The relationship between diffuse spectral reflectance of the soil and its cation exchange capacity is scale-dependent

A. Savvides[a], R. Corstanje[b,d], S.J. Baxter[a], B.G. Rawlins[c], R.M. Lark[b]

[a] *Department of Soil Science, The University of Reading, Whiteknights, Reading, RG6 6DW, UK*

[b] *Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK*

[c] *British Geological Survey, Keyworth, Nottingham NG12 5GG, UK*

[d] Now at *National Soil Resources Institute, Cranfield University, Cranfield, Bedford, MK43 0AL, UK*

Abstract

Diffuse reflectance spectroscopy (DRS) is increasingly being used to predict numerous soil physical, chemical and biochemical properties. However, soil properties and processes vary at different scales and, as a result, relationships between soil properties often depend on scale. In this paper we report on how the relationship between one such property (CEC) and the DRS of the soil depends on spatial scale. We show this by means of a nested analysis of covariance of soils sampled on a balanced nested design in a 16 km $\times$ 16 km area in eastern England. We used principal components analysis on the DRS to obtain a reduced number of variables while retaining key variation. The first principal component accounted for 99.8 % of the total variance, the second for 0.14 %. Nested analysis of the variation in the CEC and the two principal components showed that the substantial variance components are at the $>$ 2000-m scale. This is probably the result of differences in soil composition due to parent material.

We then developed a model to predict CEC from the DRS and used partial least squares (PLS) regression do to so. Leave-one-out cross-validation results suggested a reasonable predictive capability ($R^2 = 0.71$ and RMSE = 0.048 $\mathrm{mol_c kg^{-1}}$). However,

the results from the independent validation were not as good, with R^2= 0.27, RMSE = 0.056 $mol_c kg^{-1}$ and an overall correlation of 0.52. This would indicate that DRS may not be useful for predictions of CEC. When we applied the analysis of covariance between predicted and observed we found significant scale-dependent correlations at scales of 50 and 500 m (0.82 and 0.73 respectively). DRS measurements can therefore be useful to predict CEC if predictions are required, for example, at the field scale (50 m). This study illustrates that the relationship between DRS and soil properties is scale dependent and that this scale dependency has important consequences for prediction of soil properties from DRS data.

Keywords: Pedometrics; Nested Sampling; Diffuse Reflectance Spectra.

---

*Corresponding author: *E-mail address*: roncorstanje@cranfield.ac.uk (R. Corstanje).

1. Introduction

There is a large demand for data on soil for quality assessment, environmental monitoring and precision agriculture. Diffuse reflectance spectroscopy, in the near infrared (NIR, 750 – 2500 nm) or visible and near infrared (VNIR, 350 – 2500 nm) frequency bands, has been proposed as a rapid, cost-effective and non-destructive method to obtain predictions of soil properties which would be too expensive to measure directly on many samples (e.g. Viscarra Rossel et al., 2006; Zornoza et al., 2008). The method is based on the premise that the variation of the diffuse spectra of soil at these wavelengths is due to variation in the composition of the soil (Cohen et al., 2005). Predictive relationships between the spectra and target soil properties are obtained by chemometric methods such as partial least-squares regression (PLSR; Viscarra Rossel et al., 2006). Diffuse reflectance spectroscopy has been used to predict various properties of the soil including pH, cation exchange capacity, organic matter or organic carbon

content, composition of the microbial community and extracellular enzyme activities (see Viscarra Rossel et al., 2006 and Zornoza et al., 2008).

Variation of soil properties depends on factors such as parent material, climate, land use and topography. These factors all operate at different scales and will therefore influence soil processes and soil variation at different scales. As a consequence the relationship between soil variables might also be scale dependent. This has been shown in previous studies on heavy metals in soil (Goovaerts and Webster, 1994), organic carbon and urease activity (Corstanje et al., 2007), and the effects of various soil properties on the rates of emission of trace gases (Lark et al., 2004). One result of such scale-dependence is that the overall correlation between two soil properties might mask underlying relationships at different scales.

We are not aware of any previous studies on how the relationship between diffuse reflectance spectra and soil properties depends on spatial scale. Published studies report that DRS is effective for predicting some soil properties such as soil organic C, but is less effective at predicting others such as CEC or nutrient content such as N or P (Viscarra Rossel et al., 2006). However, in these studies the measurements are made on specimens collected on some support (e.g. a core) distributed across a field or landscape according to some sampling scheme. The covariation of DRS measurements and soil properties therefore contains unresolved contributions from a range of spatial scales. Diffuse reflectance spectra are surrogates for soil properties that determine, more or less directly, the nature of the interaction between soil and electromagnetic radiation. It is therefore possible that the variations of the spectra at some spatial scales are dominated by variation of one soil property, and other soil properties cause most of the spectral variation at other scales. A weak overall correlation of the spectra with measurements of the soil property might therefore mask a strong relationship at some particular scale, and if this scale coincides with the scale at which information on the soil property is actually needed (e.g. field averages) then the DRS measurements

may be of considerably more practical value than is indicated by the simple statistics on the basic sample support.

