd and Zn concentration of the two floodplains show comparable distribution ranges. For the ADW floodplain, a higher median value for the two metal

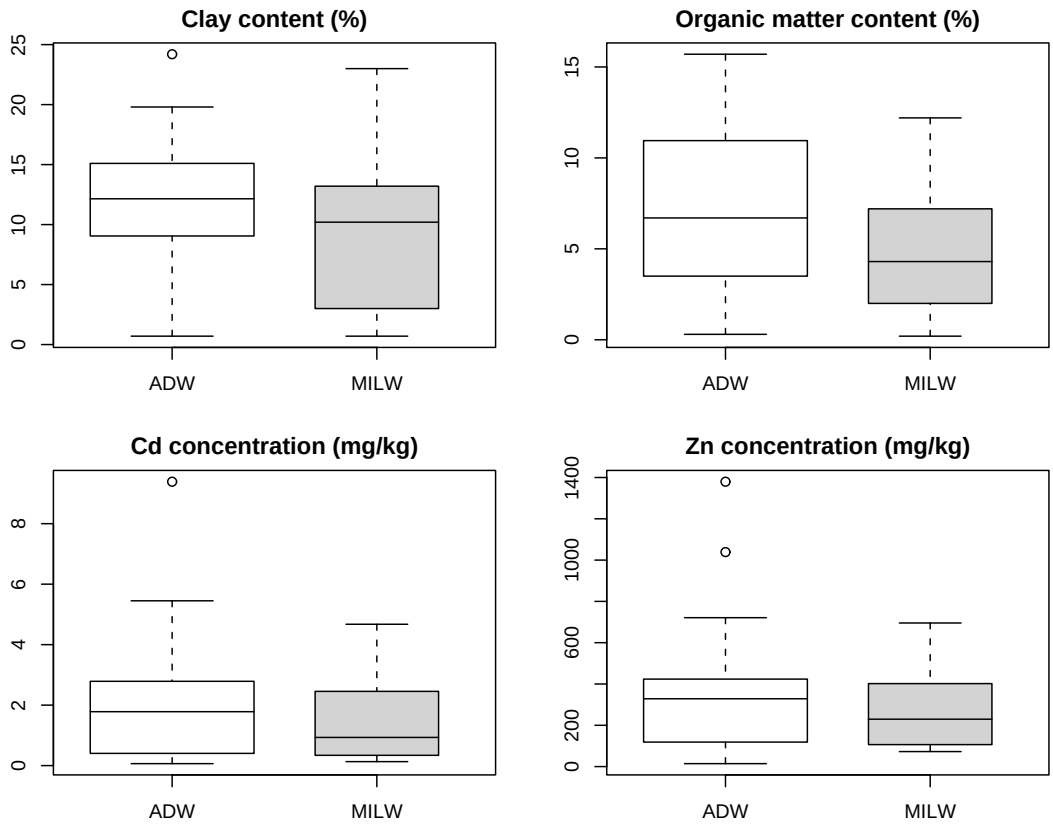

Figure 2. Box plots of clay and organic matter content and metal concentrations for the Afferdensche and Deestsche Waarden (ADW) and Millingerwaard (MILW). The ends of the box correspond to the first and third quartile of the distribution and the median is represented by the horizontal bar. The most remote, non extreme values are pictured with a small horizontal line under and above the box. Extreme, outlying observations are indicated with circles.

concentrations is observed. The box plots indicate that both the Cd and Zn concentration are characterised by a skewed distribution. Also the distributions for the clay and organic matter content in the MILW show this tendency; for the ADW a symmetric distribution is observed. Based on all 69 samples, variations of the organic matter content were positively correlated with the Cd concentration (r=0.74, p<0.01) and the Zn concentration (r=0.74, p<0.01). For the clay content, moderate correlations were found with Cd (r=0.43, p<0.01) and Zn (r=0.46, p<0.01). These results confirm earlier observations by Middelkoop (1997) that the amounts of metals contained in a sediment body depend on the exchange capacity of the sediment particles, which is largely determined by the contents of clay and organic matter.

