DISAGGREGATION OF LEGACY SOIL DATA USING AREA TO POINT KRIGING FOR MAPPING SOIL ORGANIC CARBON AT THE REGIONAL SCALE

Ruth Kerry[1, 2], Pierre Goovaerts[3], Barry G. Rawlins[4], Ben Marchant[5]

[1]Department of Geography, Brigham Young University, Provo, UT, USA and
[2]Department of Geography, University of Cambridge, Cambridge, UK. – visiting scholar
Email: ruth_kerry@byu.edu* corresponding author
[3]Biomedware Inc., 3526 W. Liberty Road. Suite 100, Ann Arbor, MI, USA.
[4]British Geological Survey, Keyworth, Nottingham, UK.
[5]Biomathematics and Bioinformatics Department, Rothamsted Research, Harpenden, UK

Abstract

Legacy data in the form of soil maps, which often have typical property measurements associated with each polygon, can be an important source of information for digital soil mapping (DSM). Methods of disaggregating such information and using it for quantitative estimation of soil properties by methods such as regression kriging (RK) are needed. Several disaggregation processes have been investigated; preferred methods include those which include consideration of scorpan factors and those which are mass preserving (pycnophylactic) making transitions between different scales of investigation more theoretically sound. Area to Point Kriging (AtoP kriging) is pycnophylactic and here we investigate its merits for disaggregating legacy data from soil polygon maps. Area to Point Regression Kriging (AtoP RK) which incorporates ancillary data into the disaggregation process was also applied. The AtoP kriging and AtoP RK approaches do not involve collection of new soil measurements and are compared with disaggregation by simple rasterization. Of the disaggregation methods investigated, AtoP RK gave the most accurate predictions of soil organic carbon (SOC) concentrations (smaller mean absolute errors (MAEs) of cross-validation) for disaggregation of soil polygon data across the whole of Northern Ireland.

The legacy soil polygon data disaggregated by AtoP kriging and simple rasterization were used in a RK framework for estimating soil organic carbon (SOC) concentrations across the whole of Northern Ireland, using soil sample data from the Tellus survey of Northern Ireland and with other covariates (altitude and airborne radiometric potassium). This allowed direct comparison with previous analysis of the Tellus survey data. Incorporating the legacy data, whether from simple rasterization of the polygons or AtoP kriging, substantially reduced the MAEs of RK compared with previous analyses of the Tellus data. However, using legacy data disaggregated by AtoP kriging in RK resulted in a greater reduction in MAEs. A jack-knife procedure was also performed to determine a suitable number of additional soil samples that would need to be collected for RK of SOC for the whole of Northern Ireland depending on the availability of ancillary data. We recommend i) if only legacy soil map data are available, they should be disaggregated using AtoP kriging, ii) if ancillary data are also available legacy data should be disaggregated using AtoP RK and iii) if new soil measurements are available in addition to ancillary and legacy soil map data, the legacy soil map data should be first

disaggregated using AtoP kriging and these data used along with ancillary data as the fixed effects for RK of the new soil measurements.

Keywords: Digital soil mapping, Legacy soil data, Area to Point kriging, Regression kriging, Soil organic carbon, Disaggregation

1. Introduction

In traditional soil survey, surveyors use their knowledge of soil forming factors combined with aerial photographs and field-based soil observations to delineate soil classes as polygons on a map. Typical soil profiles of these classes are often described and published as a memoir which may include values of key soil properties at various depths. This traditional approach gives no indication of variability in these soil properties within or between classes and it has no statistical basis which can lead to bias (Carré et al. 2007a). In the last 10-15 years the need for raster based digital soil maps has been emphasized and a digital soil mapping (DSM) approach has been developed (McBratney et al. 2003). Such maps have pixels of different size depending on the scale of interest and have values of key soil properties, such as soil organic carbon (SOC) concentration, available at several depths (McBratney et al. 2003). More specifically, DSM has been defined as the creation and population of spatial soil information by the use of field and laboratory observational methods coupled with spatial and non-spatial soil inference systems (Lagacherie and McBratney, 2007; Carré et al. 2007b).

Polygon maps representing soil classes at various levels of national or international classification systems exist in many locations. Effective methods are required for the disaggregation and incorporation of such a wealth of 'legacy soil data' into DSM at national and regional scales. Appropriate use of this historical data could ensure that additional sampling effort associated with modern digital soil mapping is minimized. De Bruin et al. (1999) and Eagleson et al. (1999) proposed approaches that use hierarchical spatial reasoning for disaggregation of soil polygons. Bui and Moran (2001) investigated some such methods empirically along with other methods for spatial disaggregation of soil polygon maps in the Murray-Darling basin, Australia.

The theoretical merits of several forms of spatial disaggregation were investigated by McBratney (1998); the author suggested transfer functions and pycnophylactic interpolation should be applied. Mass preservation in pycnophylactic methods (Tobler, 1979) means that the mass or values over all the finer pixels contained within a polygon is preserved; in other words the average of the values in the finer pixels gives the polygon value. McBratney (1998) noted that the mass preservation property of pycnophylactic splines is a useful feature for disaggregating soil data as it could make transitions between scales in DSM more sensible. This could mean that intense sampling at each scale is not essential. Mass preservation is also a feature of the recently developed geostatistical approach of Area to Point Kriging (Kyriakidis, 2004). The typical centroid-based approach to kriging from areas to points assumes that the spatial support of units is constant (Goovaerts, 2006). Hence it is not appropriate for use with polygon data of

varying shape and size (Gotway and Young, 2002). The advantage of AtoP kriging is that it incorporates the variable size and shape of polygons in variogram deconvolution and kriging. Recently, Goovaerts (2010, 2011) used AtoP and regression kriging (RK) to incorporate both point field measurements and areal data (calibration of geological map) in the spatial interpolation of heavy metals in the Swiss Jura. Sensitivity analysis indicated that these new kriging procedures improve prediction over ordinary kriging and traditional RK based on the assumption that the local mean is constant within each mapping unit. To our knowledge, the advantages associated with AtoP kriging have not been used for disaggregating legacy soil maps and for optimizing DSM or compared to current state-of-the-art methods in DSM.

Current DSM methods in more data-rich settings include RK (McBratney et al. 2003) to map variation in important soil properties such as organic carbon based on ancillary data. Soil organic carbon is arguably one of the most important soil properties due to the benefits it confers such as enhancing soil structure through aggregation, improved water holding and cation-exchange capacities and also acting as a store of terrestrial carbon. In a recent study, two types of ancillary data (altitude and airborne radiometric measurements of potassium) were shown to be effective for improved mapping of soil organic carbon across all of Northern Ireland within a RK framework (Rawlins et al., 2009). However, the authors did not incorporate disaggregated legacy data into their procedure.

In this study we use Area to Point (AtoP) kriging to disaggregate soil organic carbon (SOC) data from a polygon map and compare it with disaggregation by simple rasterization of the same data. We also compare these methods with an AtoP regression kriging (AtoP RK) which includes some hierarchical spatial reasoning in the disaggregation process (Liu et al., 2008; Yoo and Kyriakidis, 2009). In this approach, ancillary data are used to inform on within-class variation in key scorpan factors (McBratney et al. 2003) such as relief and parent material, i.e. they provide a local mean and the residuals are AtoP kriged. The errors involved with each disaggregation approach are investigated. We then use the same regression models and data as Rawlins et al. (2009), but we add legacy map SOC data disaggregated by simple rasterization (Polygon SOC) and AtoP kriging (AtoP SOC) as extra fixed effects in RK. This two-step approach (Liu et al. 2008) to incorporating legacy data into RK was used to allow direct comparison with the results of Rawlins et al. (2009), however, there is no guarantee that the pycnophylactic or mass preserving property of AtoP kriging is preserved with a two-step AtoP RK procedure. Goovaerts (2010, 2011) introduced an approach where point and areal data are incorporated in one-step (i.e. one kriging system solved) instead of the two-step approaches (AtoP kriging followed by RK kriging) used here. Although this methodology, coined Area-And-Point (AAP) kriging, is theoretically more efficient than a two-step approach, and is pycnophylactic, it is not currently available in commercial software.

The errors associated with incorporating Polygon SOC and AtoP SOC into RK using the six models of Rawlins et al. (2009) are investigated and a suitable number of samples for mapping SOC across Northern Ireland based on the available covariates is suggested. We

comment on the benefits for DSM of disaggregating data from legacy soil polygon maps using simple rasterization, AtoP kriging and AtoP RK and incorporating the former two types of disaggregated legacy data into RK.