We therefore need to study the scale-dependence of the relationship between DRS measurements and soil properties. Spatially nested sampling is an efficient way to study scale-dependent variation over large areas and very disparate scales (Youden & Melich, 1937; Webster and Oliver, 1990). The variance of soil properties sampled this way can be partitioned into scale-specific components by a nested analysis of variance, and the covariances of properties can similarly be analysed by scale (Lark, 2005). In a recent study (Corstanje et al., 2008) we investigated the covariance of soil properties such as pH, CEC and bulk density with rates of ammonia volatilization from soils collected by a nested scheme across a variable region in the eastern midlands of England. The resulting collection of soil offered the opportunity to investigate the relationship between the DRS of the soils and a basic soil property at disparate scales. That analysis is the subject of this paper.

We chose cation exchange capacity (CEC) as the target soil property for this study. The CEC of the soil is a basic physico-chemical property. It is laborious to measure, since different cations must be extracted and determined. Nonetheless, it is important if we are to predict the behaviour of the soil since it affects, among other things, the behaviour of various pollutants in soil (Wang and Keller, 2008), the ability of the soil to retain and supply important plant nutrients (Bailey et al., 2008) and the rates of important steps in biogeochemical cycles including the emission of trace gases from the soil (Jarecki et al., 2008). It would therefore be useful if DRS could be used to predict CEC, and this has been attempted previously. Viscarra Rossel et al. (2006) report several studies in which CEC was predicted with reasonable success (coefficients of determination between 0.64 and 0.88).

In this paper we report analyses to investigate the scale-dependent relationship between DRS and CEC across our study area, and an evaluation by scale of the efficacy

of DRS-based predictions of CEC.

2. Materials and Methods

2.1 The study region and the sampling scheme

A detailed account of the study region and the sampling is given by Corstanje et al. (2008). The region is approximately 16 km × 16 km and lies between the towns of Luton (south) and Bedford (north) in the eastern midlands of England. Most of this region is over Cretaceous formations: Chalk, Gault Clay and Lower Greensand, but there is also Oxford Clay (Jurassic) in the north. Superficial material, including chalky boulder clay, and other glacial drift of variable texture cover the country rock over much of the region.

We used a balanced nested sampling design, in which n_1 sampling main stations are chosen on a grid or transect of interval d_1. Two substations (level 2) were then chosen about each main station, separated from each other by fixed distance, d_2, on a line on a random bearing. We repeated this procedure until, about each substation at level $m-1$, two sample points (level m) separated from each other by distance d_m were selected. A nested analysis of the variances and covariances of variables measured on the sample points is possible, and components associated with the spatial scales determined by the distances, $d_1, d_2, \ldots, d_m$ can be estimated. As described by Corstanje et al. (2008), our main stations were on nodes of a 2-km grid, chosen so that the associated (co)variance components would be dominated by differences between the major parent materials. The substations were separated by 500 m, 50 m and 2 m. We selected 36 main stations, each with eight sample points on the nested scheme giving 288 sample points in total.

2.2 Soil preparation and analysis

This sampling exercise was done as part of a study on ammonia volatilization from soil, and this is reflected in the sample treatment. The soils were air-dried, large

plant fragments were removed, sieved to pass 0.5 mm, and then 1-kg portions were washed in 1.5 dm^3 of 10 mM $CaCl_2$ to remove nitrate and replace exchangeable cations with Ca^{2+}. The soils were then air-dried again and re-sieved to 0.5 mm.

2.3 Measurement of Cation Exchange Capacity

The cation exchange capacity of each sample was determined as described by Rowell (1994). The exchangeable calcium, magnesium and potassium ions were extracted from a weighed subsample of the air-dried soil into 1 M ammonium ethanoate buffer (pH 7). This was then displaced with ethanol and then flame photometry was used to measure the concentrations of the three ions. Ammonium was then extracted from the soil with acidified 1 M KCl and measured by steam distillation and titration. The CEC was then expressed as $mol_c\ kg^{-1}$ air-dry soil.

2.3 Measurement of diffuse reflectance spectra

Soil samples were scanned in the visible–near infrared region (350–2500 nm) using an ASD (Analytical Spectral Devices, Boulder, CO) Agri-Spec NIR Spectrometer. A 20-g subsample of each soil sample was placed in a holder with a quartz window for scanning. Soils were illuminated and scanned from below using the spectrometer connected to an ASD muglight with an internal tungsten-quartz-halogen light source and a 12 mm spot size. Data were collected every 1 nm and every spectrum was an average of 25 readings. Each sample was scanned twice; the second scan was made after rotating the sample in its holder through 90° whilst placed on the muglight. During scanning, a Spectralon 99% reflectance panel was used to optimize and white-reference the spectrometer after scanning every set of ten samples. Before further statistical analysis, we obtained an average of two spectra for each sample, truncated by removing the values below 450 nm and above 2450 nm.