The outliers in the ADW floodplain for the Cd and Zn concentration, depicted by the circles above the box plot (Figure 2), are all from subsoil samples. As older and deeper sediment layers were deposited during periods with a higher diffuse contamination load from the upstream river basin, they will have a different correlation with the clay and organic matter content. These a-typical observations, or outliers, may decrease the prediction accuracy of the regression model. To investigate both the influence of subsoil samples and the scale-level on the prediction accuracy, PLS models were calculated for three datasets. The first set consisted of all 69 samples taken from the two floodplains, representing the scale-level of the Waal tributary including both top and subsoil samples. The second set only included the 49 topsoil samples from the two floodplains. With this set the

influence of the subsoil samples on the prediction accuracy was investigated. The third set consisted of the 31 topsoil samples from the ADW floodplain, representing the scale-level of the individual floodplain.

Figure 3 shows the cross-validation RMSEP results of the PLS models for the three datasets of the four soil characteristics. On the x-as the number of latent factors are given that are used for constructing the PLS model. For the final model one wants to use a small number of factors that gives a minimal RMSEP. Including too many factors often leads to a deterioration in prediction ability (overfitting). The plots for clay and organic matter content show that exclusion of the subsoil samples does not improve the prediction capability of the model. For the clay content, the RMSEP values at the floodplain scale-level are even higher compared to the tributary scale-level including subsoil material. The three PLS models for the organic matter content show comparable RMSEP curves. This means that for this variable robust models can be made which are valid at different scale-levels. The plots of the metal concentrations show that removal of subsoil samples significantly improves the prediction capability of the PLS model. Differences between the RMSEP curves at the two scale-levels are relatively small, although for both the Cd and Zn concentration the RMSEP values tend to be smaller for the individual ADW floodplain.

