# Airborne radiometric survey data and a DTM as covariates for regional scale mapping of soil organic carbon across Northern Ireland.

B. G. Rawlins[a,*], B. P. Marchant[b], D. Smyth[c], C. Scheib[a], R. M. Lark[b] & C. Jordan[d].

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

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

[c] *Geological Survey of Northern Ireland, Belfast BT9 5BF, UK*

[d] *Agri-Food and Biosciences Institute, Belfast BT9 5PX, UK*



* Corresponding author: B. G. Rawlins

British Geological Survey

Keyworth

Nottingham NG12 5GG

UK

phone: +44 (0) 115 9363140

Fax: +44 (0) 115 936 3100

e-mail: bgr@bgs.ac.uk

## Summary

Soil scientists require cost-effective methods to make accurate regional predictions of soil organic carbon (SOC) content. We assess the suitability of airborne radiometric data and digital elevation data as covariates to improve the precision of predictions of SOC from an intensive survey in Northern Ireland. Radiometric data (K band) and, to a lesser extent, altitude are shown to increase the precision of SOC predictions when they are included in linear mixed models of SOC variation. However the statistical distribution of SOC in Northern Ireland is bimodal and therefore unsuitable for geostatistical analysis unless the two peaks can be accounted for by the fixed effects in the linear mixed models. The upper peak in the distribution is due to areas of peat soils. This problem may be partly countered if soil maps are used to classify areas of Northern Ireland according to their expected SOC content and then different models are fitted to each of these classes. Here we divide the soil in Northern Ireland into three classes, namely mineral, organo mineral and peat. This leads to a further increase in the precision of SOC predictions and the median square error is 2.2 $\%^2$. However a substantial number of our observations appear to be mis-classified and therefore the mean squared error in the predictions is larger (30.6 $\%^2$) since it is dominated by large errors due to mis-classification. Further improvement in SOC prediction may therefore be possible if better delineation between areas of large SOC (peat) and small SOC (non-peat) could be achieved.

## Introduction

Soil organic carbon (SOC) is one of the most important constituents of the soil imparting structural stability, increased water holding capacity, acting as a source of nutrients, and as a store of terrestrial carbon. The quantity of organic carbon in the top 30 cm of the soil profile typically reflects the interplay of several factors including climate (annual rainfall and temperature), elevation, local topography and land use. Soil scientists require cost-effective methods to make accurate estimates of SOC content, which could be used to estimate soil-related carbon-dioxide emissions under the UNFCCC (United Nations Framework Convention on Climate Change).

Traditional, grid-based sampling, with laboratory measurement and univariate interpolation of SOC is subject to large estimation uncertainties at unsampled points. These can be significantly reduced if intensive, secondary covariates such as data from remote sensors are used for prediction by cokriging (McBratney & Webster, 1983), regression kriging (Odeh *et al.*, 1995), or the use of linear mixed models (Lark *et al.*, 2006). A variety of covariates have been shown to improve prediction of SOC. For example, terrain attributes and land use have been shown to be correlated with SOC at multiple scales (Mueller & Pierce, 2003), whilst hyperspectral airborne data (Selige *et al.*, 2006), surface reflectance (Chen *et al.*, 2000) and electrical conductivity (Simbahan *et al.*, 2006) was shown to be correlated with SOC in arable soils over scales from a few to tens of kilometres. These secondary covariates are likely to be more or less applicable in various types of soil environment (e.g. vegetated and unvegetated), and at differing scales.

Another potential covariate which can be used in both vegetated and unvegetated environments are measurements of gamma radiation from the decay of natural radionuclides in the soil. This radiation can be measured using airborne sensors; the data correspond to the top 50 cm of a mineral-dominated soil, and depths of up to one metre in low density materials such as peat. Airborne radiometric survey has been used extensively in Australia for digital soil mapping (Cook *et al.*, 1996). Typically the data on emissions are processed to generate values from three spectral bands which correspond to the decay of potassium (K), thorium (Th) and uranium (U). In a recent study in Australia, Minasny *et al.* (2006) combined data on radiometric K, land use and terrain attributes to develop a depth-based function for estimation of SOC.

There are two reasons why we might expect spatial correlation between gamma emissions from the soil and its SOC content in the wetter landscape of north-western Europe. First, the well-established spatial correlation between gamma-ray attenuation and soil moisture (Carroll, 1981) extends to SOC because the latter accumulates in soils which are wet or waterlogged for much of the year. Water reduces the intensity of gamma-rays significantly more than air; a 10% increase in soil water leads to a reduction in K gamma radiation by the same amount (Minty, 1979). Second, for soil with a wide range of SOC contents, the mineral content (and gamma emission) will be smaller where organic matter contents are larger for soils derived from the same parent material (with similar mineral composition). As the organic matter content rises, the mineral content declines in a simple, two-component composition. It may be possible to use these relationships to improve SOC estimation in organic rich soils such as those of the Arctic Tundra (Smith *et al.*, 2004) or temperate latitudes such as Scotland and Wales (Scottish Executive, 2007), so this approach warrants further investigation.

There have been relatively few regional-scale, airborne radiometric surveys of landscapes in which SOC contents represent significant terrestrial carbon stores – such surveys have been undertaken in Finland (Lilja & Nevalainen, 2005) and Sweden (Lunden *et al.*, 2001). One example is the recently-completed Tellus survey of Northern Ireland (13 550 km$^2$), in which SOC measurements and airborne geophysical surveys (including the detection of gamma emitting radiation) were undertaken at around the same time. The aim of this paper is to determine to what extent airborne radiometric survey data and terrain attributes can be used as secondary covariates to improve estimates of SOC across the landscape of Northern Ireland. A second objective is to assess whether improvements in SOC estimation based on these covariates differs markedly for the three major soil types across this landscape. Also we explore whether the inclusion of information on radiometric K means that the number of observations of SOC required for adequate predictions is reduced. We discuss the implications of our findings for improving the estimation of SOC in cognate landscapes and some potential limitations to the application of airborne radiometric survey for this purpose.

## Methods

*Study region and surveys*

The soils of Northern Ireland have been described by Cruickshank (1997) and comprise poorly-drained gley soils (54%), peats and rankers (24%) and freely drained soils (16%) – see Figure 1. In the soil surveys of Northern Ireland described by Cruickshank (1997) soil inspection pits were dug to between 80 and 90 cm. The larger proportion of gley soils by comparison to England, Wales and Scotland reflects the wetter environment of Northern Ireland, where average annual rainfall for the vast majority of the region is greater than 1 m, with a minimum of around 0.75 m. Large areas of the region are more than 100 m above sea level, with a maximum attitude around 850 m, whilst central and eastern areas have lower elevations (<40 m).

The airborne geophysical survey of the whole of Northern Ireland was flown in the summers of 2005 and 2006. Radiometric data were collected with an Exploranium GR820 256 channel gamma spectrometer system comprising 32 litres of downward looking NaI(Tl) detectors and 8 litres of upward looking detectors. Data were collected every second (approximately 70 m flight line distance). The gamma radiation measured comes from a shallow surface layer of no more than about 30 cm in rock, although this will increase for low-density unconsolidated materials, perhaps to a maximum of a few metres in dry peat. The ground area or footprint, from which most of the contribution of gamma radiation comes, has the form of an ellipse elongated in the flight direction. For example, at 56 m altitude, 75% of the measured radiation will come from a width of about 150 m, extending to around 220 m along the flight line (Pitkin & Duval, 1980). Survey lines were spaced 200 m apart and orientated NNW or SSE (165 and 345 °), with tie-lines 2000 m apart, and at right angles to the flight line directions, acquired only in the early part of the survey. The flying height was 56 m above ground in rural areas. Procedures for processing the airborne radiometric data were based on those described in AGSO and IAEA reference manuals (Grasty & Minty, 1995; IAEA, 1991). The processing included corrections for aircraft and cosmic background radiation, aircraft altitude and spectral interactions. The corrected count rates were used to estimate the concentration of the three radioelements across specific energy ranges (MeV): K (1.37–1.57), U (1.66–1.86), Th (2.41 – 2.81). The survey yielded *ca.* 1.2 million values for equivalent K(%) and Th (mg $kg^{-1}$), U (mg $kg^{-1}$) and man-made radionuclides, predominantly $^{137}Cs$, although we do not consider the latter in this paper.

To assess whether temporal fluctuations in soil moisture status during the period of the airborne surveys was broadly representative of the long-term average, we extracted from the MIDAS database (UK Meteorological Office) monthly rainfall data (mm) throughout 2005 and 2006, along with average monthly rainfall between 1961 and 1990 for meteorological stations across Northern Ireland.

The soil geochemical survey was undertaken between July 2004 and March 2006. A sample of topsoil was collected from a site in every other square kilometre of the Irish National Grid, by simple random selection within each square, subject to the avoidance of roads, tracks, railways, urban areas and other seriously disturbed ground. There were 6862 sample sites in total. At each site soil was taken with a hand auger from between depths of 5 and 20 cm from five holes at the corners and centre of a square with a side of length 20 m and combined to form a bulked sample. All samples of soil were air-dried in a dedicated temperature controlled oven at 30 °C for 2–3 days and disaggregated. From each a 50-g sub-sample was ground in an agate planetary ball mill. The total concentrations of 55 major and trace elements were determined in each sample by wavelength and energy dispersive XRFS (X-Ray Fluorescence Spectrometry), although we only consider K (%), Th (mg $kg^{-1}$) and U (mg $kg^{-1}$) in this study. Soil organic carbon was estimated in each sample using loss-on-ignition analysis by heating a sub-sample 450 °C for eight hours and multiplying the mass difference by 0.58 (Broadbent, 1953). The coefficient of variation for this method for 174 replicate analyses of a sample standard was 3.6%.