2. Geostatistical Theory

2.1. Area to Point Kriging

Consider the problem of estimating the value of a soil property z at any location $\mathbf{u}_s$ within a study area A from a set of B areal data $\{z(v_\beta); \beta=1,\ldots,B\}$. These areal or legacy soil polygon map data are typically measured on spatial supports (mapping units) v_β of various size and shape. Area-to-Point (AtoP) kriging can be viewed as the counterpart of block kriging in that point estimates $z^*_{AtoP}(\mathbf{u}_s)$ are obtained as the following linear combination of areal (block) measurements:

$$z^*_{AtoP}(\mathbf{u}_s) = \sum_{k=1}^{K} \lambda_k(\mathbf{u}_s) z(v_k), \tag{1}$$

where K is typically smaller than the total number of areal data B; for example (K-1) is the number of blocks adjacent to the block v_β where the point estimation is conducted. The kriging weights are the solution of the following ordinary kriging system:

$$\begin{aligned} &\sum_{k'=1}^{K} \lambda_{k'}(\mathbf{u}_s)\, \overline{C}(v_k, v_{k'}) + \mu(\mathbf{u}_s) = \overline{C}(v_k, \mathbf{u}_s) \qquad \mathrm{k} = 1,\ldots,K \\ &\sum_{k'=1}^{K} \lambda_{k'}(\mathbf{u}_s) = 1, \end{aligned} \tag{2}$$

where $\mu(\mathbf{u}_s)$ is the Lagrange multiplier. Like in traditional block kriging, the block-to-point covariances $\overline{C}(v_k, \mathbf{u}_s)$ are approximated by the average of the point support covariance $C(\mathbf{h})$ computed between the location $\mathbf{u}_s$ and a set of P_k points discretizing the block v_k (Figure 1a). A similar procedure is used for the block-to-block covariances $\overline{C}(v_k, v_{k'}) = \mathrm{Cov}\{Z(v_k), Z(v_{k'})\}$ and involves averaging $C(\mathbf{h})$ computed between any two points discretizing the blocks v_k and $v_{k'}$ (Figure 1b). A key property of the AtoP kriging estimator is its coherency or pycnophylactic property: the aggregation of the P_β point estimates within any given entity v_β returns the areal datum $z(v_\beta)$:

$$z(v_\beta) = \frac{1}{P_\beta} \sum_{s=1}^{P_\beta} z^*_{AtoP}(\mathbf{u}_s) \tag{3}$$

To satisfy this constraint, the same K areal data must be used for prediction at each of the P_β discretizing point $\mathbf{u}_s$. Uncertainty about the areal data can be incorporated into the kriging system by adding a noise variance term to the diagonal elements of the kriging matrix (i.e. block-to-block covariances), leading to the filtering of that areal noise during the disaggregation procedure. This is similar to the Poisson or binomial AtoP kriging

approaches introduced by Goovaerts (2010) where the areal data are mortality rates that might be unstable and for which the pycnophylactic property is not desirable.

By analogy with ordinary kriging, the AtoP kriging variance associated with estimate (1) is computed as:

$$\sigma^2_{AtoP}(\mathbf{u}_s) = C(0) - \sum_{k=1}^{K} \lambda_k(\mathbf{u}_s)\overline{C}(v_k, \mathbf{u}_s) - \mu(\mathbf{u}_s). \quad (4)$$

where C(0) is the variance of the point process or sill of the point support variogram model (see Section 2.2).

2.2. Variogram deconvolution

The estimation of the block-to-block and block-to-point covariance terms in the kriging system (Equation 2) requires knowledge of the point support covariance $C(\mathbf{h})$ or point support variogram model $\gamma(\mathbf{h})$. Since only areal or polygon data are available, one must proceed in two steps: 1) compute and model the variogram of the areal data, and 2) deconvolute the block-support model to derive the point support variogram. In this paper, the point-support variogram model was inferred using the iterative deconvolution procedure of Goovaerts (2008) that seeks the point-support model that, once regularized, is the closest to the model fitted to the areal data.

This procedure is illustrated graphically in Figure 2a. The experimental variogram of the legacy soil polygon map (areal) data is computed (black dashed line) and then modelled by a weighted least squares fitting procedure (black solid line). A candidate point support or deconvoluted model is then chosen (solid grey line) and this is regularized using Equation (21) in Goovaerts (2008). The regularized model (black dotted line) is then compared to the model fitted to the experimental variogram of the areal or polygon data. Based on the differences between the regularized model and the areal model, the optimal point support model is rescaled and provides a new candidate model for the next iteration. The deconvoluted model which when regularized is closest to the model for the areal data is used for AtoP kriging. This deconvolution procedure is unlike conventional deconvolution methods – developed for regular mining blocks – because it takes into account the irregular shape and size of areal units (Kerry et al. 2010).

2.3. Area to Point Regression Kriging

The disaggregation of legacy data using the AtoP estimator (Equation 1) accounts only for the geometric properties of the different blocks v_β. Mapping the variability within each block v_β is likely to improve if ancillary data correlated with the soil property z (e.g. elevation or remotely sensed data) are available at a finer scale. We consider here the situation where these ancillary data are known at all N nodes of the interpolation grid and to simplify equations we present the case of a single secondary variable y. Let $\{y(\mathbf{u}_s); s=1,\ldots,N\}$ denote the grid of ancillary data and $\{y(v_\beta); \beta=1,\ldots,B\}$ be their average value within the B blocks v_β. A straightforward way to incorporate these ancillary data is to use them to derive the local mean m of the soil property z using a regression model: $m^*(v_\beta) = f[y(v_\beta)]$, then conduct AtoP kriging on the residuals. The RK estimate is written as follows:

$$z^*_{RK}(\mathbf{u}_s) = m^*(\mathbf{u}_s) + \sum_{k=1}^{K} \lambda_k(\mathbf{u}_s)\left[z(v_k) - m^*(v_k)\right], \quad (5)$$

Where the local mean $m^*(\mathbf{u}_s)$ is computed as a function of the ancillary data at that location: $m^*(\mathbf{u}_s) = f[y(\mathbf{u}_s)]$. In this paper we used the same regression model f(.) for both areal and point data, under the implicit assumption that the model is linear. The weights λ_k assigned to the K neighbouring areal data are the solution of the following simple kriging system:

$$\sum_{k'=1}^{K} \lambda_{k'}(\mathbf{u}_s)\overline{C}_R(v_k, v_{k'}) = \overline{C}_R(v_k, \mathbf{u}_s), \qquad \alpha = 1,\ldots,n(\mathbf{u}) \quad (6)$$

where the block-to-block covariances $\overline{C}_R(v_k, v_{k'}) = \mathrm{Cov}\{R(v_k), R(v_{k'})\}$ are derived from the point-support covariance of the residual random function $R(\mathrm{u})=Z(\mathrm{u})-m^*(\mathrm{u})$ using the discretization procedure described in Figure 2a.

2.4 Incorporation of Soil Map Polygon Data into Regression Kriging

Consider the situation where legacy data $\{z(v_\beta); \beta=1,\ldots,B\}$ are supplemented by a set of field measurements of the soil property z of interest $\{z(\mathbf{u}_\alpha); \alpha=1,\ldots,n\}$. In this paper, the two sets of data were combined using the following regression kriging estimate:

$$z^*_{RK}(\mathbf{u}) = m^*(\mathbf{u}) + \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha \left[z(\mathbf{u}_\alpha) - m^*(\mathbf{u}_\alpha)\right] = m^*(\mathbf{u}) + \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha r(\mathbf{u}_\alpha). \quad (7)$$

The local means $m^*(\mathbf{u})$ and $m^*(\mathbf{u}_\alpha)$ are identified either with the AtoP kriging estimated (AtoP SOC), legacy data disaggregated by AtoP kriging (Equation 1) at these locations or with the value of the legacy data at these locations disaggregated by simple rasterization (Polygon SOC, i.e. assuming that the local mean is constant within the mapping units). In other words, soil map legacy data (either AtoP SOC or Polygon SOC) are used to derive the spatial distribution of the local means of field data, and the variation in field data that is not explained by soil map legacy data (i.e. residuals) is then interpolated using kriging. The weights λ_α assigned to the $n(\mathbf{u})$ neighbouring residuals are the solution of the following simple kriging system:

$$\sum_{\beta=1}^{n(\mathbf{u})} \lambda_\beta C_R(\mathbf{u}_\alpha - \mathbf{u}_\beta) = C_R(\mathbf{u}_\alpha - \mathbf{u}) \qquad \alpha = 1,\ldots,n(\mathbf{u}), \quad (8)$$

where $C_R(\mathrm{h})$ is the covariance of the residual random function $R(\mathrm{u})=Z(\mathrm{u})-m^*(\mathrm{u})$. Assuming the independence between the local mean and the residual random function, the prediction variance for estimate (7) can be computed as the sum of the estimation variance for the local mean and the residual kriging variance:

$$\sigma^2_{RK}(\mathbf{u}) = \sigma^2_m(\mathbf{u}) + C_R(0) - \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha C_R(\mathbf{u}_\alpha - \mathbf{u}). \quad (9)$$

The variance $\sigma^2_m(\mathbf{u})$ is the AtoP kriging variance (Equation 4) or the prediction variance for that unit's local mean, depending on whether the local means are identified by

disaggregation with the AtoP kriging (AtoP SOC) (Equation 1) or simple rasterization of the areal legacy data (Polygon SOC).