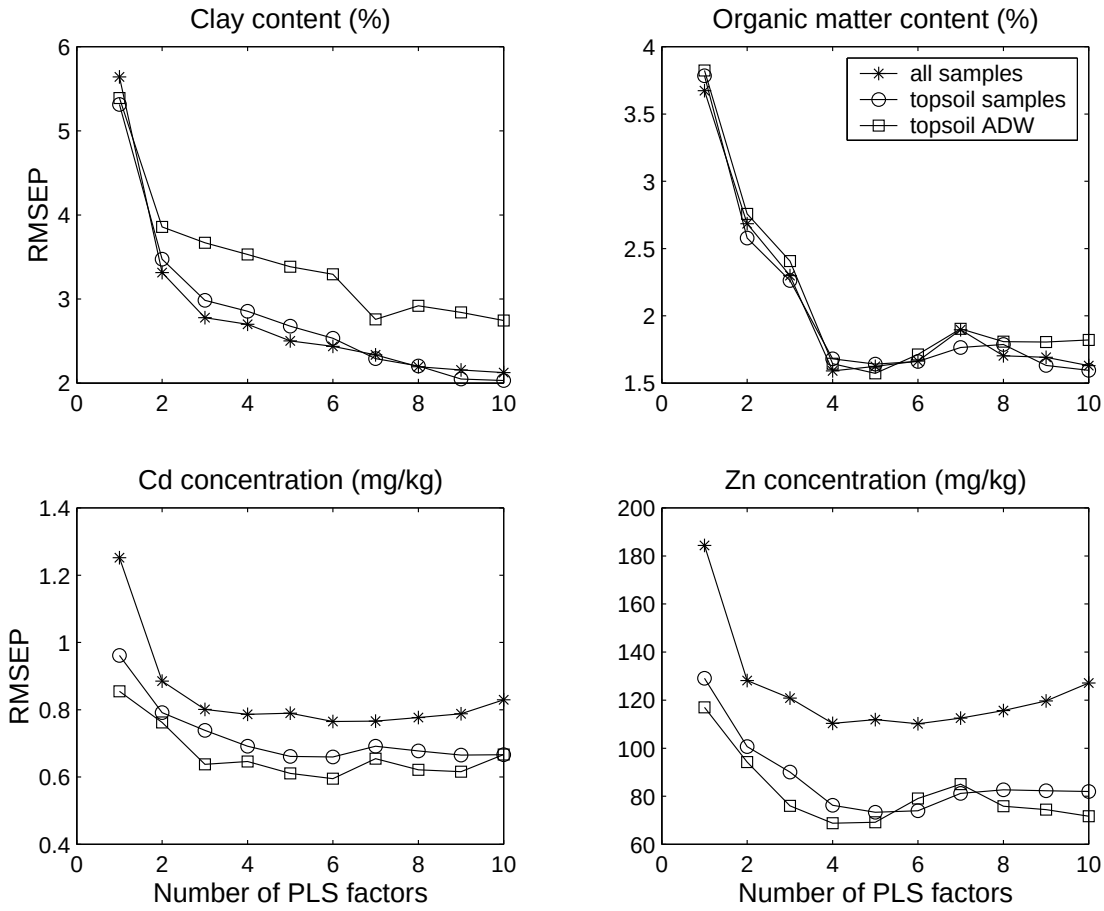

Figure 3. Root mean square error of prediction (RMSEP) for the soil characteristics against the number of PLS factors for all 69 soil samples, the 49 topsoil samples of the ADW and MILW floodplain, and the 31 topsoil samples of the ADW floodplain.

	All samples (n=69)		Topsoil ADW and MILW (n=49)		Topsoil ADW (n=31)	
	PLS factors	rel. RMSEP	PLS factors	rel. RMSEP	PLS factors	rel. RMSEP
Clay	3	0.2632	3	0.2609	2	0.3267
OM	4	0.2790	4	0.2503	4	0.2191
Cd	3	0.4564	5	0.3706	3	0.3778
Zn	4	0.3684	4	0.2399	4	0.2242

Table 1. Prediction results of the PLS models

The criteria of lowest RMSEP and an optimal number of PLS-factors were used to calculate PLS models for the three datasets of Figure 3. The location of the minimum RMSEP is not always well defined (Massart et al. 1997). For example, the PLS model of the Zn concentration for all topsoil samples (Figure 3) has its minimum at 6 factors (RMSEP = 75.5 mg/kg). But when the minimum is very shallow, one might trade in a few factors for a simpler, and probably more robust, model with the same prediction error (4 factors and RMSEP = 79.4 mg/kg). Results for the four soil characteristics are summarised in Table 1. In order to be able to compare differences between datasets, relative RMSEP values for a soil property were calculated as the RMSEP corresponding to the optimal number of PLS-factors divided by the mean value of the soil property in a given dataset. For all cases, the expected prediction error as estimated by the relative RMSEP is significantly smaller than half the mean value of the soil parameter of interest. Figure 4 shows the plots of the predicted vs. measured value for the total dataset of the four examined soil