The topographic data were a series of 50-m spaced observations for elevation (m) covering Northern Ireland (Ordnance Survey of Northern Irelands data) based on airborne, photogrammetric acquisition; 65% of the data are accurate to ± 1 metre. Simple linear interpolators are often used to create continuous Digital Elevation Models (DEMs; Moore *et al.*, 1991) from stereo, aerial photo-derived point elevation data. We used inverse distance weighted (IDW) interpolation to form a DEM surface in ESRI ArcMap$^{TM}$.

We used the DEM to estimate Compound Topographic Index (CTI) in ArcInfo WorkStation$^{TM}$ which has been shown to be correlated with SOC content (Moore *et al.*, 1993). We extracted values for elevation (m) and CTI for the soil sampling locations by point intersection.

We used a spatial join procedure to associate each soil sampling observation with its nearest radiometric survey observation. The median distance between the sample sites and the corresponding radiometric measurement was 52 m, with an interquartile range of 47 m showing that the soil sampling locations fall entirely within the support of the airborne detector. We used digital versions of the (1:50,000) soil maps of Northern Ireland to form a three-fold classification of the soil sampling locations (see Figure 1): *organic soils* (SOC $>$ 20 % and $>$ 50 cm in thickness; peats), *organo mineral soils* (with an organic surface horizon overlying mineral subsoil; peaty podzols, rankers and humic-gleays) and *mineral soils* (no organic horizon and SOC$<$ 10 %; brown-earths, podzols, gleys and rankers).

*Exploratory analysis*

The appropriateness of using the airborne measurements of K, Th and U as auxiliary information in a regional survey of SOC was initially explored by (i) comparing the summary statistics with those of the corresponding ground based variables (Table 1) and (ii) by calculating the correlation coefficients between the airborne and ground based variables and the estimates of SOC based upon LOI analysis (Table 2). Similarly we also explored the correlation of SOC with the two terrain variables. These exploratory analyses were the basis for deciding on the variables to include in our models of SOC variation.

We calculated total annual rainfall for a subset of 20 meteorological stations across Northern Ireland for both 2005 and 2006, the years in which the airborne surveys were flown. A comparison of these data with the long-term, average annual rainfall (1961-1990) indicated that they were of a similar magnitude. We also plotted monthly rainfall totals for observations from several meteorological stations during 2005 and 2006 and compared these to the long-term monthly averages (1961-1990). There was no compelling evidence that rainfall throughout 2005 or 2006 was spatially or temporally anomalous and so we feel justified in assuming that the soil moisture regime over the period of the airborne survey was representative of its long-term variation.

*Spatial analysis: linear mixed models*

A prediction set of 3000 observations was randomly extracted from the SOC data set and the remaining observations were used as a validation set. We considered linear mixed models of the form

$$\mathbf{z} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\eta}, \tag{1}$$

where $\mathbf{X}$ is an $n \times p$ design matrix containing values of $p$ auxiliary variables or fixed effects, $\boldsymbol{\beta}$ is the length $p$ vector containing the coefficients of the fixed effects and $\boldsymbol{\eta} \sim \mathcal{N}(0, \mathbf{V})$ is a vector of spatially correlated random residuals with a Gaussian distribution and covariance matrix $\mathbf{V}$. We assume that the spatial correlation of $\boldsymbol{\eta}$ can be represented by an isotropic nested nugget and Matérn variogram model

$$\begin{aligned} \gamma(h) &= c_0 + c_1 \left\{ 1 - \frac{1}{2^{\nu-1}} \Gamma(\nu) \left( \frac{h}{a} \right)^{\nu} K_{\nu} \left( \frac{h}{a} \right) \right\} \text{ for } h > 0, \\ \gamma(h) &= 0 \text{ for } h = 0, \end{aligned} \tag{2}$$

where $h$ is the lag distance separating observation pairs, $\nu$ is a smoothing parameter,

$K_\nu$ a modified Bessel function of the second kind of order $\nu$ (Abramowitz & Stegun, 1972) and $\Gamma$ is the gamma function. Bessel and gamma functions may be calculated by many standard numerical packages such as IMSL (1994). The smoothness parameter gives the Matérn function greater flexibility for modelling the spatial covariance than more commonly-used models such as the exponential and spherical models (Webster & Oliver, 2007). For each fixed effects matrix we calculated $\hat{\boldsymbol{\beta}}$, the estimate of $\boldsymbol{\beta}$, by ordinary least squares, subtracted $\mathbf{X}\hat{\boldsymbol{\beta}}$ from the data and used the method of moments (Matheron, 1962) to fit the vector of variogram parameters $\boldsymbol{\alpha} = (c_0, c_1, a, \nu)$. Following Cressie (1985), when fitting the variogram model to the experimental variogram by least squares we applied weights

$$w_i = \frac{N\left(h_i\right)}{\hat{\gamma}\left(h_i\right)^2}, \tag{3}$$

where $N(h_i)$ is the number of observation pairs within the bin centred on lag $h_i$ and $\hat{\gamma}(h_i)$ is the experimental variogram for $h_i$. Generally the REML estimator is recommended for fitting linear mixed models (Lark *et al.*, 2006) because there is known to be some bias when the method of moments is applied. However for 3000 observations this bias will be very small and does not justify the prohibitive computation time required by REML.

We fitted a number of linear mixed models with different fixed effects to the prediction data to compare (i) the effectiveness of including different auxiliary variables as fixed effects (ii) the effectiveness of fitting a single model for all soil types with the effectiveness of fitting different models for each of our three broad soil classes and (iii) how the precision of the different linear mixed models varies with the intensity of SOC observations. The variables to be included as fixed effects were selected according to the findings of our exploratory analyses.

Each linear mixed model was fitted first to all the prediction data. The linear mixed models assume that the random effects are Gaussian and therefore we log transform the data if the skew is greater than 1 and this assumption is implausible (Webster & Oliver, 2007). We note that strictly we should test the skew once we have subtracted the fixed effects from the data. However this could lead to the data being transformed for some but not all of the models so in order to make fair comparisons between the different models we decide whether to transform based upon the skew of the raw data. The entire prediction set had skew equal to 2.03. Therefore a log transform was applied and the skew was reduced to 0.92. The empirical best linear unbiased predictor (E-BLUP) was used to predict SOC content at the validation sites (Lark *et al.*, 2006). To calculate the mean prediction in the original units – and the mean squared error MSE between the predictions and observed SOC values at the validation sites – an unbiased inverse log transform (Cressie, 2004) was performed.

The prediction and validation data were then divided into the three soil classes according to our three-fold classification and separate models were fitted to the prediction data from each class and the appropriate soil class model was used to predict SOC content at each validation location by the E-BLUP. The skew for the SOC in mineral soil was 3.94 which was reduced to 0.80 by applying a log transform. No transform was applied to data from the organo mineral and peat soil classes because they had skew of 0.86 and -0.53 respectively.

The values of SOC are substantially larger on peat soils than mineral soils and sites at which the soil is mis-classified are likely to have large prediction errors which will dominate the MSE. We therefore also record the median square error (MdSE) and the MdSE based upon the back transform to the median prediction in the original units.

This test was then repeated using random subsets of $n_{\mathrm{s}}$ of the prediction data where $n_{\mathrm{s}} = 100, 200, \ldots 2900$, to explore how the quality of the predictions at the validation sites varies with the number of SOC observations.

## Results

### *Exploratory analysis*

The airborne radiometric estimates of K and Th were strongly correlated with their measurements in the soil survey (r=0.86 and 0.80 respectively; Table 2) which demonstrates that the airborne survey is an effective method for estimating these elements. Airborne radiometric estimates of K exhibited a strong negative correlation with ground-based estimates of SOC (r= -0.51); this correlation was stronger than for Th, U and total counts. We therefore chose to include radiometric K as a fixed effect in a linear mixed model of SOC. Furthermore, plots of radiometric K against SOC (not shown) suggested that this relationship may be nonlinear and therefore the square of radiometric K ($K^2$) was also included as a fixed effect.

We also explored the correlation of SOC with the two terrain parameters: altitude and compound topographic index (CTI). First we transformed the SOC data to a more Gaussian distribution by taking natural logarithms; the transformed variable had a skewness coefficient of 0.94. The correlation coefficients ($r$) between log SOC and altitude and CTI were, respectively: 0.6 and -0.03. Given the strong positive correlation between SOC and altitude we chose to include the latter as a fixed effect in the estimation of SOC.



The histogram of the SOC observations across all soil types in Northern Ireland is shown in Figure 2a. The distribution of SOC is bimodal. The main peak occurs between 0 and 10 % and there is a secondary peak around 50 %, which corresponds to peat soils. Bimodal distributions are not suited to geostatistical analyses and therefore this histogram vindicates our decision to analyse mineral, organo mineral and peat soils separately. We expect that mineral soils will have SOC less than 20 % and peat soils will have SOC greater than 20 %; this is the threshold adopted in Northern Ireland. According to the (1:50,000) soil map of Northern Ireland, 5552 (81.0 %) of the observations are from mineral soils, 382 (5.6 %) are from organo mineral soils and 915 (13.4 %) are from peat soils. Figures 2 b, c and d show the distribution of SOC for mineral, organo mineral and peat soils.