As mentioned in the introduction, the regression kriging estimate is not optimal in that it is computed in two steps: 1) estimation of the local mean, and 2) kriging of residuals. This allowed direct comparison with the work of Rawlins et al. (2009), but unlike the one-step approach introduced in Goovaerts (2011), there is no guarantee than the regression kriging estimates still fulfill the coherency or pycnophylactic property of the original AtoP estimates. In addition, the regression kriging variance tends to exceed the kriging variance of Goovaert's (2011) AAP kriging estimator.

3. Materials and Methods

3.1. Soil Sampling

The soil sampling was undertaken between July 2004 and March 2006 comprising the collection of a sample of topsoil 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. This was part of the Tellus survey of Northern Ireland http://www.bgs.ac.uk/gsni/tellus/.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 soil samples were air-dried in a dedicated temperature controlled oven at 30º C for between 2 and 3 days, disaggregated and sieved to less than 2 mm. From each a 50-g sub-sample was ground in an agate planetary ball mill. Loss on ignition (450ºC) was determined for the air-dried disaggregated fraction. These values were multiplied by 0.58 to give percent SOC equivalents. Replicate analyses were done to ensure accuracy and repeatability of results. The differences between replicate analyses were small. The SOC values from these 6862 soil sample data from the Tellus survey had a positively skewed distribution (Figure 2d) with a skew of 1.99 which was reduced to 0.94 upon log transformation. When these Tellus soil sample data were used in regression kriging logSOC was used. The SOC values from the Tellus survey points are shown in log form in Figure 3a to enable visual distinction given the large range of SOC values in Northern Ireland. A prediction set of 3000 samples was selected randomly and the other samples (n=3862) were used for validation.

3.2. Legacy Soil Map

A 1:250,000 digital map (courtesy of the Agricultural, Food and Biosciences Institute of Northern Ireland; AFBINI) of soil associations (polygons) and associated typical percent SOC concentrations was used here as legacy soil data. This soil map and associated memoir (Cruikshank, 1997) resulted from the first full survey of the soils of Northern Ireland completed by the Department of Agriculture and Rural Development (DARD) between 1988 and 1997. In this survey, the soils of Northern Ireland were systematically sampled on a regular 5 km grid. The soils were described, analysed and classified into soil series. Soil profiles were characterized from each major soil series in the agriculturally-important areas (areas below 200 m) and the physical and chemical properties of these profiles were determined. The survey identified 308 distinct soil

series, the locations of which were published as a 1:50,000 scale soil series maps. The 1:250,000 polygon map is a generalization for the 1:50,000 maps. The soil carbon values provided for the soil associations in the 1:250,000 map are from 0-30 cm depth and are typical values for the soil association in that polygon rather than actual values recorded within that polygon. The log transformed % SOC concentrations associated with this map are shown in Figure 3b. The 1:250,000 scale map was used in this analysis over the 1:50,000 scale maps as it was thought a more appropriate scale when considering the whole of northern Ireland, and SOC percentages for peat areas were not included for the 1:50,000 maps.

According to Cruickshank (1997) the proportions of the dominant soil types in Northern Ireland comprise 54% gleys, 24% rankers and peats and 16% freely drained soils, with minor soil types accounting for the other 6%. The dominance of gleyed soils and peats reflect the wet climate of Northern Ireland (mean annual precipitation of around 1 m) which in combination with a varied topography results in a wide range of SOC contents across the study area. The SOC concentrations associated with the 1:250,000 polygon map (Figure 2b and Figure 3b) were disaggregated first by simple rasterizing (500 m pixels) to give what we subsequently refer to as Polygon SOC, where all pixels within a polygon received the same SOC concentration. They were also disaggregated using AtoP kriging and AtoP RK, the theory of which is described above and procedures for calculating which are described below.

3.3. Ancillary data

Two ancillary datasets were available for use as covariates (Rawlins et al., 2009). The first was airborne radiometric K (%) for the whole of Northern Ireland collected as part of the Tellus airborne geophysical survey in 2005 and 2006 (Figure 3f). The flight lines were spaced 200 m apart and the spacing between measurements along the flight line was between 60 and 70 m yielding around 1.2 million values; the detector was an Exploranium GR820 256 channel gamma spectrometer. Data were processed to correct for aircraft and cosmic background radiation, aircraft altitude and spectral interactions. The second covariate was altitude from a Digital Elevation Model (Figure 3e) which was available at 50 m resolution based on airborne, photogrammetric acquisition (Ordnance Survey of Northern Ireland's ® data). We used a GIS spatial join procedure to associate each soil sampling observation with the values of its nearest airborne radiometric K (%) survey observation and its altitude in metres.

In data-rich scenarios such as Northern Ireland, ancillary data, recent soil measurements and disaggregated legacy soil data can be used together to predict soil organic carbon using RK. Only those ancillary data with a strong spatial correlation with SOC values, and for which there were sound theoretical reasons for this correlation, were considered as suitable covariates. Prior to RK, the correlations between SOC and the various ancillary data were calculated and the theoretical basis for them identified.

Gamma-ray attenuation and soil moisture typically show strong spatial correlation and this can be extended to SOC because it accumulates in wet or waterlogged soil. Figure 3f shows the spatial variation in radiometric K whilst the distribution of log SOC from the

Tellus survey is shown in Figure 3a; their Pearson correlation coefficient (*r*) is -0.67. There is also typically a spatial correlation between SOC and altitude; precipitation tends to increase and temperature decrease with increasing altitude. Both of these factors aid the accumulation of SOC because mineralization decreases with temperature and soils which are waterlogged for more of the year have slower rates of soil carbon decomposition. Figure 3e shows the variation in altitude compared to log SOC from the Tellus survey (Figure 3a); the Pearson correlation coefficient (*r*) between altitude and logSOC is 0.60. The Pearson correlation (*r*) between the Polygon SOC values (rasterized) and the Tellus SOC concentrations was 0.61. The polygon map SOC values were also disaggregated by AtoP kriging as described in Section 2.1 and 3.4. Figure 3c shows the log of the AtoP SOC values to allow comparison to log SOC from the Tellus survey (Figure 3a) and the Pearson correlation (*r*) between these values was 0.68 showing that there is some correspondence between the legacy soil map SOC and recent soil survey data and that relationship is strengthened when legacy data are disaggregated using AtoP kriging (AtoP SOC) instead of being assumed constant within each mapping unit (Polygon SOC).

3.4. Geostatistical analysis

3.4.1. Disaggregation of soil polygon data using AtoP Kriging and AtoP Regression Kriging

Geostatistical disaggregation of the soil Polygon SOC data was done using two methods available in the software SpaceStat (BioMedware Inc, 2011): AtoP Kriging (hereafter called AtoP) and AtoP RK. Both methods use only the soil polygon data and do not require any new soil sample data, however, the latter method uses appropriate ancillary data to inform on the within polygon variation in SOC.

The frequency distribution of untransformed SOC concentrations from the legacy soil polygon map (Figure 2b) is somewhat different to that of the Tellus soil sample data (Figure 2d) and did not particularly benefit from a log transform as did the Tellus soil sample data. The legacy soil polygon map SOC values relate to soil types; mineral soils with low SOC (mean SOC, 4.8 %), organo-mineral soils (mean SOC, 14.3 %) and organic soils (mean SOC, 44.8 %). Therefore, we computed the variogram for the legacy soil map areal or polygon data from class residuals of these three broad groups (Figure 2c). The large outliers in the distribution of class residuals (Figure 2c) suggests that some soils in the legacy map have been mis-classified. A model was fitted to this variogram of the class residuals and deconvoluted as described above. We then applied AtoP kriging to the polygon class residuals to estimate their values at the Tellus soil sample locations. The variogram deconvolution and AtoP kriging of the class residuals were undertaken using the SpaceStat ^{TM} software (BioMedware, Michigan). We then added the class mean from the soil group to these residuals to give Area to Point kriged SOC which we subsequently refer to as AtoP SOC. The mean absolute errors (MAEs) between AtoP SOC and the measured SOC values at the Tellus soil sample locations were calculated as were the mean errors (MEs), mean squared deviation ratios (MSDRs) and median squared deviation ratios (MeSDRs). The main focus in each corss-validation study mentioned below is on the MAEs, but the latter cross-validation statistics were computed

to check that the data were unbiased (ME) and that an appropriate model had been used (i.e.MSDR close to 1 or MeSDR close to 0.455)..