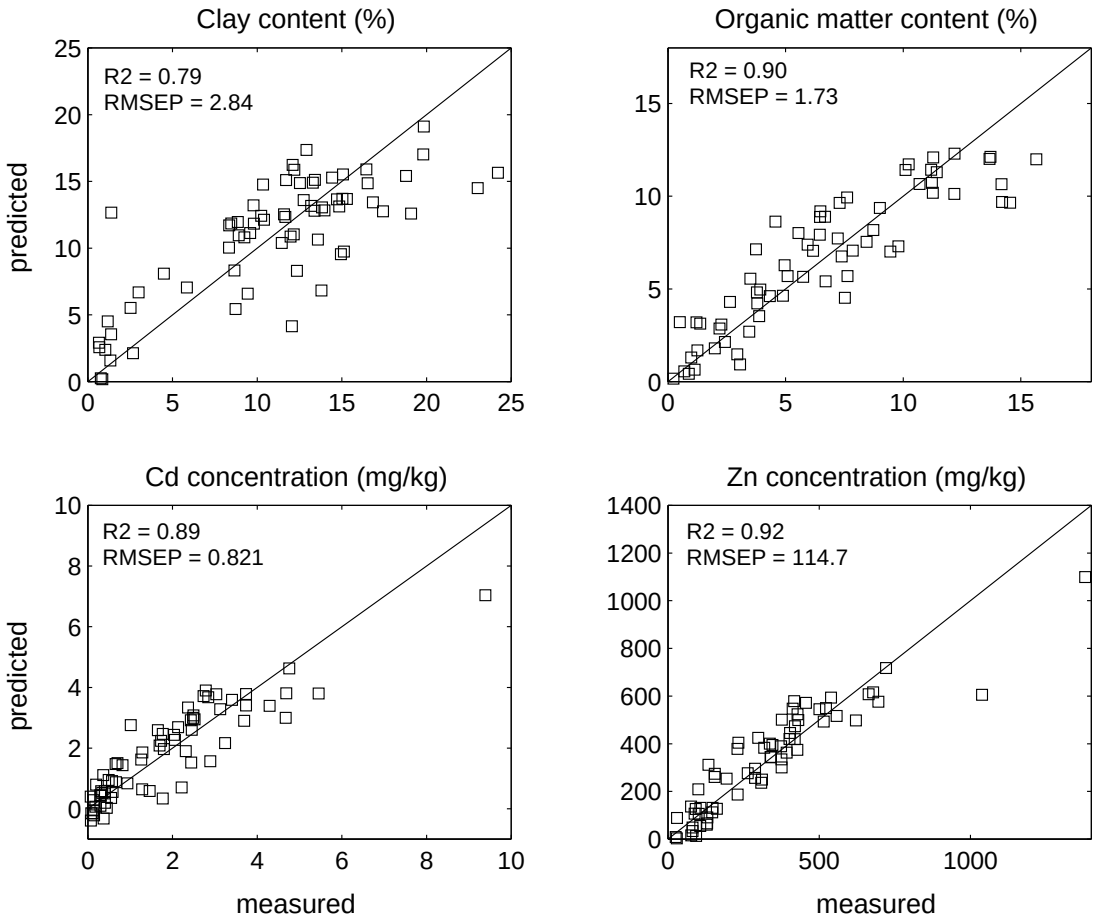

Figure 4. Plots of measured soil characteristics against the predicted for all 69 samples based on the PLS models of Table 1.

properties. The number of PLS factors used for constructing the models are given in Table 1. Also shown are the RMSEP and linear correlation coefficient of the model in each plot. For both the soil characteristics and the metal concentrations the points fall in the vicinity of the 1:1 line with a few outliers. The two points with the highest measured Zn concentrations can be seen in Figure 2 as well, an correspond to subsoil samples.

The required prediction accuracy will depend on the objective of a specific application. For example, in the Netherlands contaminated soil material from river floodplains is divided into four soil quality classes. Based on the ranges of these classes, satisfactory predictions can be made for the Zn contamination level. For the Cd concentration, predictions for the lower classes are less accurate as the ranges of these classes are relatively narrow, but for samples with higher Cd contamination levels good results can be achieved. This allows a rapid determination of the relative contamination level for a large number of samples in the laboratory.

Further improvement of the prediction capability could be achieved by pretreatment of the spectral data (e.g., linearization, smoothing, feature selection) to correct for baseline deviations, light scattering and lack of reproducibility in data acquisition (Masserschmidt et al. 1999). For example, using the first derivative of the reflectance improved the prediction results for the clay content significantly. This is especially the case when the method is applied to data collected directly in the field or with remote sensing instruments. In those situations results are affected by the effects of natural soil surface conditions (e.g., roughness, moisture, stoniness, etc.), the influences of the atmosphere and the illumination conditions (Leone and Sommer 2000). These influences will be further studied on the basis of spectral data collected directly in the field.

4 CONCLUSIONS

This study showed that high resolution spectra of soil samples taken from river floodplains can be used for predicting elevated metal concentrations of Cd and Zn. Soil spectroscopy in the VNIR region with a partial least squares (PLS) model is shown to be a very promising method for the determination of both soil properties and metal concentrations in floodplain soils. The results obtained when comparing PLS models of three datasets differing in spatial variability showed that the samples taken from deeper and older sediment layers have a large influence on the prediction accuracy for metal concentration. The influence of the scale-level was of less importance. The results from this laboratory study indicate the potential of using remote sensing images for the classification of contaminated areas in river floodplains. To understand the variability of soil contamination in the floodplain landscape more completely, it will be informative to analyse an extended dataset more completely in a geographic context. This can be achieved by using a relational GIS database, where other types of information such as vegetation and geomorphology are also included. Apart from remote sensing applications, this method also allows a rapid determination of metal concentrations for a large number of samples in the laboratory.

ACKNOWLEDGEMENTS

The authors wish to thank Wim Haanstra (Arcordis, Arnhem) for the use of the NIRS 6500 spectrophotometer and Tom Pape and Peter Buurman of Wageningen University for offering the opportunity to do the laser diffraction measurements.

REFERENCES

Ben-Dor E., Banin A., 1995. Near-Infrared Analysis as a Rapid Method to Simultaneously Evaluate Several Soil Properties. Soil Sci. Soc. Am. J. 59, 364-372.