The largest peak in the organo mineral soils is for SOC less than 10 % but there is a secondary peak greater than 50 %. Thus a substantial proportion of our observations appear to be misclassified by the soil map. This is likely to be because the soil map is unsuitable for recognising very local variations in soil type. The soil classification is partially successful in separating the observations with large and small SOC and the secondary peak is not evident in the mineral soil distribution. However, some larger than expected SOC values remain in the mineral soil set (4 % of mineral soils have SOC greater than 20 %). The majority of peat soil observations have SOC greater than 40 % although 26 % of observations have SOC less than 20 %.

*Spatial analysis: linear mixed models*

The effectiveness of six different linear mixed models of SOC variation were compared. The fixed effects for these models were:

Model 1 Constant (i. e. the mean),

Model 2 Constant and altitude,

Model 3 Constant and K,

Model 4 Constant, K and $K^2$,

Model 5 Constant, altitude and K,

Model 6 Constant, altitude, K and $K^2$.

Figures 3-6 show the variograms fitted to residuals of all soils, mineral soils, organo mineral soils and peat soils respectively. On the all soils variogram the sill variance is largest when the fixed effects consist of a constant (the overall mean). The sill variance is reduced when altitude is also included in the fixed effects and further reductions are achieved by including radiometric K. The smallest variances are seen for Model 6. A similar pattern is seen on the mineral soils variogram although the semi-variances are substantially smaller for each model than the all soils variograms. For the organo mineral and peat variograms the sill variances are again largest with constant fixed effects and a greater reduction in these variances is achieved by including radiometric K rather than altitude. There is some evidence of spatial correlation in the Model 1 (constant mean) variograms on organo mineral and peat soils but the proportion of spatially correlated variation becomes smaller as other variables are added to the fixed effects. This indicates that the spatial correlation has been resolved by the inclusion of other variables in the fixed effects.

Figures 7-9 illustrate the different components of a single Model 6 linear mixed model fitted on all soils types. The log SOC observations are presented in Figure 7, the contributions from the fixed effects are presented in Figure 8 and the residuals between the observations and fixed effect contributions are presented in Figure 9. The fixed effects contain many of the large scale features of the observations such as large SOC values in the NE, SE, SW and a cluster slightly NW of centre. The residuals show less spatial structure, as would be expected from the almost all nugget variogram for Model 6 in Figure 3.

The MSEs between the observations and predictions at the validation sites are shown in Table 3 and the MdSEs in Table 4. We indicate the number of soil classes into which the prediction set is divided for fitting of the linear mixed model. The first two rows of each table show the errors upon predicting SOC over the entire validation set. In the remaining six rows the validation set is divided according to the soil classification and the errors for each classification are shown.

As variables are added to the fixed effects the MSEs and MdSEs generally decrease in the same pattern as the semi-variances in Figures 3-6 with the largest decreases occurring when radiometric K is added to the fixed effects. This illustrates that SOC is correlated with altitude but more effective information comes from radiometric K.

The MSEs and MdSEs are smaller when separate models are fitted to each soil class than when the soil classes are combined. For example the smallest MSE for the model fitted to the entire prediction set is 41.08 $\%^2$ whereas the smallest MSE when three separate models are fitted is 30.38 $\%^2$. The greatest improvement for three models over one model is seen for peat, particularly when the fixed effects include radiometric K. The MSEs for mineral soils when three models are fitted are slightly larger than those from a single model. We suspect that this is an artefact due to the large squared differences between predictions and observations which have been mis-classified as mineral. To illustrate this Table 5 contains the MSEs for mineral soils when only observations with SOC less than 20 % are included in the validation set. Removing the observations from mineral soils which we assume are mis-classified reduces the MSEs, and the MSEs for three models are now substantially less than those for one model. The MdSEs are consistently smaller than the MSEs further illustrating that the MSEs are dominated by classification errors.

Figure 9 compares root MSEs for SOC from Model 6 (the model which generally had the smallest MSE), Model 2 (the model with the smallest MSE of those models which did not include radiometric K as a fixed effect) and Model 1 (the model with no auxiliary information) against $n$. In each of these plots, separate models are fitted for each soil class. The results are combined to give the MSE across Northern Ireland in (a) and plots (b), (c) and (d) show the absolute differences over mineral, organo mineral and peat soils respectively. In each plot a substantial improvement upon including radiometric K in the fixed effects is evident. We also note that the precision of our predictions decreases very slowly as $n$ decreases.

## Discussion

The distribution of SOC will be bimodal over any study region which contains both organic and mineral soils. Thus if we wish to map SOC over such a region we must address the problem of applying geostatistics to bimodal distributions. By dividing the study region into three soil classes based on a (1:50,000) soil map of Northern Ireland we substantially improved the precision of SOC predictions across Northern Ireland. However a proportion of soils appeared to be mis-classified. The errors at these sites dominate the MSEs which are much larger than the corresponding MdSEs which are more robust to a proportion of large errors due to mis-classification. Therefore further improvements in regional predictions should be based upon improving our ability to differentiate peat and non peat soils. Some mis-classification may be inevitable because it is not practical to create soil maps which resolve very small scale deposits of organic or mineral soil. This classification may be significantly improved by using airborne hyperspectral data (McMorrow *et al.*, 2004) or satellite data (e.g. ASTER or Landsat).

Even without such improvements, the approach we describe could be applied to the data from the Tellus survey in combination with data on soil bulk density to improve current estimates of carbon pools across Northern Ireland, in which peat soils are estimated to account for more than 50% of the total (Cruickshank *et al.*, 1998).

Further work is required to elucidate the factors influencing the airborne estimates of radiometric K and its spatial correlation with SOC for each of the three soil types. The two dominant factors which influence this are: i) the variation in mineral-K content; this decreases with increasing quantities of soil organic matter, and ii) increasing soil-moisture resulting in greater attenuation of the gamma signal from the soil. This would require contemporaneous measurements of soil moisture content which we do not have from the original survey.

There are likely to be limitations to the widespread application of airborne radiometric data as a covariate for mapping SOC. First, it relies upon a spatial correlation between SOC and long-term (i.e. annual) soil moisture content which is only likely to occur in certain combinations of climate, topography and land use where SOC has accumulated above some minimum threshold. The soils of northern Europe include significant areas with soils common to those in northern Ireland, particularly the Gleysols and Histosols of northern Scandanavia, the Baltic States and Russia. Further work is required to establish the utility of airborne radiometric data as a covariate for mapping SOC, particularly for the large area of Podzols across northern Europe (European Soil Bureau Netowrk, 2005). Second, patterns of antecedent rainfall conditions and the quantity of precipitation during the airborne survey may cause unusually large temporal and spatial variations in soil moisture contents across the study area. This may reduce the degree of spatial correlation between SOC and the gamma radiation which is due to greater attenuation of the latter where the ground is wetter and where carbon accumulates. To address this potential limitation, it may be possible to ensure surveys are not flown when significantly atypical soil moisture conditions occur, based on antecedent rainfall data and medium-term precipitation forecasts. In addition, laboratory-based soil column experiments, in which in-situ measurements of gamma radiation are made under controlled soil moisture conditions for soils with range of SOC contents, could be used to calibrate the relationship between airborne radiometric data and SOC based on antecendent rainfall data. Finally, where the soil parent material contains very little K (0.5 %), such as quartzite, the accuracy of airborne radiometric estimation of soil-K may be insufficient. However, 90 % of European soils contain more than 0.83 % K (or 1 % $K_2O$; Salminen, 2005) suggesting that for the vast majority of soils, accuracy near the limit of detection is unlikely to be problematic.

One of the main applications of SOC maps is the estimation of carbon stocks. Where soil carbon is concentrated in the upper horizons, radiometry may be particularly useful as a covariate as it measures gamma radiation from the upper 35 cm of the solum. However, in deep organic-rich soils or areas of Arctic tundra where SOC may be transported to depth by cryoturbation (Ping *et al.*, 1997), the utility of gamma radiometry will be diminished because no information is provided for the deeper parts of the soil profile.

## Conclusions

Our results show that the precision of regional predictions of SOC across Northern Ireland are substantially improved by including auxiliary information on radiometric K from airborne surveys as fixed effects in a linear mixed model of SOC variation. To a lesser extent the precision is also improved by including altitude in the linear mixed model. We have also seen that the number of observations of SOC may be substantially reduced with little cost in terms of the precision of the predictions. However the MSEs between predictions and observations are still large even when radiometric K and altitude are included in the linear mixed model. This is because SOC in Northern Ireland has a bimodal distribution which is not suited to geostatistical analyses. After separating the soil into three main classes (mineral, organo mineral and organic), the MSEs are substantially reduced. However, large independent validation errors occur at certain mis-classified sites where, for example, the soil map shows a mineral soil but its SOC content indicates it is organic. Improvements in our ability to differentiate between mineral and organic-rich soils are required to make better predicitions of the SOC at the regional scale.