A similar procedure was followed for AtoP RK. First, linear regression was performed in which the Polygon SOC values were the dependent variable whilst the independent variables were altitude, radiometric K and squared radiometric K that were averaged within each polygon. This combination of independent variables was generally optimum from the analyses published by Rawlins et al. (2009) and these variables are related to the scorpan factors relief and parent material (see explanations of correlations above) making them particularly appropriate theoretically for disaggregating soil polygon information. The residuals from regression were then used to compute the areal support variogram which was deconvoluted as above. Area to Point kriging was then applied to the regression residuals to estimate their values at the Tellus soil sampling locations using the deconvoluted variogram. These residuals were then added to the SOC estimates obtained by applying the areal regression model to the same independent variables available at the Tellus soil sampling locations. The MAEs between the AtoP RK SOC estimates and those measured at the Tellus soil sample points were calculated.

3.4.2. Incorporating Disaggregated Soil Polygon Map data into Regression Kriging

Once soil polygon data were disaggregated using either simple rasterization (Polygon SOC) or AtoP kriging (AtoP SOC), they could be used as a fixed effect in estimating SOC concentrations by RK of new soil sample data from the Tellus survey. Rawlins et al. (2009) used RK to incorporate altitude and radiometric K in various combinations into the estimation of SOC. The six models (referred to subsequently as Models 1-6) used by Rawlins et al. (2009) are summarized in Table 1. Here we added the Polygon SOC values and the AtoP SOC values as extra fixed effects in each of the six models used by Rawlins et al. (2009). In each case, the logSOC values for the prediction set of 3000 of Tellus soil survey data were used as the dependent variable to compute the regression models. The residuals from these models were kriged to the 3862 validation sites and were added to regressed values at each location then backtransformed. The MAEs associated with each model based on these separate prediction and validation sets were then calculated.

3.4.3. Jack-knife procedure

A jack-knife procedure was undertaken to investigate the impact of the interpolation algorithm and sample size on the prediction errors. One hundred repeated random selections of sample subsets of size 100, 200, 300 … 2000 were created from the original 6862 Tellus soil data. The 100 random subsets were used for prediction to the remaining Tellus soil survey locations using RK with Models 1-6 (see Table 1), and Models 1-6 plus Polygon SOC or AtoP SOC. As above, MAEs were calculated to determine the relative magnitudes of estimation error.

4. Results and Discussion

4.1. Disaggregation of Legacy Soil Polygon Map Data

Table 2 shows the MAEs based on disaggregation of the SOC data from the legacy soil polygon map. When all soil types are considered, simple disaggregation (Polygon SOC) has slightly smaller MAEs than AtoP SOC. However, while the MAEs are similar for

mineral and peat soils for AtoP SOC and Polygon SOC (around 4 and 18 %, respectively), the AtoP disaggregation method has substantially smaller MAEs (13.0 %) for organo-mineral soils than the Polygon SOC approach (15.2%). The lowest MAEs in each soil class were those for AtoP RK where ancillary data (altitude and radiometric K) account for within polygon variation of soil SOC suggesting that such data related to scorpan factors add value to the disaggregation approach.

Simple rasterization of the polygon map produces sharper boundaries in SOC content between soil types and assumes there is no variation in SOC concentration within a given soil class. Although sharp boundaries occur between some soil types, there is always some variation within soil types. Also, it is more common for changes between soil types to be gradual rather than sharp. This is reflected in the AtoP approach which accounts for the spatial configuration of the soil polygons and the underlying trends in the spatial distribution of SOC within and between soil classes. Differences in the MAEs suggest that there are regions, particularly for organo-mineral soils where the AtoP kriging may be more appropriate than simple rasterization as a disaggregation procedure.

Table 2 shows that AtoP RK produces the lowest MAEs of all disaggregation methods for all soils and all soil types. This shows that if data from a polygon map are to be disaggregated without use of any extra soil sample information, incorporating ancillary data that relate to soil-forming factors can bring significant benefits. Incorporating ancillary data that are related to soil-forming or scorpan factors (McBratney et al. 2003) into the disaggregation of soil legacy data is more theoretically sound than simple rasterization or AtoP kriging alone, however, when the simpler two-step AtoP RK procedure is used as here, the pycnophylactic property can be lost.

These results show that where current soil data are scarce and funding is not available for new soil survey, the disaggregation of data from legacy soil maps is a cost-effective alternative strategy for DSM. This disaggregation of legacy soil maps relies on the soil polygons with or without quantitative ancillary data which are relatively inexpensive to collect but provide some indication of within polygon variations of soil properties. No new soil measurements need to be collected for such disaggregation approaches.

4.2. Incorporation of Disaggregated Legacy Soil Polygon Map Data into Regression Kriging

When new soil data are available, it is expected that incorporation of disaggregated legacy data with ancillary data in RK of the new soil data will be more fruitful than merely using a polygon map and ancillary data to disaggregate the polygon data. Using 3000 Tellus soil sample data for prediction and 3862 for validation, Table 3 shows the MAEs for RK with and without AtoP SOC and Polygon SOC. Of the other cross-validation statistics (not shown), the MEs were close to zero showing no real bias in the data. Some MSDR values were close to one, however, given that some soils were mis-classified as peat and the MSDR is the ratio of the squared errors to the kriging variance, these large errors dominated the squared errors and produced some large MSDR values. In such cases a MeSDR close to 0.455 provides a better evaluation of whether an appropriate model has been used. The MeSDRs were close to 0.455.

Table 3 shows that when all soil types are considered, incorporating Polygon SOC into each model reduces the MAEs, and inclusion of AtoP SOC reduces the MAEs further. This is also the case for peat soils. The patterns are less consistent for mineral and organo-mineral soils, the reduction in MAE is smaller when Polygon SOC and AtoP SOC are included in RK for some models, but not for others. However, even when MAEs do not show a distinct benefit from incorporation of Polygon SOC and AtoP SOC in RK, the values are very similar to those when they are not incorporated. When certain combinations of altitude and radiometric K data are used, the advantage of including Polygon SOC and AtoP SOC values decreases. The best combinations of ancillary data (lowest MAEs) for each of the three soil types and all soils (Table 3) were those in Model 4 (constant plus K plus K^2) and Model 6 (constant plus altitude plus K plus K^2). It should be noted that in Table 3 the MAE for all soils for Model 6 is greater than each of the three individual soil types because for all soils it was assumed that the mean is constant whereas for the individual soils the mean was only constant within areas of the same broad soil class.

If the MAEs from cross-validation of RK (Table 3) which were computed using 3000 soil data are compared with those from mere disaggregation of legacy soil data from the polygon map by AtoP RK (Table 2; i.e no new soil sample data), the MAEs of the latter are similar to the former for all soil types (around 4 %), but are about double those of the former for individual soil types (Table 3).

4.3. Jack-knife analysis

The cross-validation results above were based on a single, random prediction subset of Tellus data from 3000 locations. The results demonstrated that it is worthwhile including disaggregated polygon data into RK of SOC. The benefits of a jack-knife procedure, which involves repeatedly selecting random subsets of various size are that it provides a better overall indication of the value of: 1) incorporating disaggregated legacy soil polygon map data into RK and, 2) which models produce more accurate estimates. Figure 4 shows the MAE jack-knife results for models 1-6 for all soil types plotted on the same scale. The MEs (not shown) were close to zero showing no bias except for very small sample sizes (<30) and MEs were also an order of magnitude lower for sample sizes ≥300. As for with the results documented in section 4.2, depending on the random nature of whether mis-classified samples were included in the jack-knife procedure for a given sample size or not, the MSDRs were sometimes but not always close to one. The MeSDRs were close to 0.455 and gave a better indication of whether a suitable model had been used for RK.

In Figure 4, the MAEs are somewhat smaller at all sample sizes when Polygon SOC is included, but there is a more marked reduction in MAEs when AtoP SOC is incorporated into RK. The MAEs are smallest, generally for Models 4 and 6 (around 4 %). Figure 5 also shows, for mineral soil types, that MAEs are smaller at all sample sizes when Polygon SOC is included and smaller still when AtoP SOC is included in RK. The only exception to this is for sample sizes greater than about 1000 for models 4 and 6, but MAEs are generally smallest for model 6 (around 1.6 %). For organo-mineral soils (Figure 6), the patterns are reversed for models 3 and 4 with MAEs being largest at all

sample sizes when AtoP SOC is included in RK and smallest when compared to the approach adopted by Rawlins et al.. However, the MAEs are smallest (around 2.8 %), especially for small sample sizes for Model 6. For peat soils (Figure 7) the advantage of incorporating Polygon SOC and AtoP SOC into RK is the most marked of any of the soil types; Model 4 has the lowest MAEs overall (around 3 %). Figure 7 also shows the largest differences between MAEs when comparing the different approaches (Rawlins et al., Polygon SOC and AtoP SOC) when compared with Figures 4-6. This, and the relatively small MAEs that were obtained for peat soils when AtoP SOC was incorporated into models 4 and 6, clearly indicate that in Northern Ireland where there is a large variation in SOC concentrations one of the major benefits of legacy data is that it identifies the distribution of peat. This is particularly the case when legacy data is disaggregated using AtoP kriging.