2.3 Statistical analysis

We used nested analysis of covariance to study the correlation of the DRS measurements

and soil CEC at different scales, and to assess predictions of CEC by partial least-squares regression on the DRS at a set of validation sites. In the following section we describe the general nested analysis. We then describe how this was used to compare the DRS measurements and CEC data and then to assess predictions of CEC.

2.3.1 Nested analysis Nested analysis of covariance is described by Lark (2005) and we give here only a summary for the balanced case. The randomization of directions in the nested sampling scheme allows us to treat the values of two soil properties, u and v, as random variables Z^u and Z^v, which comprise the following components;

$$\begin{aligned} Z^u_{ij\ldots m} &= \mu_u + A^u_i + B^u_{ij} + \cdots + \varepsilon^u_{ij\ldots m} \\ Z^v_{ij\ldots m} &= \mu_v + A^v_i + B^v_{ij} + \cdots + \varepsilon^v_{ij\ldots m}. \end{aligned} \quad (1)$$

The values μ_u, μ_v are the overall means of u and v, respectively. The random variables A^u_i, A^v_i are, respectively, the differences between the corresponding overall means, μ_u and μ_v, and the corresponding means of the ith main station. Similarly B^u_{ij}, B^v_{ij} are the differences, within the ith main station, between the mean values of the ith main station and jth substation. The variables $A^u_i,\ B^u_{ij}, \ldots$ and $A^v_i,\ B^v_{ij}, \ldots$ have zero mean, and the variables associated with each scale in the nested scheme (e.g. A^u_i and B^u_{ij}) have covariance matrices $\mathbf{C}_i,\ \mathbf{C}_j, \ldots$. The objective of multivariate nested analysis is to estimate these covariance matrix components, which are additive components of $\mathbf{C}$, the overall covariance matrix of the two random variables, since they are associated with the scales of interest in the sampling scheme. Because estimates of the covariance matrix components by method-of-moments are not guaranteed to be non-negative definite, and therefore admissible as covariance matrices for real random variables, Lark (2005) used a residual maximum likelihood (REML) algorithm due to Calvin & Dykstra (1992). We used this method in the present study.

The estimated covariance matrix components were then converted to correlation matrices by dividing each element, $\mathbf{C}_{k,l}$ by the square-root of the product of the corresponding elements on the main diagonal, $\mathbf{C}_{k,k}$ and $\mathbf{C}_{l,l}$. We obtained confidence

intervals for the scale-dependent correlations with Fisher's z-transform, following Lark (2005).

2.3.2 Spectral reduction and correlation with CEC

We used principal components analysis on the spectra to obtain a smaller number of variables that represent the key variations in spectral variation among our soils. This was done using GenStat (Payne et al., 2008) to analyse the correlation matrix of the spectral reflectance in the 2001 channels. Principal component analysis finds p linear combinations of a set of k variables that are uncorrelated (see, for example, Webster and Oliver, 1990). The first component has the largest variance of any possible such linear combination. The second component has the largest variance of any linear combination that is orthogonal to the first and so on. The sum of the variances of the principal components is equal to the sum of the variances of the original variables, but if there are correlations between the latter then a large proportion of the total variance of the full data set is represented by substantially fewer than p of the principal components. In fact in our case the first principal component accounted for 99.8% of the total variance, and the second for 0.14%. This shows that the spectra are very redundant. We used the nested analysis of covariance to investigate the scale-dependent correlation between these two principal components of the spectra and soil CEC.

2.3.3 Prediction and validation

The fact that more than 99% of the variance of the observations in 2001 channels can be accounted for by the first principal component indicates that there is a good deal of redundancy in the spectra, that is to say different channels are so strongly correlated that they present little independent information. This is a common situation in the analysis of spectra, and partial least squares (PLS) methods are widely used to obtain predictive regressions of variables of interest (such as soil properties) on such very redundant predictor variates.

In this study we used PLS to obtain predictive relationships between the diffuse

reflectance spectra and soil CEC, using a subset of the data for estimation of the regression model, and the remainder to test the predictions. We used the PLS regression (PLSR) algorithm in the ParLes package (Viscarra Rossel, 2008). In PLSR we have n observations of k soil variables (predictands) in the $n \times k$ matrix $\mathbf{Y}$ and n values of a p-variate predictor (e.g. the DRS) in the $n \times p$ matrix $\mathbf{X}$. In PLSR these variables are decomposed into common orthogonal factors (similar to the principal components discussed above) from which the original variates can be reconstituted by means of loading matrices for the predictands and predictors. The algorithm finds an orthogonal decomposition such that the first few factors account for as much variation in the predictands and predictors as possible. The decomposition can be expressed by the following equation