## Acknowledgements

The Tellus surveys were financed in Northern Ireland by the the Department of Enterprise, Trade and Investment and the 'Building Sustainable Prosperity' scheme of the Rural Development Programme. Topographic data based upon Ordnance Survey of Northern Irelands ® data with the permission of the Controller of Her Majesty's Stationery Office, ©Crown copyright and database rights License Number DMOU205. Rainfall data for Northern Ireland were provided from the MIDAS database of the UK Meteorological Office. This paper is published with the permission of the Director of the British Geological Survey (Natural Environment Research Council) and the Director of the Geological Survey of Northern Ireland. B.P. Marchant and R.M. Lark's contributions were part of Rothamsted's program in Mathematical and Computational Biology which is funded by the Biotechnology and Biological Sciences Research Council of the United Kingdom. The authors wish to acknowledge the contributions of all staff from the Geological Survey of Northern Ireland (including Dr C. Van Dam) and the British Geological Survey who were involved in the organisation, collection, analysis or processing of the airborne radiometric and soil survey data from the Tellus survey.

## References

Broadbent, F.E. 1953. The soil organic fraction. *Advances in Agronomy*, **5**, 153–183. York.

Carroll, T.R. 1981, Airborne soil moisture measurement using natural terrestrial gamma radiation. *Soil Science*, **132**, 358–366.

Chen, F., Kissel, D.E., West, L.T., Rickman, D., Luvall, J.C. & Adkins, W. 2005.

Mapping surface soil organic carbon for crop fields with remote sensing. *Journal Of Soil And Water Conservation*, **60**, 51–57.

Cook, S.E., Corner, R.J., Groves, P.R. & Grealish, G. J. 1996. Use of airborne gamma radiometric data for soil mapping. *Australian Journal of Soil Research*, **34**, 183–194.

Cressie, N. 2006. Block kriging for lognormal spatial processes. *Mathematical Geology*, **38**, 413–443.

Cruickshank, M M, Tomlinson, R W, Devine, P M, & Milne, R M. 1998. Carbon in the vegetation and soils of Northern Ireland. *Biology and Environment: Proceedings of the Royal Irish Academy*, **98**, 9–21.

Cruickshank, J.G. 1997. *Soil and Environment: Northern Ireland.* Queen's University of Belfast, Belfast.

European Soil Bureau Network, 2005. *Soil Atlas of Europe* Office for official publicaitons of the European Communities, Luxembourg.

Grasty, R.L. & Minty, B.R.S. 1995. *A guide to the technical specifications for airborne gamma-ray surveys.* Australian Geological Survey Organisation: Record 1995/20.

IAEA, 1991. *Airborne gamma ray spectrometer surveying.* International Atomic Energy Agency, Technical Report Series, No. 323.

Jordan, C. & Higgins, A. 2007. *A generalised map of major soil types for Northern Ireland at a scale of 1:250,000.* Agri- Food and Biosciences Institute, Belfast. Based on Cruickshank, J.G. 1997. *Soil and Environment: Northern Ireland.* Queen's University of Belfast, Belfast.

Lark, R.M., Cullis, B.R. & Welham, S.J. 2006. On spatial prediction of soil properties in the presence of a spatial trend: the empirical best linear unbiased predictor

(E-BLUP) with REML. *European Journal of Soil Science*, **57**, 787–799.

Lilja, H. & Nevalainen, R. 2005. Developing a digital soil map for Finland. In: *Digital Soil Mapping: An Introductory Perspective* eds P. Lagacherie, A.B McBratney & M. Voltz. pp. 67-74. Elsevier, Amsterdam, The Netherlands.

Lunden, B., Wang, G. & Wester, K. 2001. A GIS based analysis of data from Landsat TM, airborne geophysical measurements, and digital maps for geological remote sensing in the Stockholm region, Sweden. *International Journal of Remote Sensing*, **22**, 517–532.

Matérn, B. 1960. Spatial variation. Meddelanden från Statens Skogsforskningsinstitut, 49. [2nd Edition (1986), Lecture Notes in Statistics, No. 36, Springer, New York].

McBratney, A.B. & Webster, R. 1983. Optimal Interpolation and Isarithmic Mapping of Soil Properties.5. Co-Regionalization and Multiple Sampling Strategy. *Journal of Soil Science* **34**, 137–162.

McMorrow, M.J.M., Cutler, M.E.J., Evans, M. G. & Al-Roichdi, A. 2004. Hyperspectral indices for characterizing upland peat composition. *International Journal of Remote Sensing*, **25**, 313–325.

Minasny, B., McBratney, A.B., Mendonca-Santos, M.L., Odeh, I.O.A. & Guyon, B. 2006. Prediction and digital mapping of soil carbon storage in the Lower Namoi Valley. *Australian Journal Of Soil Research*, **44**, 233–244.

Minty, B.R.S. 1997. Fundamentals of airborne gamma-ray spectrometry. *AGSO Journal of Australian Geology and Geophysics*, **17**, 39–50.

Moore, I.D., Gessler, P.E., Nielsen, G.A., and Petersen, G.A. 1993. Terrain attributes: estimation methods and scale effects. In , (ed A.J. Jakeman, M.B. Beck & M. McAleer), pp. 189 - 214. John Wiley & Sons, Chichester.

Moore, I.D., Grayson, R.B. & Ladson, A. R. 1991. Digital Terrain Modelling: a Review of Hydrological, Geomorphological and Biological Applications. *Hydrological Processes*, **5**, 3-30.

Mueller, T.G. & Pierce, F.J. 2003. Soil carbon maps: enhancing spatial estimates with simple terrain attributes at multiple scales. *Soil Science Society of America Journal*, **67**, 258–627.

Odeh, I.O.A., McBratney, A.B. & Chittleborough, D.J. 1995. Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression kriging. *Geoderma*, **67**, 215–226.

Ping, C.L., Michaelson, G.J. & Kimble, J.M. 1997. Carbon storage along a lattiudinal transect in Alaska. *Nutrient Cycling in Agroecosystems*, **49**, 235–242.

Pitkin, J.A. & Duval, J.S. 1980. Design parameters for aerial gamma surveys. *Geophysics*, **45**, 1427–1439.

Rawlins, B.G., Webster, R. & Lark, R.M. 2007. Understanding airborne radiometric survey signals across part of eastern England. Earth Surface Processes and Landforms, **32**, 1503–1515.

Saliminen, R. 2005. *Geochemical Atlas of Europe: Part 1.* Geological Survey of Finland, Espoo.

Scottish Executive, 2007. ECOSSE Estimating Carbon in Organic Soils Sequestration and Emissions. http://www.scotland.gov.uk/Resource/Doc/170721/0047848.pdf. Accessed 20th September 2007.

Selige, T., Bohner, J. & Schmidhalter, U. 2006. High resolution topsoil mapping using hyperspectral image and field data in multivariate regression modeling procedures. *Geoderma*, **136**, 235–244.

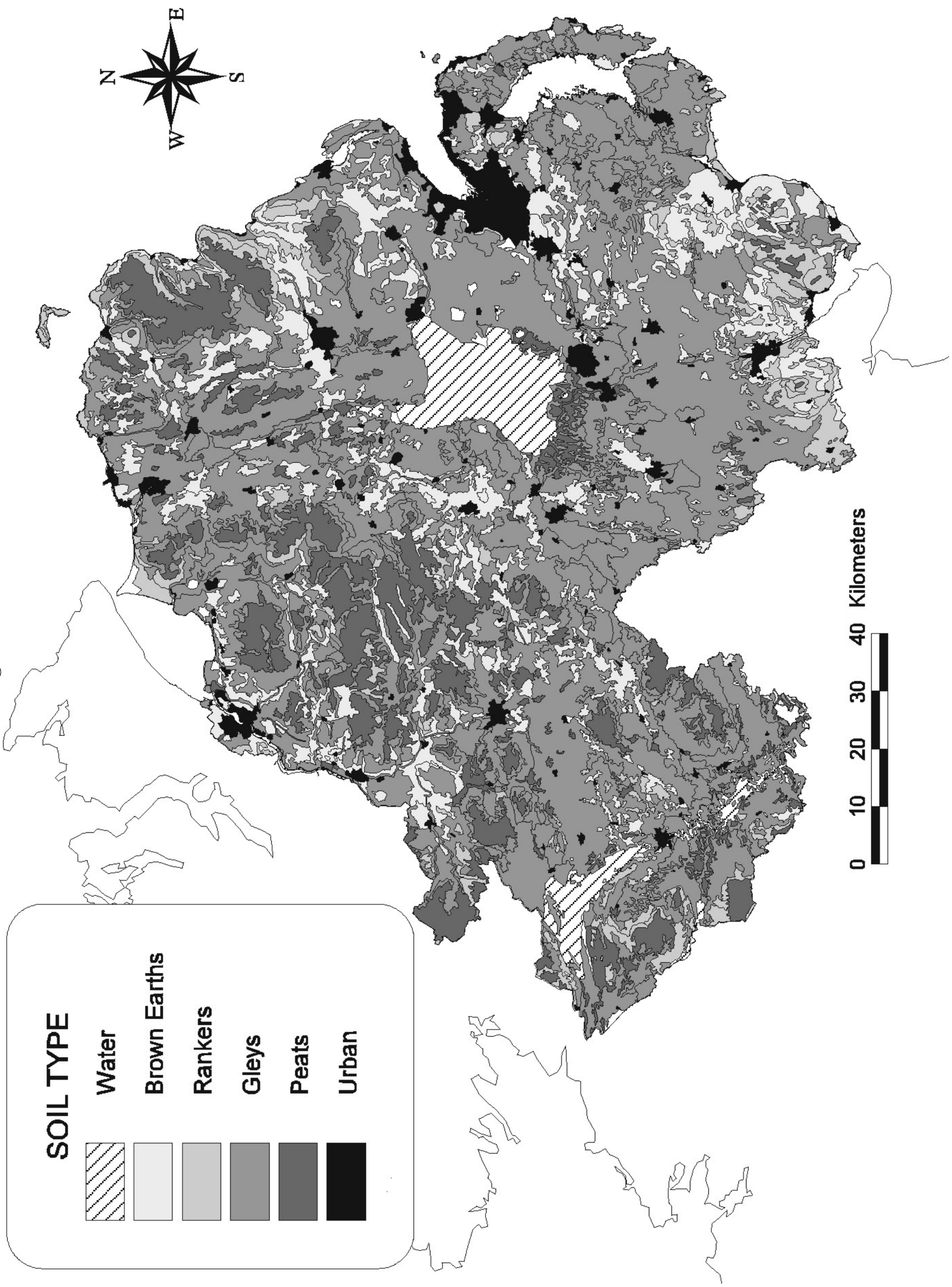
Simbahan, G.C. & Dobermann, A. 2006. Sampling optimization based on secondary information and its utilization in soil carbon mapping. *Geoderma*, **133**, 345–362.