The MAE values for the best performing models (4 and 6) from the jack-knife procedure are similar to those reported from cross-validation in Table 3, but they show that they are more consistent in general than suggested by cross-validation alone, and that beyond a certain point collecting additional soil samples at other locations provides little benefit in terms of reducing MAEs.

Overall, model 6 had the lowest MAEs (Figures 4-7) so we chose to examine these results in greater detail (Figure 8). They show that incorporating Polygon SOC into RK leads to a small reduction in MAE, about 0.05 and 0.1 % reductions for all soil types and peat, respectively. However, for mineral and organo-mineral soils, incorporating Polygon SOC has little effect for sample sizes less than 1200. However, including AtoP SOC in RK reduced MAEs by 0.35 – 0.4% on average for all soils and peat and by about 0.03-0.05% for mineral and organo-mineral soils. Figure 8 and Figure 4 show that a suitable sample size for RK when incorporating AtoP SOC data is around 300 where the curve levels out and there is not a great advantage in terms of reducing MAE of collecting more samples. In many areas, airborne radiometric survey data are unavailable for DSM, whilst altitude data is usually available at some resolution, so we examined in greater detail the results for Model 2 in which only altitude is included as a covariate (Figure 9). The advantage of incorporating AtoP SOC data over Polygon SOC in RKs is clear and very marked for all soil types and sample sizes. For AtoP SOC, MAEs are reduced by about 1 % for all soil types whereas for the Polygon SOC, MAEs are only reduced by about 0.2 % compared with Rawlins et al's models. These reductions in MAEs are at least double the MAE reductions reported above for Model 6. Figure 9 also shows that if only altitude data were available as a covariate, around 600 measurements of SOC concentration would be suitable given where the curve levels out; 300 more than was the case for when radiometric K data were included. However, without radiometric K data (Model 2), the MAEs are consistently larger compared to when it is included as a covariate (Model 6).

The results of the jack-knife procedure show the consistent improvement of MAEs when Polygon SOC and AtoP SOC are incorporated into regression kriging and they show that it is better to incorporate legacy soil data as AtoP SOC than Polygon SOC. Using more ancillary data can reduce the number of soil samples that need to be collected for RK. Practitioners may need to evaluate the relative costs and benefits of collecting extra soil

samples versus including extra ancillary data. Certainly, if the legacy data being used are old or lacking specific information on methodology, the benefits of collecting extra soil samples as opposed to ancillary data are clear.

5. Conclusions

Our analysis shows that for disaggregating legacy SOC data from a polygon map, an AtoP RK approach is more effective than simple rasterization. An AtoP RK approach is theoretically sound because it allows for within class variability, spatial autocorrelation and scorpan factors as represented by ancillary data. In the case of estimating SOC concentrations across Northern Ireland, the AtoP RK approach does not require the collection and use of new soil measurements, but could produce overall MAEs that are similar to those in which 3000 new soil SOC measurements which were distributed throughout the study area are included. The cross-validation and jack-knife results show that incorporating AtoP SOC into RK where some new soil samples are required in addition to legacy data is desirable to reduce errors further, especially for mineral and organo-mineral soil types. Nevertheless, the jack-knife results suggest that between 300 and 600 new soil measurements would be optimal depending on the availability of ancillary data for the study area. Incorporating Polygon SOC into RK with various numbers of spatially distributed soil measurements reduces errors at all sample sizes and for most models. Using AtoP SOC reduces MAEs further for Northern Ireland.

These results clearly show that legacy data is useful and should not be disregarded in DSM approaches, and that there are differences in the performance of methods for disaggregating legacy data in DSM. Nevertheless, some important cautionary notes should be made and some additional research conducted before these methods are widely employed with a wide range of soil properties and in vastly different pedological settings.

The legacy soil data used here were collected between 1988 and 1997, and at similar soil depths to new soil data, and are thus likely to be more useful than older legacy data or data that have been collected at markedly different depths. This is especially the case for properties like SOC which can change over time and markedly with depth. The temporal aspect is likely to be less of an issue with more permanent soil properties like soil texture, but differences in the methods of determining soil texture between legacy data and new sample data may also reduce the usefulness of legacy data. The current research has shown the usefulness of the AtoP approach for disaggregating legacy data and then incorporating it into RK, but many locations do not have such a wide range of SOC values within small areas and this approach needs to be tested in areas with far less variability in SOC and in areas with older legacy data. Also the two-step approach used in this study does not ensure that the pycnophylactic property of AtoP kriging is preserved. We did this to allow direct comparison with previous analysis of the Tellus data, but recommended that practitioners use the one-step approach outlined by Goovaerts (2011) when appropriate software becomes available if the pycnophylactic property is very important to their study. The AtoP approach is clearly most useful for DSM in a data-rich scenario where current soil survey data, ancillary data and legacy data are available, but this research highlights the relative magnitude of the errors associated with scenarios when less data is available. Therefore these approaches should be

adaptable to situations where less data is available to give a first rough estimate of how SOC varies in the area and perhaps inform future sampling protocols.

Acknowledgements
We thank Michael Young of the Geological Survey of Northern Ireland for arranging access to the Tellus data. The Tellus project was funded by the Department of Enterprise, Trade and Investment and by the Building Sustainable Prosperity Scheme of the Rural Development Programme (Department of Agriculture and Rural Development of Northern Ireland). Topographic data are based upon Ordnance Survey of Northern Ireland's data with the permission of the Controller of Her Majesty's Stationery Office, © Crown copyright and database rights Licence Number DMOU205. This paper is published with the permission of the Executive Director of the British Geological Survey (NERC). Dr. Goovaerts' work was funded by grant R44-CA132347-02 from the National Cancer Institute. Dr Marchant's contribution was funded by BBSRC through its core grant to Rothamsted Research.

References
Bui, E.N, Moran, C.J., 2001. Disaggregation of polygons of surficial geology and soil maps using spatial modelling and legacy data. Geoderma 103, 79-94.

Carré, F., McBratney, A.B., Minasny, B., 2007a. Estimation and potential improvement of the quality of legacy soil samples for digital soil mapping. Geoderma 141, 1-14.

Carré, F., McBratney, A.B., Mayr, T., Montanarella, L., 2007b. Digital soil assessments: Beyond DSM. Geoderma 142, 69-79.

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

De Bruin, S., Wielemaker, W. G., Molenaar, M., 1999. Formalisation of soil-landscape knowledge through interactive hierarchical disaggregation. Geoderma 91, 151–172.

Eagleson, S., Escobar, F., Williamson, I.P., 1999. Spatial hierarchical reasoning applied to administrative boundary design using GIS. 6th Southeast Asian Surveyors Congress, 1–6 November 1999, Fremantle, WA. (http://www.geom.unimelb.edu.aurresearchrSDI–research).

Goovaerts, P., 2008. Kriging and semivariogram deconvolution in the presence of irregular geographical units. Mathematical Geosciences 40, 101-128.

Goovaerts, P. 2010. Combining areal and point data in geostatistical interpolation: Applications to soil science and medical geography. Mathematical Geosciences 42(5), 535-554.

Goovaerts, P., 2011. A coherent geostatistical approach for combining choropleth map and field data in the spatial interpolation of soil properties. European Journal of Soil Science 62(3), 371-380.

Gotway, C.A., Young, L.J., 2002. Combining incompatible spatial data. Journal of the American Statistical Association 97, 632-648.

Kerry, R., Goovaerts, P., Haining, R.P. and Ceccato, V., 2010. Applying geostatistical analysis to crime data: car-related thefts in the Baltic States. Geographical Analysis 42, 53-77.

Kyriakidis, P., 2004. A geostatistical framework for area-to-point spatial interpolation. Geographical Analysis 36, 259-289.

Lagacherie, P., McBratney, A.B., 2007. Chapter 1. Spatial soil information systems and spatial soil inference systems: perspectives for digital soil mapping. In: Lagacherie, P., McBratney, A.B., Voltz, M. (Eds.), Digital Soil Mapping: An Initial Perspective. Developments in Soil Science 31. Elsevier, Amsterdam, p. 250

Liu, X.H., Kyriakidis , P.C., Goodchild, M.F. 2008. Population density estimation using regression and area-to-point residual kriging. International Journal of Geographical Information Science 22(4), 431–447.

McBratney, A.B., 1998. Some considerations on methods for spatially aggregating and disaggregating soil information. Nutrient Cycling in Agroecosystems 50, 51–62.

McBratney, A., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3–52.

Rawlins, B. G., Marchant, B. P., Smyth, D., Scheib, C., Lark, R. M., Jordan, C., 2009. Airbourne radiometrics survey data and a DTM as covariates for regional scale mapping of soil organic carbon across Northern Ireland. European Journal of Soil Science 60, 44-54.