$$\begin{aligned} \mathbf{X} &= \mathbf{T}\mathbf{P}' + \mathbf{E} \\ \mathbf{Y} &= \mathbf{T}\mathbf{Q}' + \mathbf{F}, \end{aligned} \tag{2}$$

where $\mathbf{T}$ is an $n \times l$ matrix of factor scores, and $\mathbf{P}$ and $\mathbf{Q}$ are, respectively, $p \times l$ and $k \times l$ matrices of loadings. The number l is the number of factors that are assumed to be informative, and is selected according to a criterion such as the Akaike Information Criterion, see Viscarra Rossel (2008) for details. The matrices $\mathbf{E}$ and $\mathbf{F}$ contain residuals, i.e. the contributions of the excluded factors. The number of l used in the PLSR model was determined through leave-one-out cross-validation. We selected l= 4 on the basis of the RMSE and AIC criterion.

In this study we randomly divided the main-stations of our nested sampling scheme into a prediction set of 29 (232 observations) for estimation of the regression model to predict CEC and a validation set of 7 (56 observations) to test the predictions. We used the PLSR algorithm in the ParLes software to fit the predictive model. In Table 1 we report the results from the cross-validation on the prediction set and prediction at the separate validation sites.

Since the validation observations have a nested structure we were able to analyse

the covariation of the observed and predicted values of CEC by nested analysis of covariance. Correlations at each scale and their confidence intervals were computed.

3. Results

Summary statistics of the soil cation-exchange capacities are shown in Table 1. In Figure 1 we present the computed scale dependent variance components for CEC and the first two principal components of the DRS; PC1 and PC2. These show that variation in both the DRS and CEC is scale-dependent. The largest component of variance in CEC, which comprises 66% of the total, was at $>$ 2000 m. Variation at this scale will be predominantly due to differences between parent materials. The observed values of CEC were smaller over Lower Greensand, where the average CEC was 0.104 $\mathrm{mol_c\ kg^{-1}}$. Most soils over the Lower Greensand are sandy loams or loamy sands, so the CEC is relatively small because the soil contains relatively little clay. In contrast the average CEC over the Gault clay, Chalk and chalky till were larger: 0.219, 0.161 and 0.188 $\mathrm{mol_c\ kg^{-1}}$ respectively. The second-largest component of variance was at the 500-m scale (15 % of the total), followed by the 50-m scale (12 %) and $<$2-m scale (6 %). In general, then, the variation of CEC appears to be dominated by parent material differences. The smaller variance components at the finer scales will be due to variation in factors such as organic matter content of the soil.

The largest components of variance of both the first two principal components of the DRS were at the coarsest scale of $>$ 2000 m — 60 and 50 % of the total for PC1 and PC2 respectively. As with CEC, this suggests that the variation of the DRS in this data set is dominated by differences between the parent materials. The second-largest component of variance of both principal components was at the 500-m scale, with 22 % for both PC1 and PC2. Variation at this scale will be due to differences in land use and management practices as well as some short-range variation in parent materials such as superficial deposits. The components of variance for PC1 at the finer scales of 50 m and 2 m were small (6 and 8 %, respectively) for PC1, but slightly larger (12 and

16 %, respectively) for PC2.

The results from the nested analysis of covariance of CEC and each of the first two principal components of DRS are presented in Figure 2. The correlations of CEC and PC1 are weak, and the correlation at 500 m only (0.39) is significantly different from zero. The overall correlation between these two variables was weak and positive (0.19). In the case of PC2, we found significant, strong and positive correlations at all scales except the finest scale of 2-m. The observed scale dependent correlations were 0.60, 0.65 and 0.86 for scales 50-, 500- and >2000-m respectively and the overall correlation was 0.47.

The results of the PLS model fitting and validation are summarized in Table 2. Cross-validation of the fitted PLS model suggested that the predictions are reasonable (R^2 = 0.71 and RMSE = 0.048 $\mathrm{mol_c kg^{-1}}$). However, the tests on the separate validation data are less encouraging, with R^2= 0.27 and RMSE = 0.056 $\mathrm{mol_c kg^{-1}}$. If we consider the scale-dependent correlations (Figure 3), then these show that the overall correlation of the predicted and measured values of CEC in our validation set (0.52) masks stronger correlations (0.82 and 0.73 respectively) at the scales 50 and 500 m, while the correlation at 2m is zero. A correlation at the coarsest scale is not reported because the estimated covariance matrix was positive semi-definite, but not positive definite, so the estimated correlation is 1.0, see Lark (2005).

4. Discussion and Conclusions

We have seen that both CEC and the principal components of the DRS show scale-dependent variation with the variance components increasing with distance. This indicates that the variation of the DRS is dominated by aspects of the composition of the soil associated with parent material. By contrast, for example, Corstanje et al. (2008) found that about 20% of the variance in urease activity in this soil occurred at the 2-m scale. In this environment, broad-scale variations in parent material have a larger impact on the DRS than factors opertating at finer scales, such as geomorpho-

logical or biological controls on soil composition. This was also observed for topsoil geochemistry in eastern England (Rawlins et al., 2003). This is likely to explain the scale-dependent relationship between CEC and the spectra. The soil spectrum responds to components of the soil which are themselves correlated with its CEC.