Ben-Dor E., Inbar Y., Chen Y., 1997. The reflectance spectra of organic matter in the visible near infrared and short wave infrared region (400-2,500 nm) during a controlled decomposition process. Remote Sensing of Environment 61, 1-15.

Buurman P., Pape Th., Muggler C.C., 1997. Laser grain-size determination in soil genetic studies 1. Practical problems. Soil Science 162, 211-218.

Folkart A.M., 1998. How far can imaging spectroscopy go in identifying the cause of vegetation stress around landfill sites? Proceedings of the 1st EARSeL workshop on imaging spectroscopy. Zurich, Zwitserland, p. 447-453.

Geladi P., Kowalski B.R., 1986. Partial Least-Squares Regression: A Tutorial. Analytica Chimica Acta 185, 1-17.

Henderson T.L., Szilagyi A., Baumgardner M.F., Chih-chien T.C., Landgrebe D.A., 1989. Spectral band selection for classification of soil organic matter content. Soil Sci. Soc. Am. J. 53, 1778-1784.

Henderson T.L., Baumgardner M.F., Franzmeier D.P., Stott D.E., Coster D.C., 1992. High dimensional reflectance analysis of soil organic matter. Soil Sci. Soc. Am. J. 56, 865-872.

Houba V.J.G., van der Lee J.J., Novozamsky I., Walinga I., 1989. Soil and plant analysis, a series of syllabi. Part 5, Soil analysis procedures. Wageningen Agricultural University, The Netherlands.

Jago R.A., Cuttler M.E.J., Curran P.J., 1999. Estimating canopy chlorophyll concentration from field and airborne spectra. Remote Sensing of Environment 68, 217-224.

Kooistra L., Leuven R.S.E.W., Wehrens R., Buydens L.M.C., Nienhuis P.H., 2000. GIS as a tool for incorporating the spatial variability of pollutants in ecological risk assessment of Dutch floodplains. In R.S.E.W. Leuven, I. Poudevigne and R.M. Teeuw (eds.), Application of geographical information systems and remote sensing in river studies. Backhuys Publishers, Leiden, The Netherlands (in press).

Llewellyn G.M., Kooistra L., Curran P.J., 2000. The effect of soil contamination on grassland spectra. RSS 2000. Remote Sensing Society (in press).

Leone P.L., Sommer S., 2000. Multivariate analysis of laboratory spectra for the assessment of soil development and soil degradation in the Southern Apennines (Italy). Remote Sensing of Environment 72, 346-359.

Massart D.L., Vandeginste B.G.M., Buydens L.M.C., De Jong S., Lewi P.J., Smeyers-Verbecke J., 1997. Handbook of Chemometrics and Qualimetrics: Part A. Elsevier Science, Amsterdam, The Netherlands.

Masserschmidt I., Cuelbas C.J., Poppi R.J., De Andrade J.C., De Abreu C.A., Davanzo C.U., 1999. Determination of organic matter in soils by FTIR/diffuse reflectance and multivariate calibration. Journal of Chemometrics 13, 265-273.

Middelkoop H., 1997. Embanked floodplains in the Netherlands. Geomorphological evolution of various time scales. Netherlands Geographical Studies 224. PhD Thesis, Utrecht University, Utrecht, The Netherlands.

Morra M.J., Hall M.H., Freeborn L.L., 1991. Carbon and nitrogen analysis of soil fractions using near-infrared reflectance spectroscopy. Soil Sci. Soc. Am. J. 55, 288-291.

Palacios-Orueta A., Pinzon J.E., Ustin S.L., Roberts D.A., 1999. Remote sensing of soil properties in the Santa Monica Mountains: II. Hierarchical Foreground and Background Analysis. Remote sensing of the Environment 68, 138-151.

Schouten C.J.J., Rang M.C., De Hamer B.A., Van Hout H.R.A., 2000. Strongly polluted deposits in the Meuse river floodplain and their effects on river management. In A.J.M. Smits, P.H. Nienhuis and R.S.E.W. Leuven (eds.), New Approaches to River Management. Backhuys Publishers, Leiden, The Netherlands, p. 33-50.

Sommer S., Hill J., Mégier J., 1998. The potential of remote sensing for monitoring rural land use changes and their effects on soil conditions. Agriculture, Ecosystems and Environment 67, 197-209.