Smith, L.C., MacDonald, G.M., Velichko, A.A., Beilam, D.W., Borisova, O.K., Frey, K.E., Kremenetski, K.V. & Sheng, Y. 2004. Siberian peatlands a net carbon sink and global methane source since the early Holocene. *Science*, **303**, 353–356.

Webster, R. & Oliver, M.A. 2007. *Geostatistics for Environmental Scientists*: second edition. John Wiley & Sons, Chichester.

### List of Figures and Captions

Figure 1:


N
E
S
W
0 10 20 30 40 Kilometers
SOIL TYPE
Water
Brown Earths
Rankers
Gleys
Peats
Urban

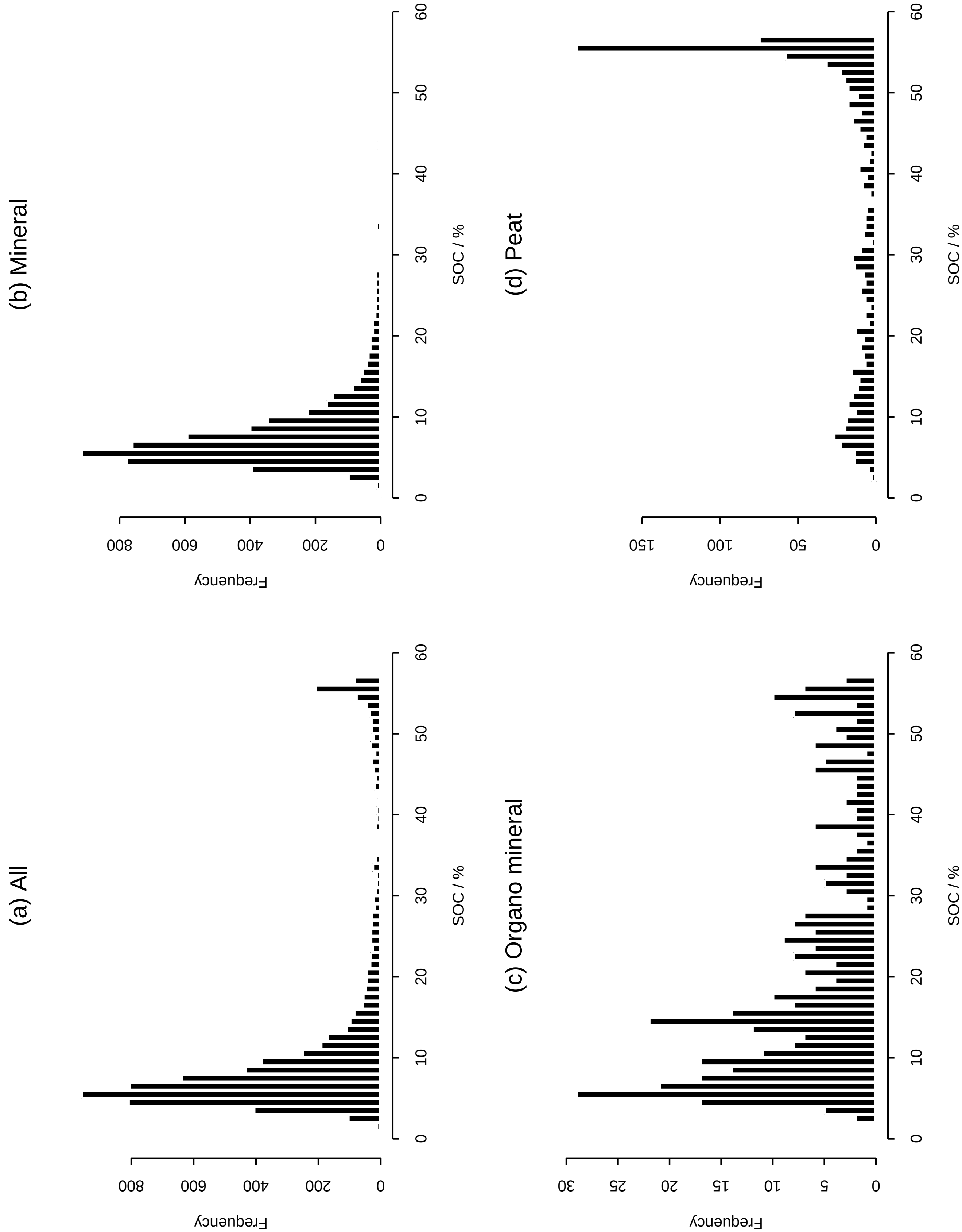



Figure 2:

Figure 3:

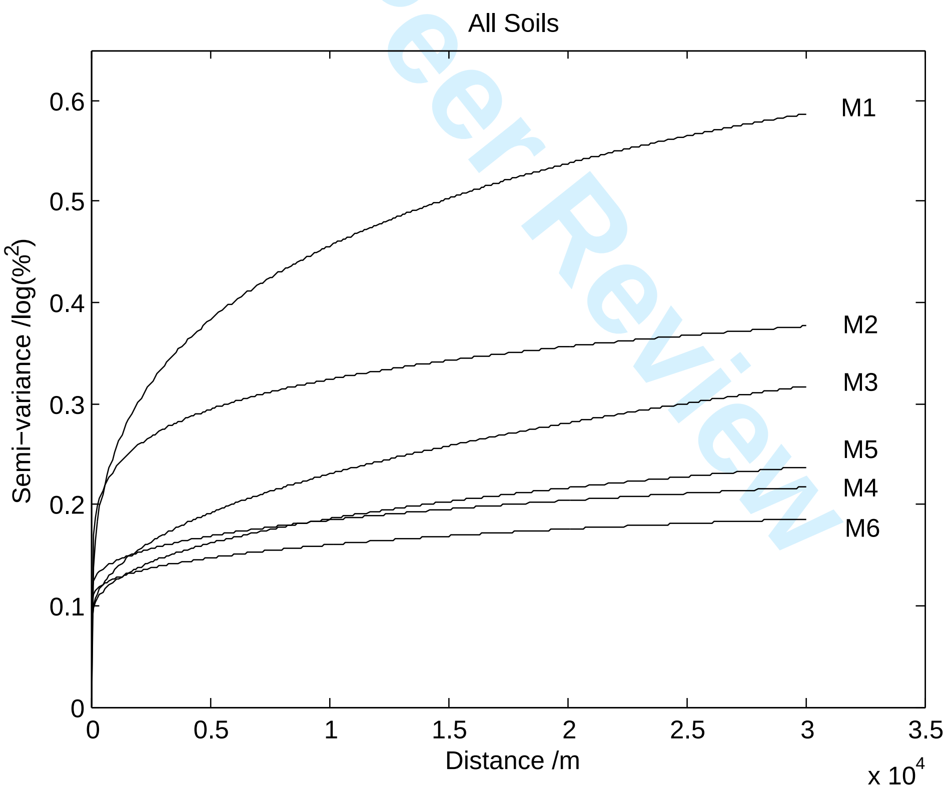

All Soils
M1
M2
M3
M5
M4
M6
0
0.1
0.2
0.3
0.4
0.5
0.6
Semi−variance /log(%$^2$)
0
0.5
1
1.5
2
2.5
3
3.5
Distance /m
x 10$^4$

Figure 4:

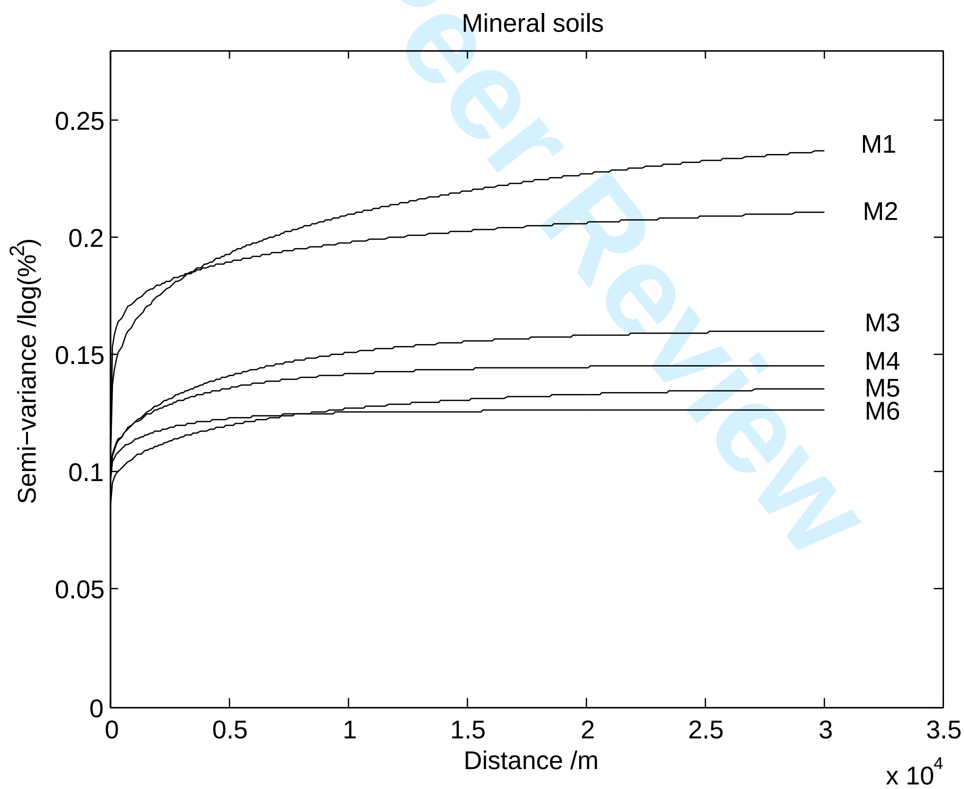

Mineral soils
Semi−variance /log(%$^2$)
Distance /m
x $10^4$
M1
M2
M3
M4
M5
M6
0
0.05
0.1
0.15
0.2
0.25
0
0.5
1
1.5
2
2.5
3
3.5

Figure 5:

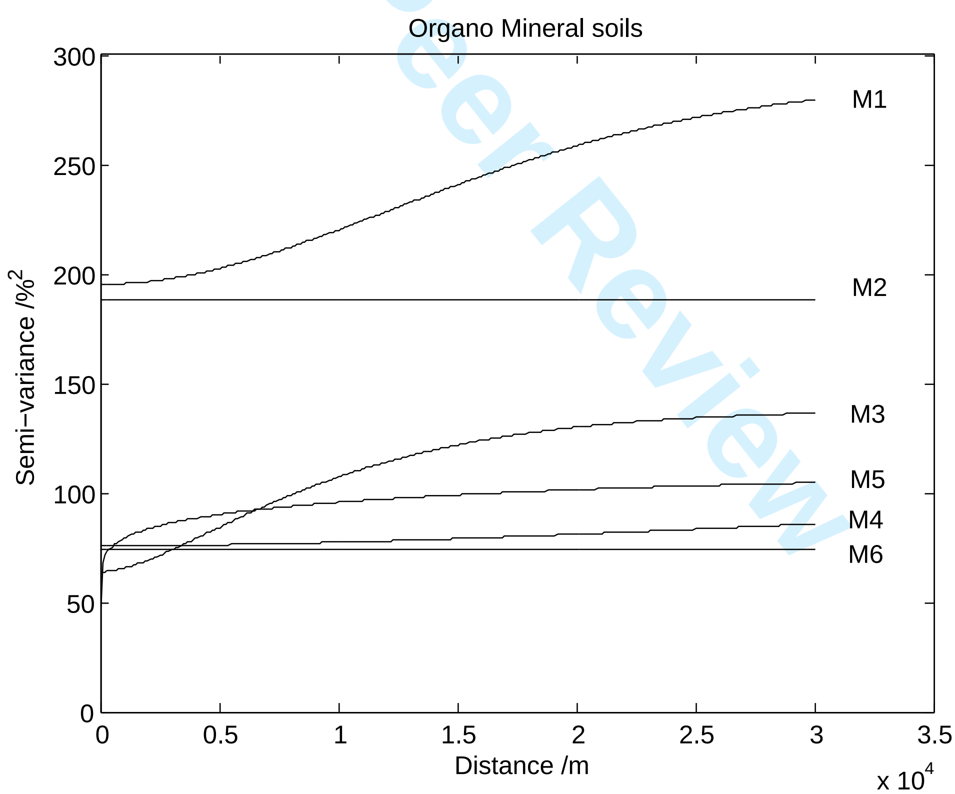
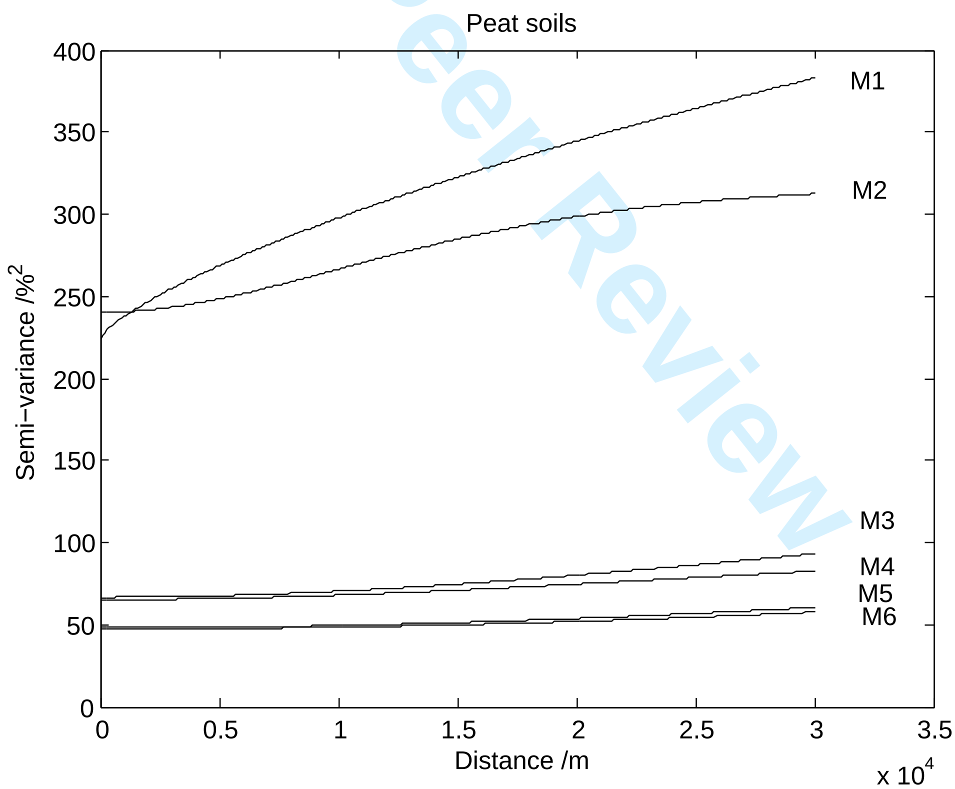

Organo Mineral soils
Semi−variance /%$^2$
Distance /m
x 10$^4$
0
50
100
150
200
250
300
0
0.5
1
1.5
2
2.5
3
3.5
M1
M2
M3
M5
M4
M6

Figure 6:


Peat soils
Semi−variance /%2
Distance /m
x 10^4
400
350
300
250
200
150
100
50
0
0
0.5
1
1.5
2
2.5
3
3.5
M1
M2
M3
M4
M5
M6