In this landscape, for example, the large iron oxide content of the Lower Greensand is likely to be spectrally distinctive, and we have noted that soils on the Lower Greensand are generally lighter-textured with smaller CEC than those located over different parent materials. It is therefore likely that the better predictive relationships between CEC and the DRS will be seen at scales where these surrogate relationships are expressed, while the correlations at other scales are very poor. This is the behaviour that our nested sampling and analysis reveals in this case study.

The important point that these results illustrate is that a poor overall correlation between DRS and a target soil property, or poor overall validation statistics for predictions, do not necessarily indicate that the spectra are not suitable for predictive purposes. For example, although the overall correlation of the predicted and measured spectra in our validation set was only 0.52, the correlations of the components at the 50- and 500-m scales were much stronger. The very weak correlation at the 2-m scale masks the relationship at coarser scales.

In practice we might often be interested in predicting a soil property only at coarser scales. For example, if we want to estimate the mean CEC for each of a set of fields, then the relatively good correlation of DRS and CEC at the 50-m scale suggests that the spectral measurements might be useful, and the average predicted CEC for a set of soil specimens collected within a field should give a reasonable prediction of the true field mean. Similarly, cokriging estimates for blocks with sides 50 m or longer, from a set of measurements of CEC and a denser set of DRS spectra should be reasonable, because the variation at fine scales, where the DRS and CEC are weakly correlated, is averaged out.

Some additional issues are raised by this study. First, it provides evidence for the importance of assessing predictions from spectra on separate validation data sets, and not giving undue weight to cross-validation assessments. Second, the scale-dependence in the errors from our PLSR predictions suggests that there is a need to develop the algorithm to allow for models where the errors are not assumed, as in standard implementations of PLSR as we used here, to be independent random variables. While the regression coefficients are still unbiased, they are not necessarily the ones that give us minimum variance predictions. There may therefore be advantages in extending the PLSR algorithm to deal with such circumstances.

To conclude, the relationship between DRS and soil properties has been shown to be scale-dependent for one case study. An important consequence of this is that assessments of the predictive value of statistical models that use DRS to predict soil properties should account for scale-dependence. If this is not done then weak relationships between spectral properties and the target soil property at one or more scale might obscure strong relationships at other scales, which might well be scales at which the soil information is needed. The nested sampling and analysis scheme used in this study is one way to identify such behaviour.

Acknowledgements.

The field work described here, along with RC's contribution, was undertaken as part of a project funded by the United Kingdom's Biotechnology and Biological Sciences Research Council (BBSRC), grant no. BB/C506813/1. RML's contribution was part of the programme of the Centre for Mathematical and Computational Biology at Rothamsted Research, funded by the BBSRC. AS acknowledges the technical assistance of Karen Gutteridge and Chris Speed at the University of Reading. This paper is published with the permission of the Executive Director of the British Geological Survey (Natural Environment Research Council).

References

Bailey, J.S., Ramakrishna, A. and Kirchof, G., 2008. Relationships between important soil variables in moderately acidic soils (pH <= 5.5) in the highlands of Papua New Guinea and management implications for subsistence farmers. Soil Use Manage., 24: 281–291.

Calvin, J.A. and Dykstra, R.L., 1992. An algorithm for restricted maximum likelihood estimation in balanced multivariate variance components models. J. Stat. Comput. Simul., 40: 233–246.

Cohen, M.J., Prenger, J.P. and DeBusk, W.F., 2005. Visible-near infrared reflectance spectroscopy for rapid, non-destructive assessment of wetland soil quality. J. Environ. Qual., 34: 1422–1434.

Corstanje, R., Schulin R. and Lark, R.M. ,2007. Scale-dependent relationships between urease activity and soil organic carbon. Eur. J. Soil Sci., 58: 1087–1095.

Corstanje, R., Kirk, G.J.D., Pawlett, R. M., Read, R. and Lark, R.M., 2008. Spatial variation of ammonia volatilization from soil and its scale dependent correlation with soil properties. Eur. J. Soil Sci., 59: 1260–1270.

Goovaerts, P. and Webster, R., 1994. Scale-dependent correlation between topsoil copper and cobalt concentrations in Scotland. Eur. J. Soil Sci., 45: 79–95.

Jarecki, M.K., Parkin, T.B., Chan, A.S.K., Hatfield, J.L. and Jones, R., 2008. Greenhouse gas emissions from two soils receiving nitrogen fertilizer and swine manure slurry. J. Environ. Qual., 37: 1432–1438.

Lark, R.M., 2005. Exploring scale-dependent correlation of soil properties by nested sampling. Eur. J. Soil Sci., 56: 307–317.