Tobler, W.R., 1979. Smooth pycnophylactic interpolations for geographical regions. Journal of the American Statistical Association 74, 519–536.

Yoo, E.H., Kyriakidis, P., 2009. Area-to-point Kriging in spatial hedonic pricing models. Journal of Geographical Systems 11, 381-406.

Table 1. Fixed effects for the six models applied by Rawlins et al. (2009)

Model	Fixed effects
1	Constant (i.e. the mean)
2	Constant plus altitude
3	Constant plus K
4	Constant plus K plus K^2
5	Constant plus altitude plus K
6	Constant plus altitude plus K plus K^2

Table 2. Mean absolute errors (MAEs) for different methods of disaggregating SOC polygon data (soil polygon data used for prediction and Tellus validation data n=3862 used for validation).

Disaggregation method	MAE			
	All soils	Mineral soils	Organo-mineral soils	Peat soils
Polygon SOC	6.81	4.16	15.20	17.16
AtoP SOC	7.44	4.99	13.03	17.92
AtoP RK SOC	4.72	2.82	7.99	13.31
AtoP KED SOC	7.24	4.32	15.83	22.37

Table 3. Mean absolute errors (MAEs) from cross-validation of regression kriging for a) Rawlins et al. (2009) models 1-6, b) Rawlins et al. models 1-6 plus polygon SOC, and c) Rawlins et al. models 1-6 plus AtoP SOC (Tellus data n=3000 used for prediction and Tellus data n=3862 used for validation) .

		Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
All soils	(a) Rawlins et al.	6.46	6.08	4.43	3.86	4.53	4.02
	(b) + Polygon SOC	6.13	5.94	4.32	3.82	4.43	3.96
	(c) + AtoP SOC	5.60	5.37	4.11	3.66	4.18	3.77
Mineral soils	(a) Rawlins et al.	1.93	1.85	1.62	1.57	1.61	1.57
	(b) + Polygon SOC	1.88	1.82	1.62	1.56	1.60	1.56
	(c) + AtoP SOC	1.81	1.75	1.62	1.58	1.59	1.57
Organo-mineral soils	(a) Rawlins et al.	3.32	3.40	2.66	2.64	2.86	2.69
	(b) + Polygon SOC	3.35	3.35	2.72	2.65	2.82	2.68
	(c) + AtoP SOC	3.43	3.27	2.86	2.81	2.78	2.70
Peat soils	(a) Rawlins et al.	4.35	4.20	3.58	3.15	3.62	3.30
	(b) + Polygon SOC	4.21	4.17	3.46	3.11	3.56	3.25
	(c) + AtoP SOC	3.89	3.90	3.19	2.82	3.36	3.04

Figure 1. (a) Representation of the discretization procedure used to estimate: (a) Block-to-point and (b) Block-to-Block covariances as the average of point-to-point covariances.

Figure 2. (a) Variogram deconvolution procedure shown with experimental variogram (dashed black line) and model for soil polygon map data (solid black line), deconvoluted variogram model (point support, solid grey line) and theoretically regularized variogram model (dashed grey line). Histograms of (b) organic carbon for legacy soil polygon map data, (c) organic carbon soil class residuals for legacy soil polygon map data and (d) organic carbon from samples (n=6862) in the Tellus survey.

Figure 3. The distribution of measured and estimated SOC data and covariates across Northern Ireland: (a) Tellus survey measured log SOC (n=6862), (b) Polygon map log SOC, (c) Area to Point (AtoP) kriged logSOC, (d) the key for Log SOC in a-c, (e) altitude (m) and (f) airborne radiometric K (%).

Figure 4. Jack-knife results for all soil types for models 1-6 (M1-M6)

Figure 5. Jack-knife results for mineral soil types for models 1-6 (M1-M6).

Figure 6. Jack-knife results for organo-mineral soil types for models 1-6 (M1-M6)

Figure 7. Jack-knife results for peat soils for models 1-6 (M1-M6)

Figure 8. Jack-knife results for model 6 for all soil types

Figure 9. Jack-knife results for model 2 for all soil types

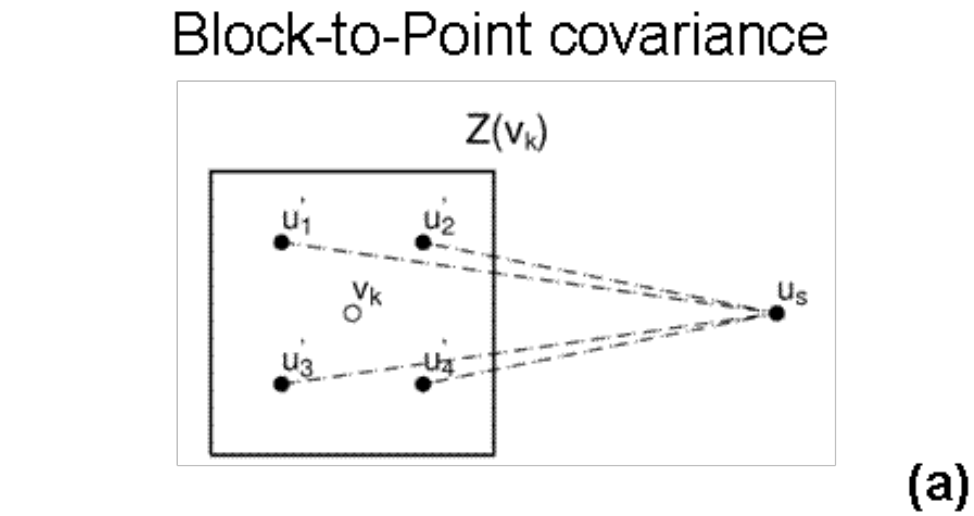

Figure 1. (a) Representation of the discretization procedure used to estimate: (a) Block-to-point and (b) Block-to-Block covariances as the average of point-to-point covariances.