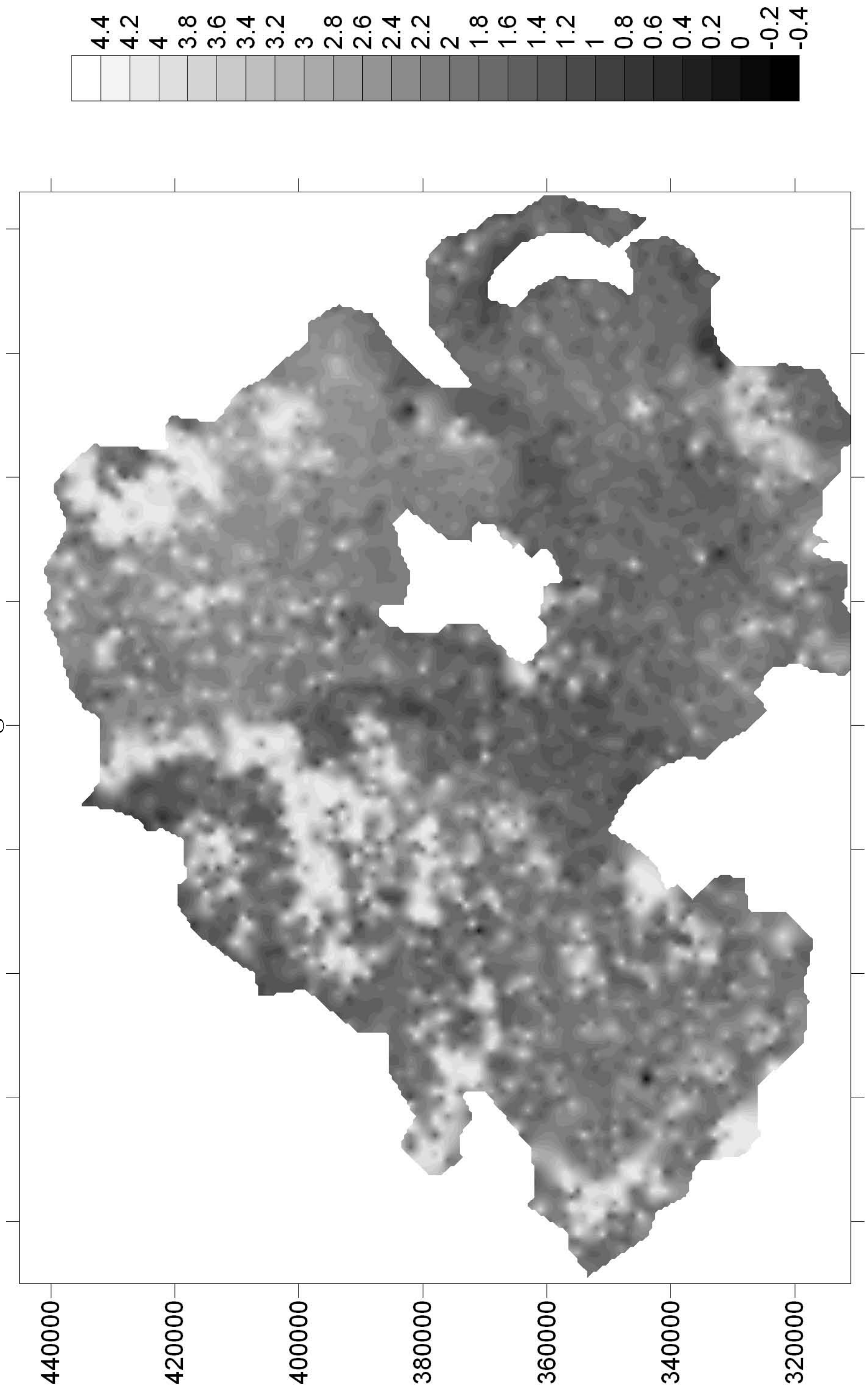

4.4
4.2
4
3.8
3.6
3.4
3.2
3
2.8
2.6
2.4
2.2
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
200000
220000
240000
260000
280000
300000
320000
340000
360000
440000
420000
400000
380000
360000
340000
320000

Figure 7:

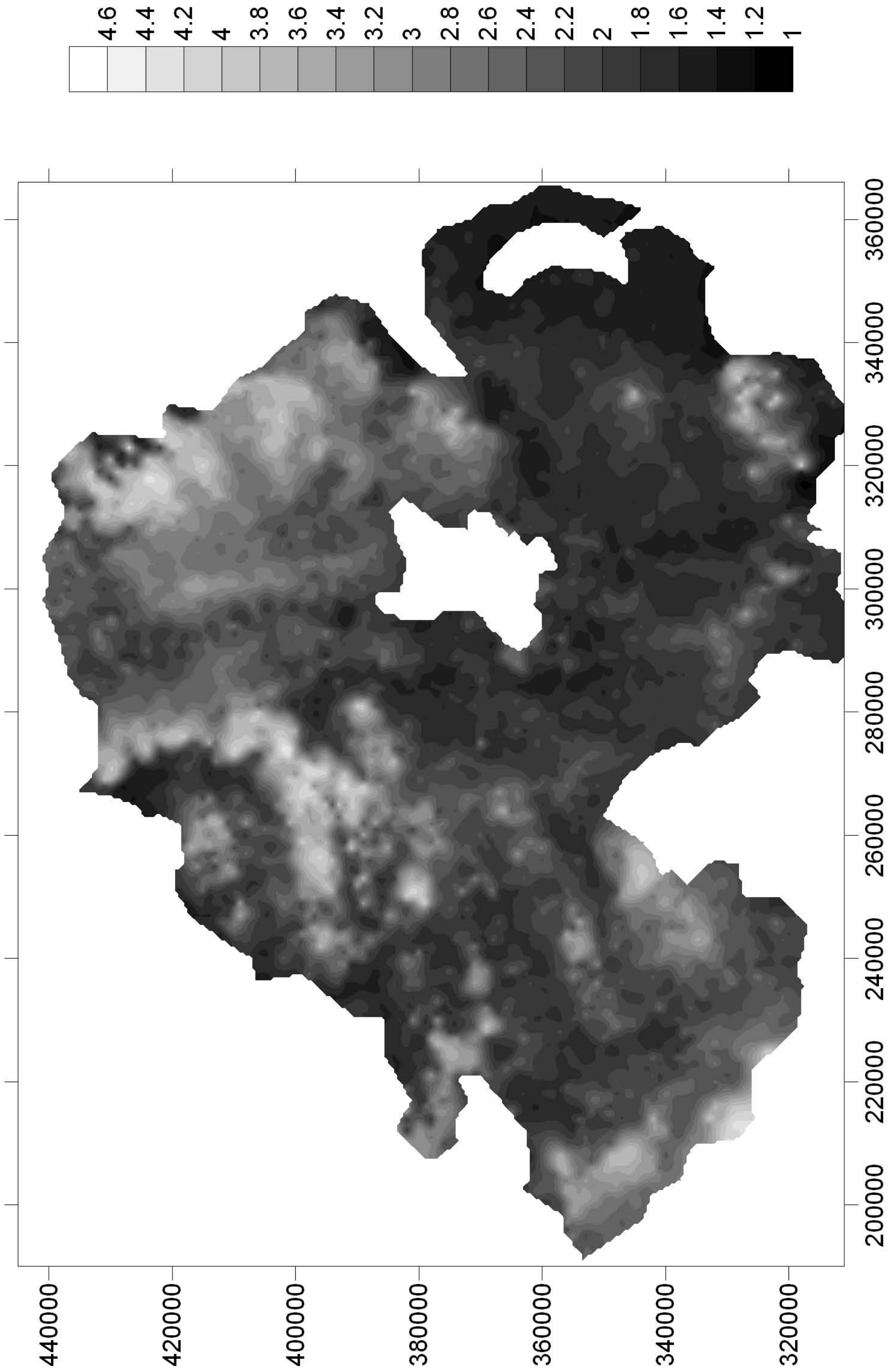

4.6
4.4
4.2
4
3.8
3.6
3.4
3.2
3
2.8
2.6
2.4
2.2
2
1.8
1.6
1.4
1.2
1
200000
220000
240000
260000
280000
300000
320000
340000
360000
440000
420000
400000
380000
360000
340000
320000

Figure 8:

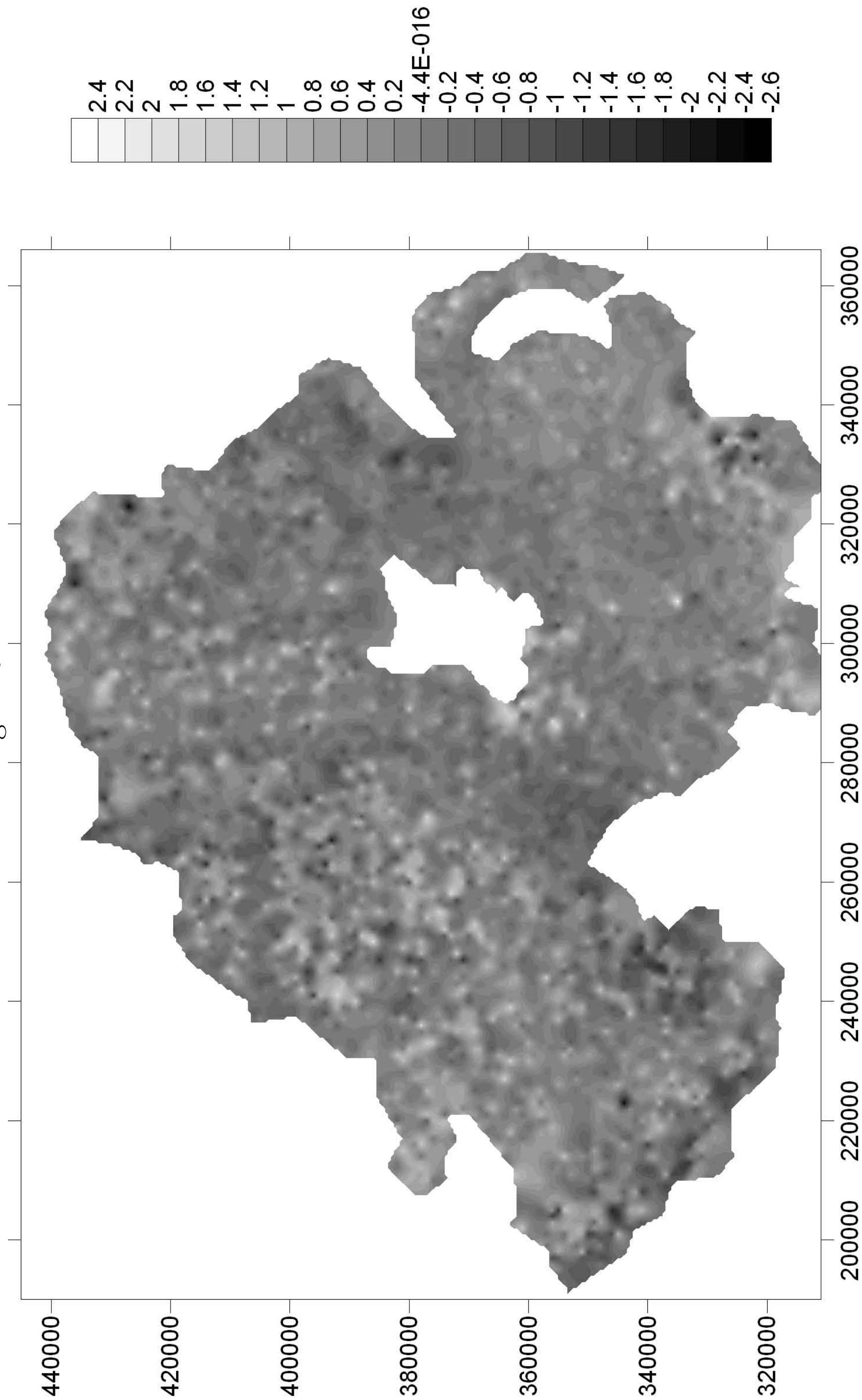

2.4
2.2
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
-4.4E-016
-0.2
-0.4
-0.6
-0.8
-1
-1.2
-1.4
-1.6
-1.8
-2
-2.2
-2.4
-2.6
440000
420000
400000
380000
360000
340000
320000
200000
220000
240000
260000
280000
300000
320000
340000
360000

Figure 9:

Figure 10:

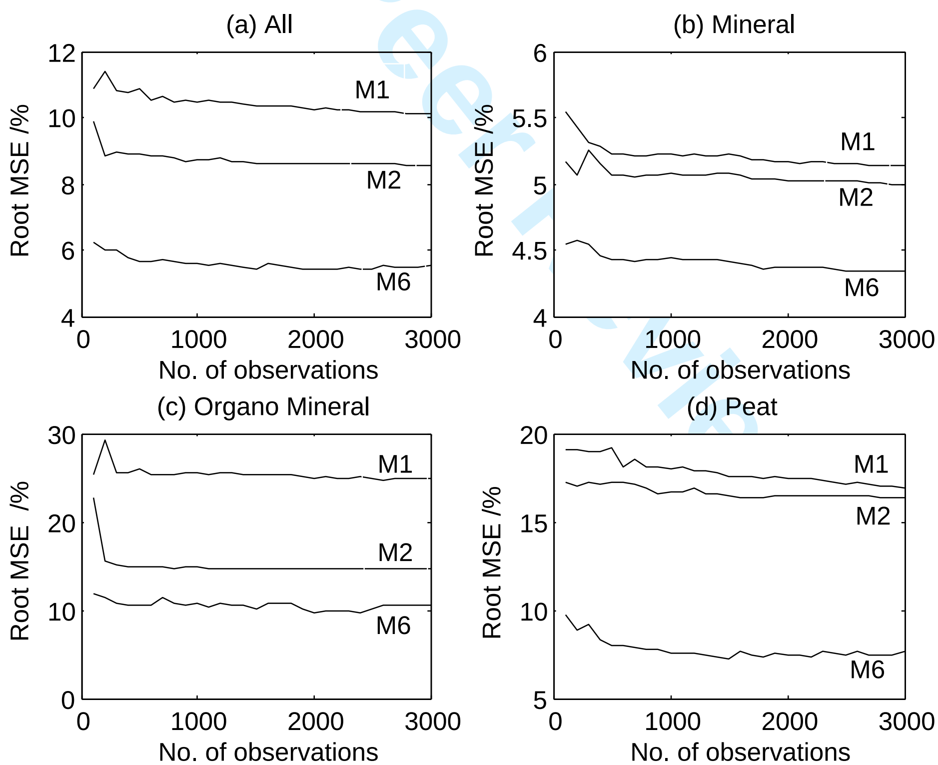

(a) All
(b) Mineral
(c) Organo Mineral
(d) Peat
M1
M2
M6
Root MSE /%
No. of observations
0
1000
2000
3000
4
6
8
10
12
4.5
5
5.5
0
10
20
30
15

**Table 1** Summary statistics for the soil and the nearest radiometric survey location (n=6862). Units are % for K and soil organic carbon (SOC), mg $kg^{-1}$ for Th and U. By SS we denote the soil geochemical survey data, and by Rad the radiometric data.

| Element | K | | Th | | U | | |
|---|---|---|---|---|---|---|---|
| Dataset | SS-K | Rad-K | SS-Th | Rad-Th | SS-U | Rad-U | SS-SOC |
| Mean | 1.38 | 0.91 | 5.04 | 3.44 | 2.48 | 0.79 | 13.57 |
| Median | 1.47 | 0.89 | 5.00 | 3.19 | 2.30 | 0.68 | 7.49 |
| Variance | 0.52 | 0.38 | 8.10 | 7.96 | 8.10 | 0.64 | 212 |
| Standard deviation | 0.72 | 0.61 | 3.00 | 2.82 | 2.85 | 0.80 | 14.6 |
| Skewness | 0.07 | 0.50 | 2.48 | 2.52 | 29.2 | 2.82 | 1.99 |
| $\mathrm{Log}_e$ skewness | | | | | | | 0.94 |

**Table 2** Correlation matrix for K, Th, U and soil organic carbon (SOC) for the soil survey (SS) data and the nearest neighbouring radiometric data (Rad).

| | Dataset and element | | | | | | |
|---|---|---|---|---|---|---|---|
| | SS-K | SS-Th | SS-U | SS-SOC | Rad-K | Rad-Th | Rad-U |
| SS-K | 1 | | | | | | |
| SS-Th | 0.75 | 1 | | | | | |
| SS-U | 0.18 | 0.45 | 1 | | | | |
| SS-SOC | -0.63 | -0.45 | -0.03 | 1 | | | |
| Rad-K | 0.86 | 0.71 | 0.25 | -0.51 | 1 | | |
| Rad-Th | 0.67 | 0.80 | 0.39 | -0.36 | 0.79 | 1 | |
| Rad-U | 0.51 | 0.66 | 0.36 | -0.29 | 0.60 | 0.70 | 1 |

**Table 3** MSEs for SOC predictions from Models 1-6 (M1- M6) at all validation sites and at validation sites classified as mineral, organo mineral (O M) and peat soils. Units are $\%^2$.

| Soils | Soil classes[a] | M1 | M2 | M3 | M4 | M5 | M6 |
|---|---|---|---|---|---|---|---|
| All | 1 | 119.42 | 116.67 | 60.18 | 41.08 | 67.69 | 49.37 |
| | 3 | 78.70 | 73.14 | 36.92 | 30.60 | 34.24 | 30.38 |
| Mineral | 1 | 34.45 | 30.11 | 17.74 | 14.95 | 17.89 | 15.44 |
| | 3 | 26.34 | 24.95 | 20.97 | 19.57 | 19.79 | 18.85 |
| O M | 1 | 215.24 | 319.65 | 107.04 | 93.01 | 139.65 | 108.08 |
| | 3 | 238.53 | 216.96 | 127.05 | 102.98 | 112.88 | 110.74 |
| Peat | 1 | 523.05 | 485.37 | 262.17 | 156.18 | 298.53 | 202.27 |
| | 3 | 286.82 | 265.99 | 83.69 | 58.82 | 77.79 | 58.04 |

[a] The number of soil classes into which the prediction set is divided when fitting the linear mixed model.

**Table 4** MdSDs for SOC predictions from Models 1-6 (M1-M6) at all validation sites and at validation sites classified as mineral, organo mineral (O M) and peat soils. Units are $\%^2$.

| Soils | Soil classes[a] | M1 | M2 | M3 | M4 | M5 | M6 |
|---|---|---|---|---|---|---|---|
| All | 1 | 4.77 | 4.69 | 3.37 | 2.99 | 3.41 | 3.01 |
| | 3 | 3.22 | 3.03 | 2.45 | 2.20 | 2.38 | 2.26 |
| Mineral | 1 | 2.79 | 2.84 | 1.96 | 1.91 | 2.01 | 1.97 |
| | 3 | 1.94 | 1.83 | 1.65 | 1.53 | 1.51 | 1.51 |
| O M | 1 | 51.00 | 40.13 | 22.83 | 22.85 | 21.07 | 18.71 |
| | 3 | 102.37 | 93.12 | 49.17 | 37.81 | 39.57 | 45.14 |
| Peat | 1 | 406.92 | 282.18 | 169.17 | 101.81 | 130.21 | 99.70 |
| | 3 | 192.38 | 190.50 | 27.70 | 12.07 | 26.33 | 10.06 |

[a] The number of soil classes into which the prediction set is divided when fitting the linear mixed model.

**Table 5** MSEs for SOC predictions from Models 1-6 (M1- M6) on validation sites classified as mineral soils when observations greater than 20 % are removed from the validation set. Units are $\%^2$.

| Soils | Soil classes[a] | M1 | M2 | M3 | M4 | M5 | M6 |
|---|---|---|---|---|---|---|---|
| Mineral | 1 | 21.85 | 15.73 | 10.60 | 8.94 | 8.81 | 8.00 |
| | 3 | 6.93 | 6.71 | 5.87 | 5.83 | 5.69 | 5.76 |

[a] The number of soil classes into which the prediction set is divided when fitting the linear mixed model.