Lark, R. M., Milne, A. E., Addiscott, T. M., Goulding, K. W. T., Webster, C. P. and OFlaherty, S., 2004. Scale- and location-dependent correlation of nitrous oxide emissions with soil properties: an analysis using wavelets. Eur. J. Soil Sci., 55: 611–627.

Payne, R.W. (Editor). GenStat® for Windows™ 11th Edition Introduction.

Rawlins, B. G., Webster, R. and Lister, T.R., 2003. The influence of parent material on topsoil geochemistry in eastern England. Earth Surf. Processes Landforms, 28: 1389–1409.

Rowell, D.L., 1994. Soil Science: Methods and Applications. Longman, London.

Viscarra Rossel, R.A., 2008. ParLes: Software for chemometric analysis of spectroscopic data. Chemometrics Intell. Lab. Syst., 90: 72–83.

Viscarra Rossel, R.A., Walvoort, D.J.J., McBratney, A.B., Janik, L.J. and 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.

Wang, P. and Keller, A.A., 2008. Soil particle-size dependent partitioning behavior of pesticides within water-soil-cationic surfactant systems. Water Res., 42: 3781–3788.

Webster, R. and Oliver, M.A., 1990. Statistical Methods in Soil and Land Resource Survey. Oxford University Press, Oxford.

Youden, W.J. and Mehlich, A., 1937. Selection of efficient methods for soil sampling. Contrib. Boyce Thompson Inst. Plant Res., 9: 59-70.

Zornoza, R., Guerrero, C., Mataix-Solera, J., Scow, K.M., Arcenegui, V. and Mataix-Beneyto, J., 2008. Near infrared spectroscopy for determination of various phys-

ical, chemical and biochemical properties of Mediterranean soils. Soil Biol. Biochem., 40: 1923–1930.

Table 1. Summary statistics - soil cation exchange capacity (n=288).

Statistic	value /mol_c kg^{-1}
10^{th} percentile	0.074
Median	0.18
90^{th} percentile	0.28
Mean	0.18
Variance	0.0067

Table 2. Statistics for prediction performance by the partial least squares model. These are determined by cross-validation on the prediction set, or on the independent validation set.

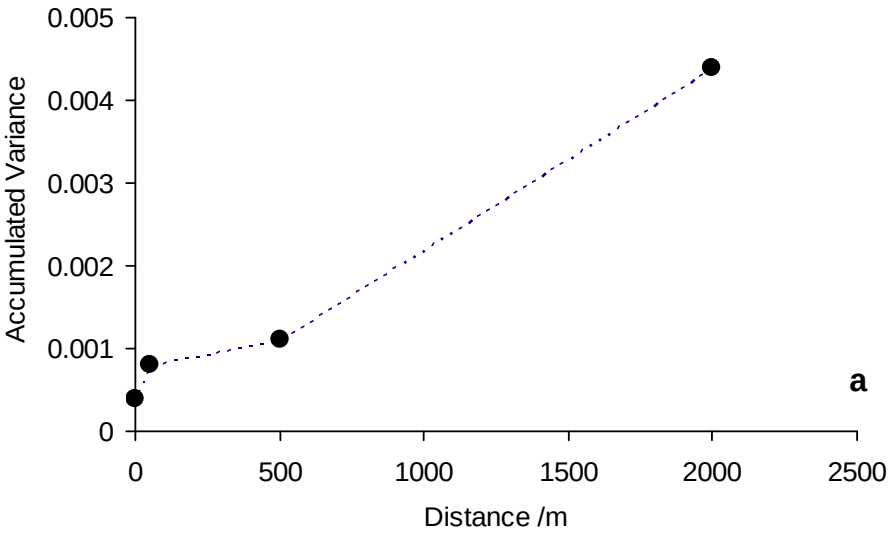
Model	R^2	R^2_{adj}	RMSE
Cross-validation	0.71	0.71	0.048
Independent	0.27	0.24	0.056