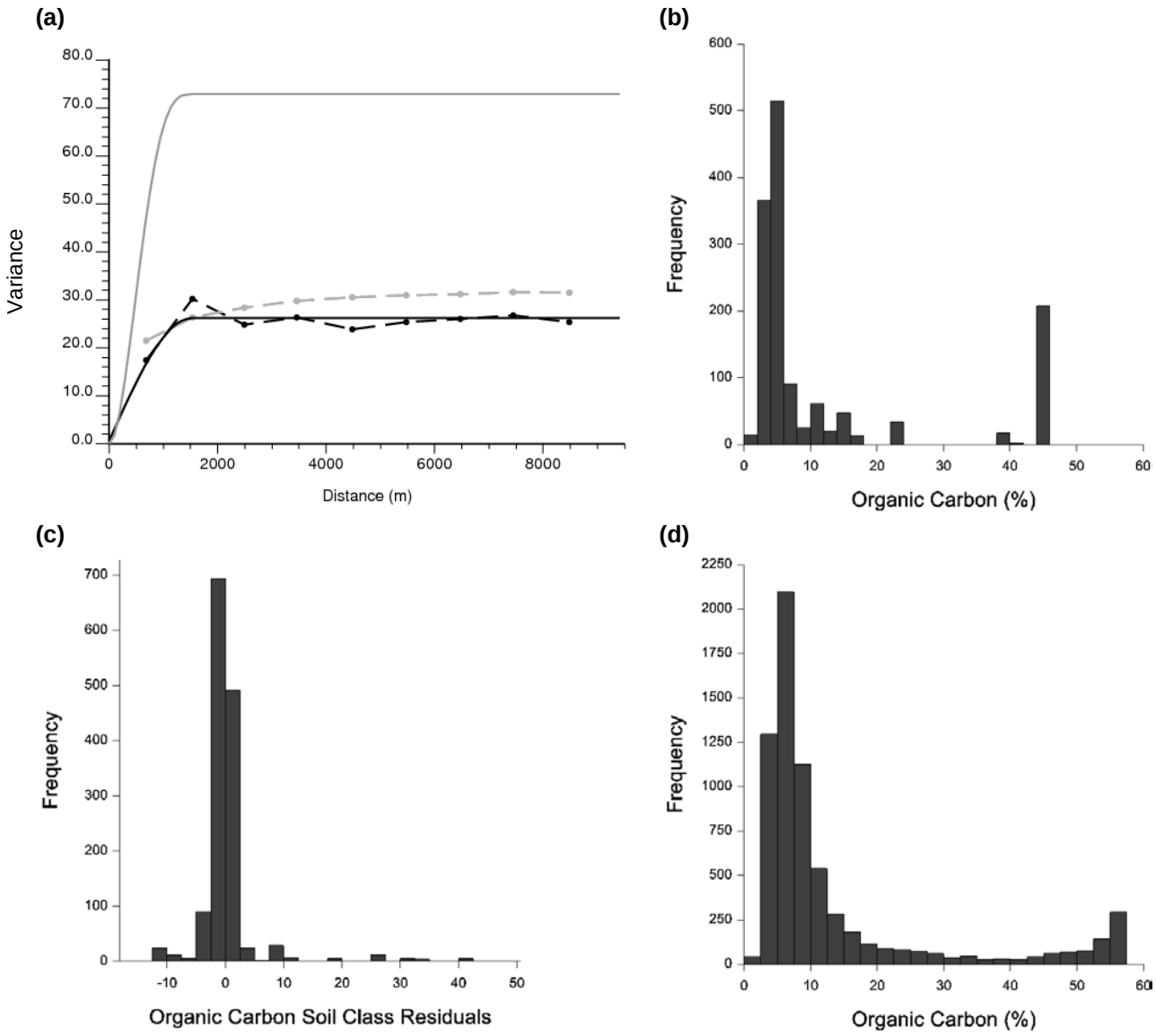

Figure 2. (a) Variogram deconvolution procedure shown with experimental variogram (dashed black line) and model for soil polygon map data (solid black line), deconvoluted variogram model (point support, solid grey line) and theoretically regularized variogram model (dashed grey line). Histograms of (b) organic carbon for legacy soil polygon map data, (c) organic carbon soil class residuals for legacy soil polygon map data and (d) organic carbon from samples (n=6862) in the Tellus survey.

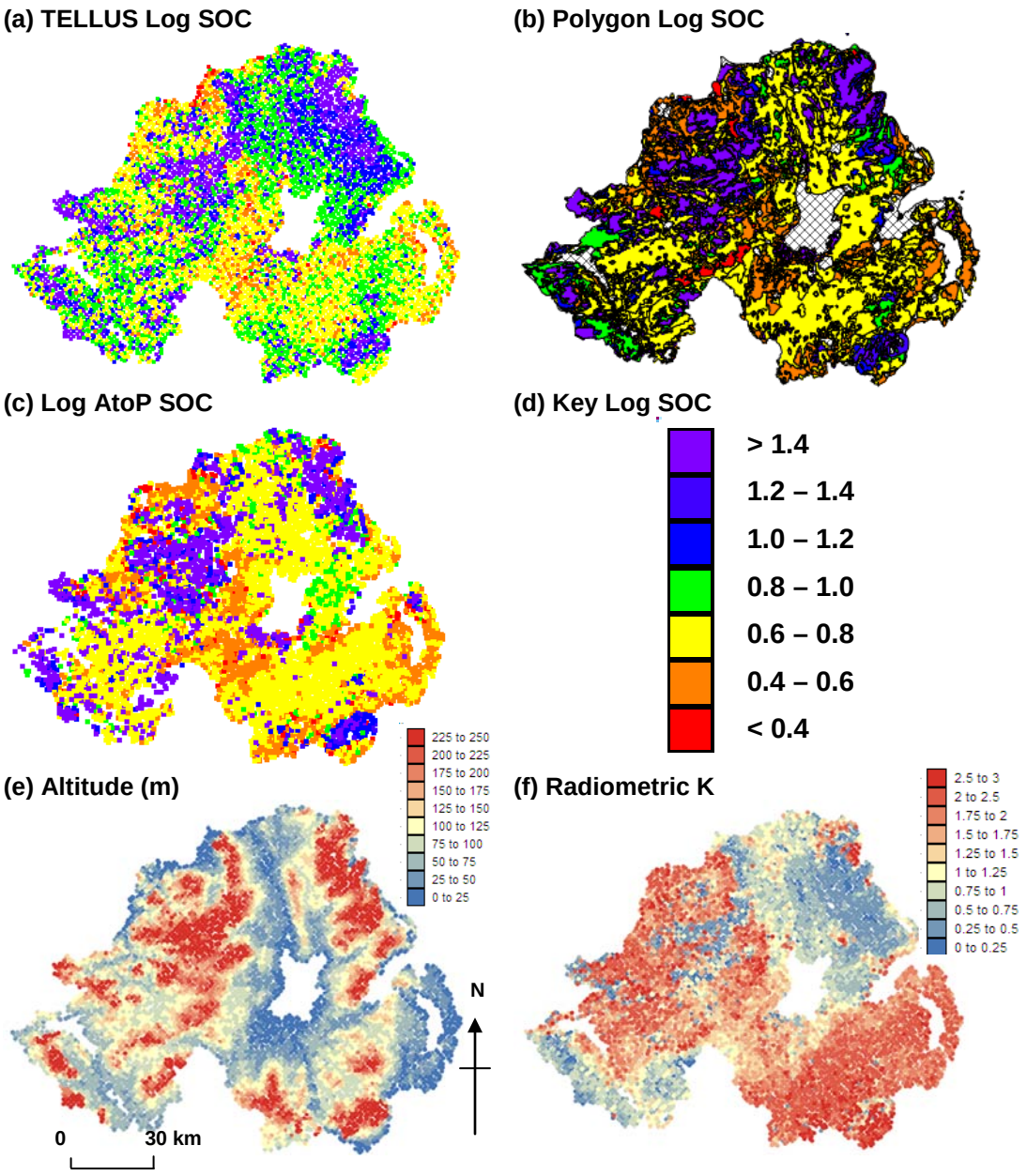

Figure 3. The distribution of measured and estimated SOC data and covariates across Northern Ireland: (a) Tellus survey measured log SOC (n=6862), (b) Polygon map log SOC, (c) Area to Point (AtoP) kriged logSOC, (d) the key for Log SOC in a-c, (e) altitude (m) and (f) airborne radiometric K (%).

A colour Figure3 should appear online only, please use black and white version below for printed journal

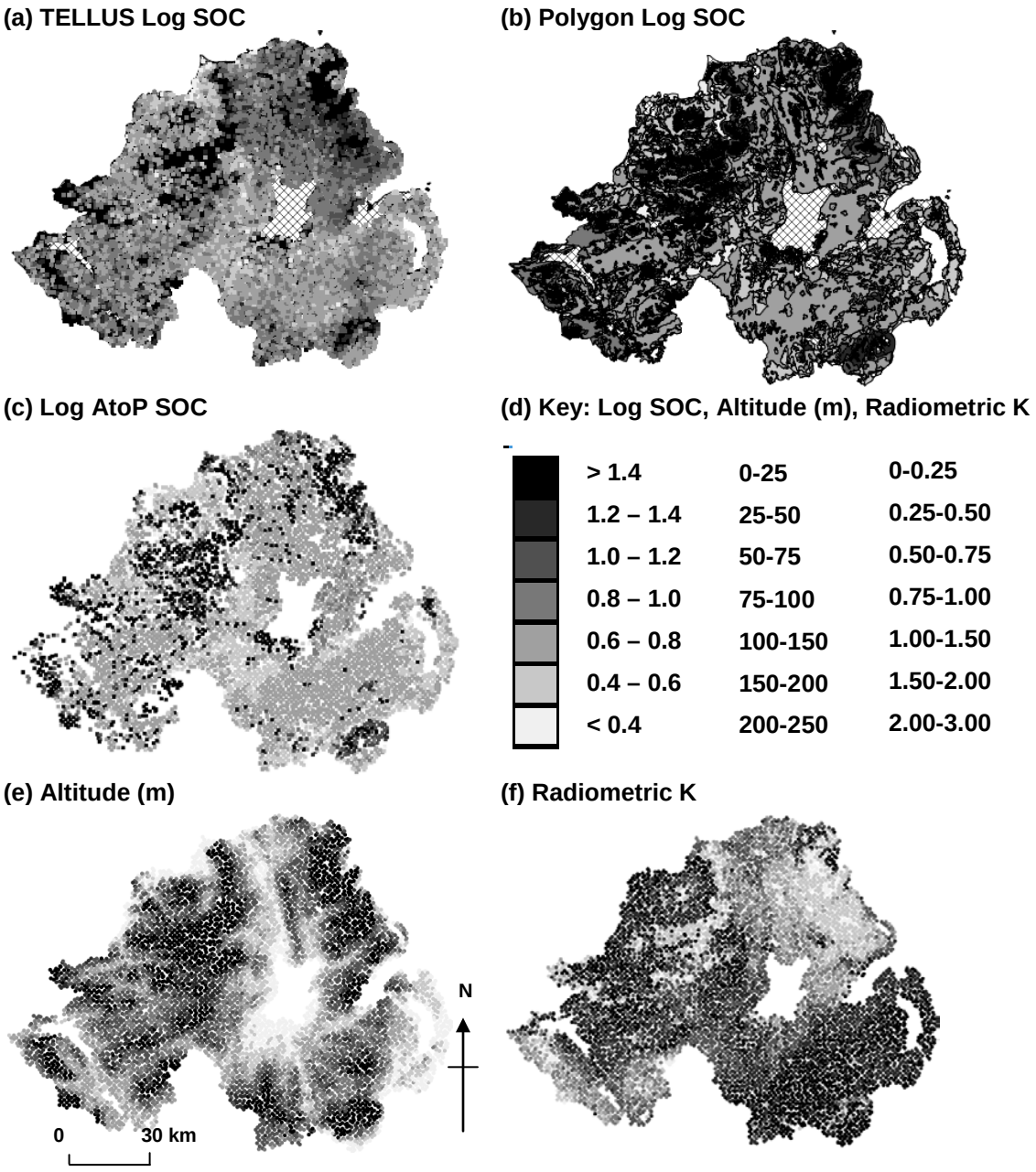

Figure 3. The distribution of measured and estimated SOC data and covariates across Northern Ireland: (a) Tellus survey measured log SOC (n=6862), (b) Polygon map log SOC, (c) Area to Point (AtoP) kriged logSOC, (d) the key for Log SOC in a-c, altitude and radiometric K, (e) altitude (m) and (f) airborne radiometric K (%).