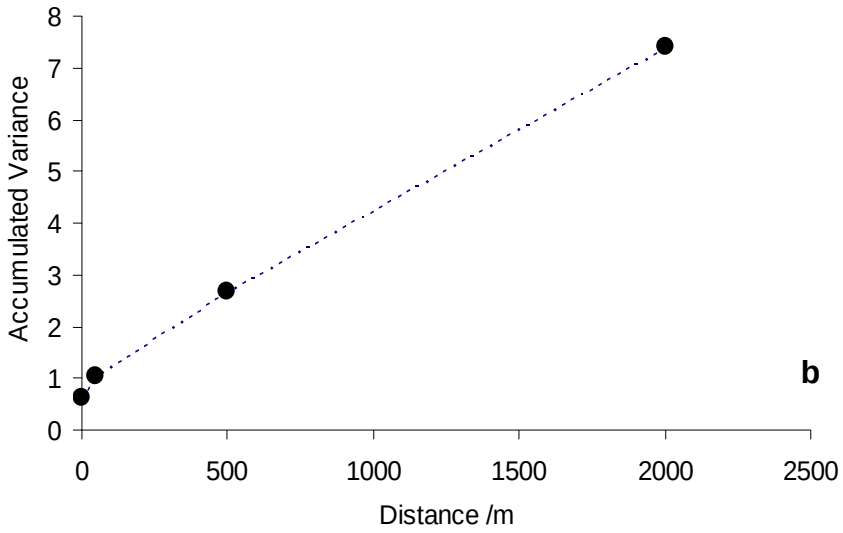
Figures

Figure 1. Accumulated components of variation for CEC (a) and the first two principal components of the diffuse reflectance spectra (PC1 and PC2; represented by b and c respectively). Components for the largest scale are plotted here against 2000 m but apply to distances > 2000 m.

Figure 2. Scale-dependent correlations between CEC and the first two principal components of the diffuse reflectance spectra (PC1 and PC2; represented by a and b respectively). The correlations are plotted with 95 % confidence intervals (bars). Components for the largest scale are plotted here against 2000 m but apply to distances > 2000 m.

Figure 3. Scale-dependent correlations between predicted and measured CEC from the validation set. The predicted CEC were obtained using a PLS model on the DRS. See text for details. The correlations are plotted with 95 % confidence intervals (bars). Components for the largest scale are plotted here against 2000 m but apply to distances > 2000 m.


Accumulated Variance
0.005
0.004
0.003
0.002
0.001
0
0
500
1000
1500
2000
2500
Distance /m
a


Accumulated Variance
8
7
6
5
4
3
2
1
0
0
500
1000
1500
2000
2500
Distance /m
b

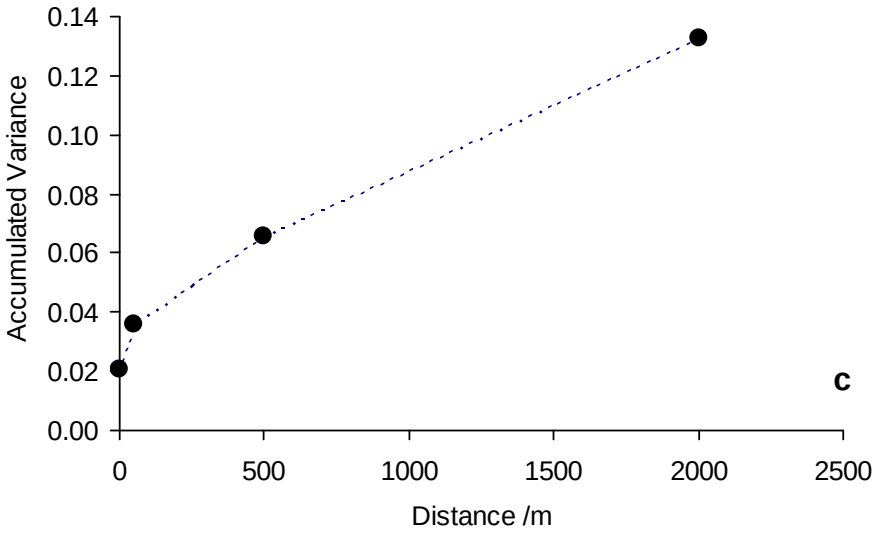
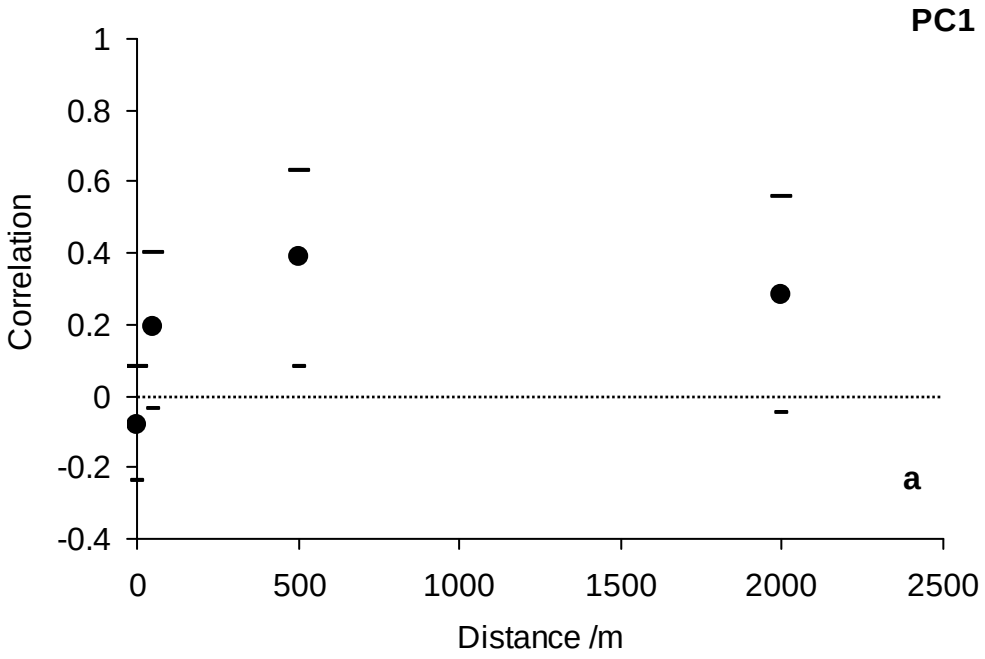

Accumulated Variance
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
0
500
1000
1500
2000
2500
Distance /m
c

Fig1


PC1
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
Correlation
a
0
500
1000
1500
2000
2500
Distance /m

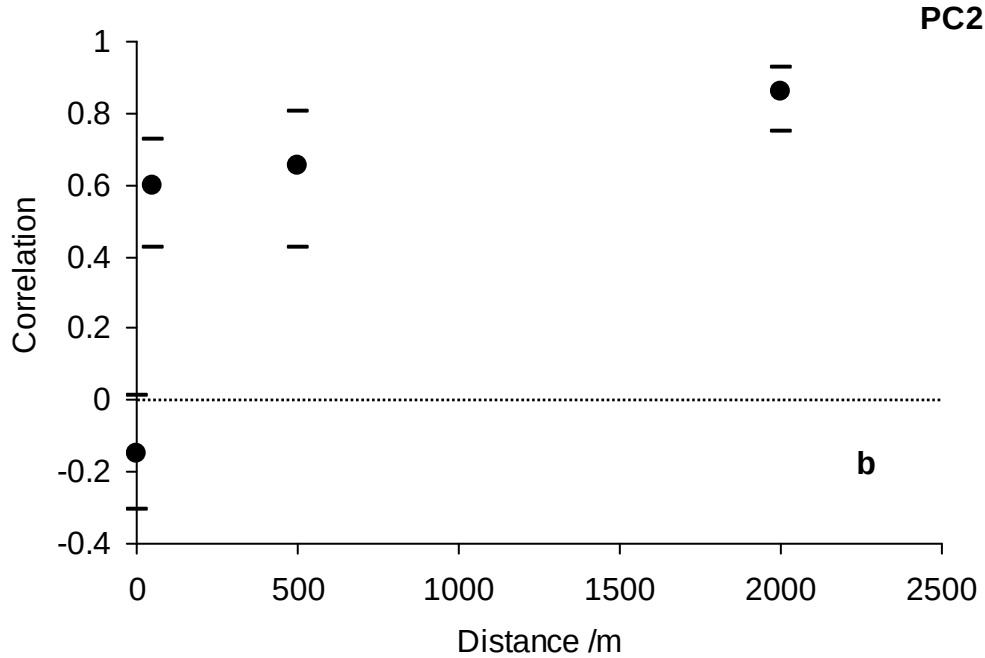

PC2
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
Correlation
b
0
500
1000
1500
2000
2500
Distance /m

Fig2

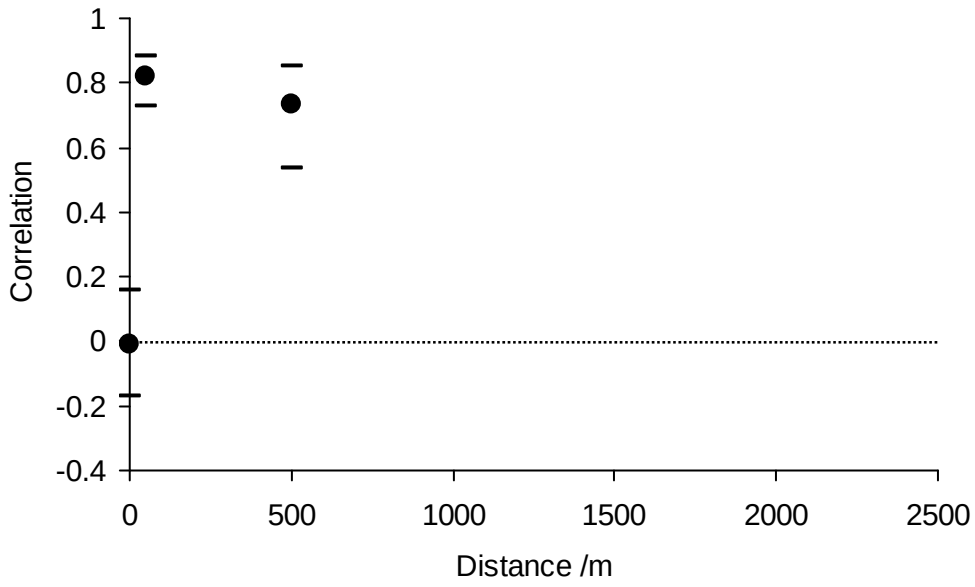

1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
Correlation
0
500
1000
1500
2000
2500
Distance /m

Fig3