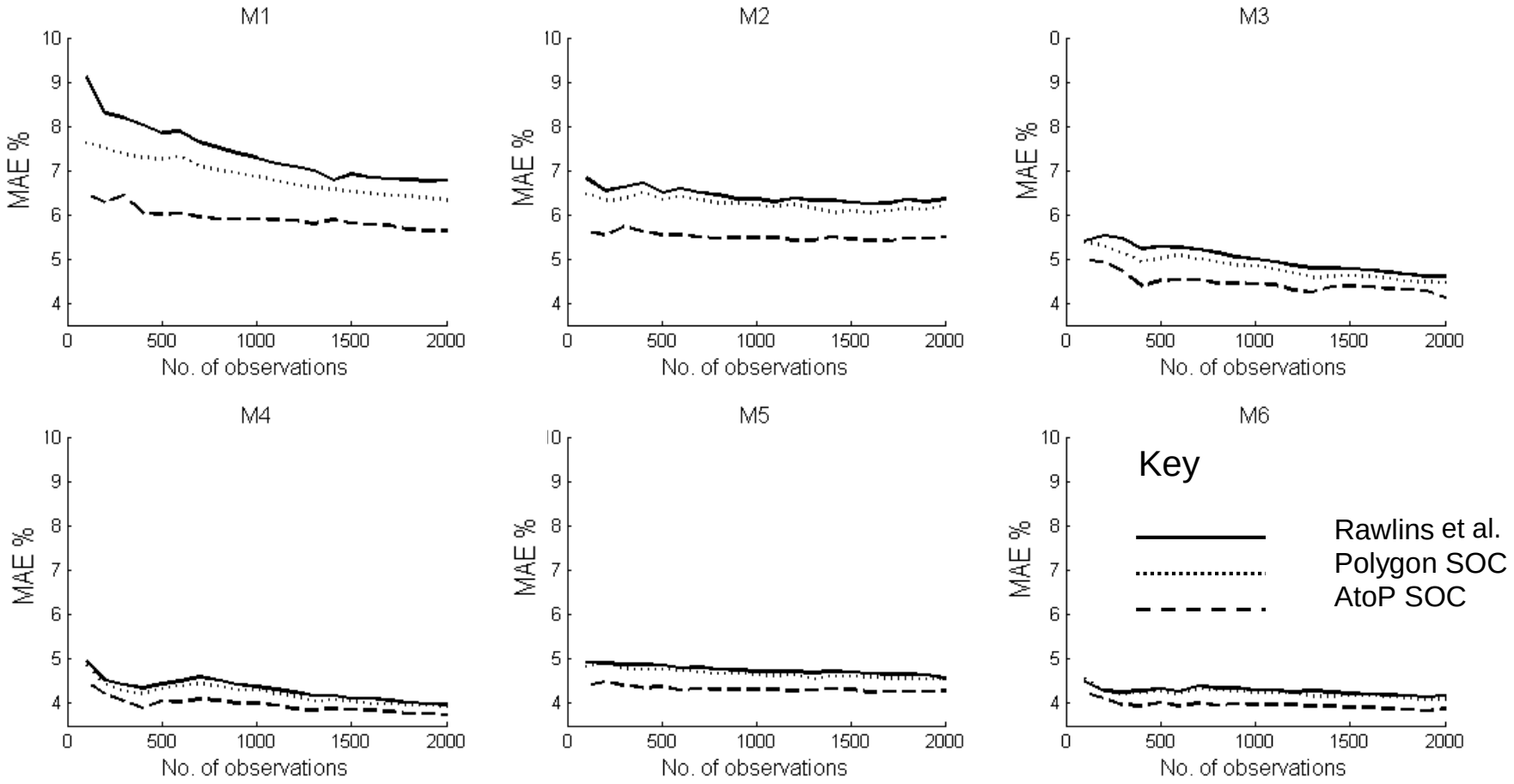

Figure 4. Jack-knife results for all soil types for models 1-6 (M1-M6)

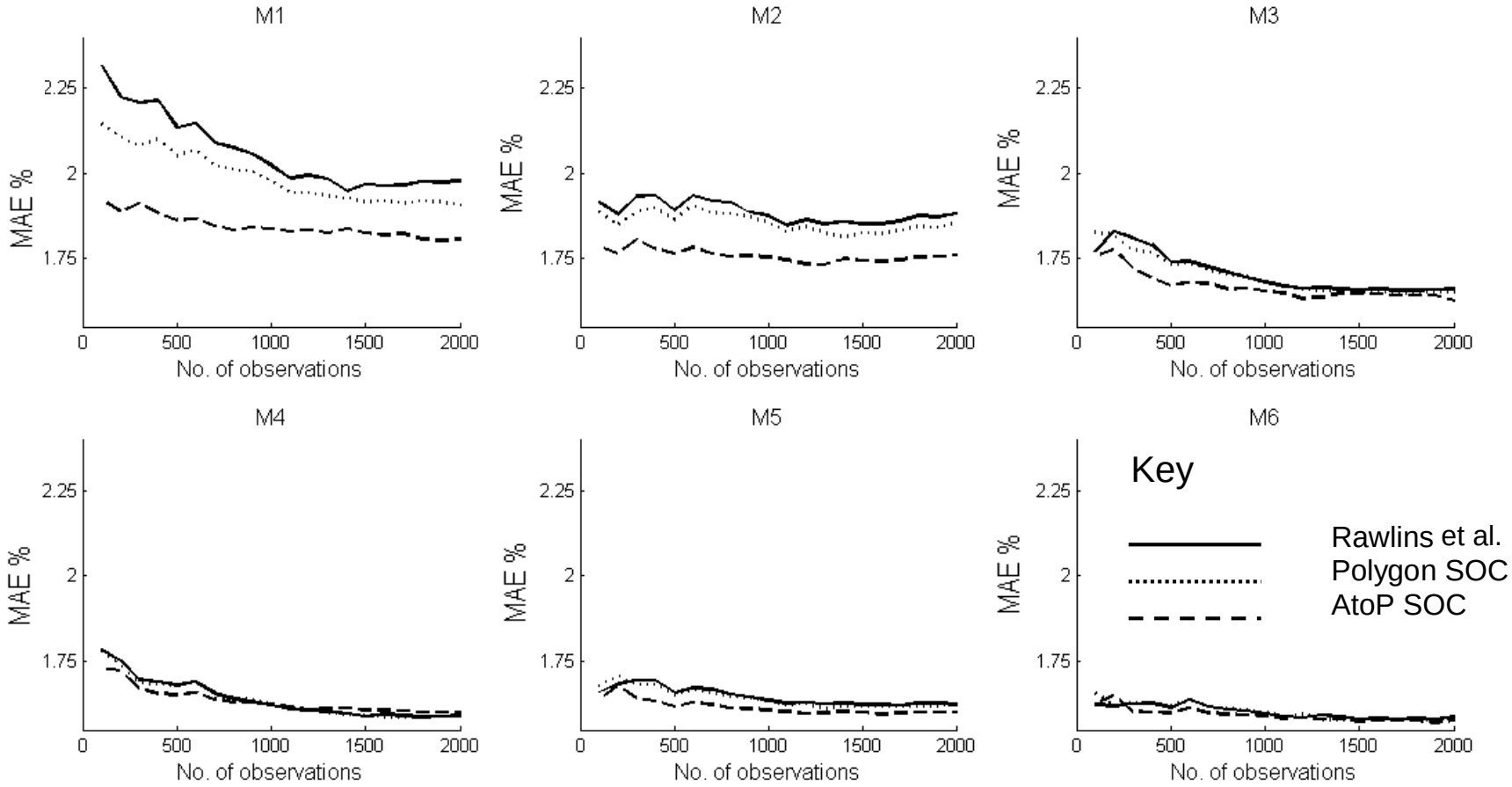

Figure 5. Jack-knife results for mineral soil types for models 1-6 (M1-M6).

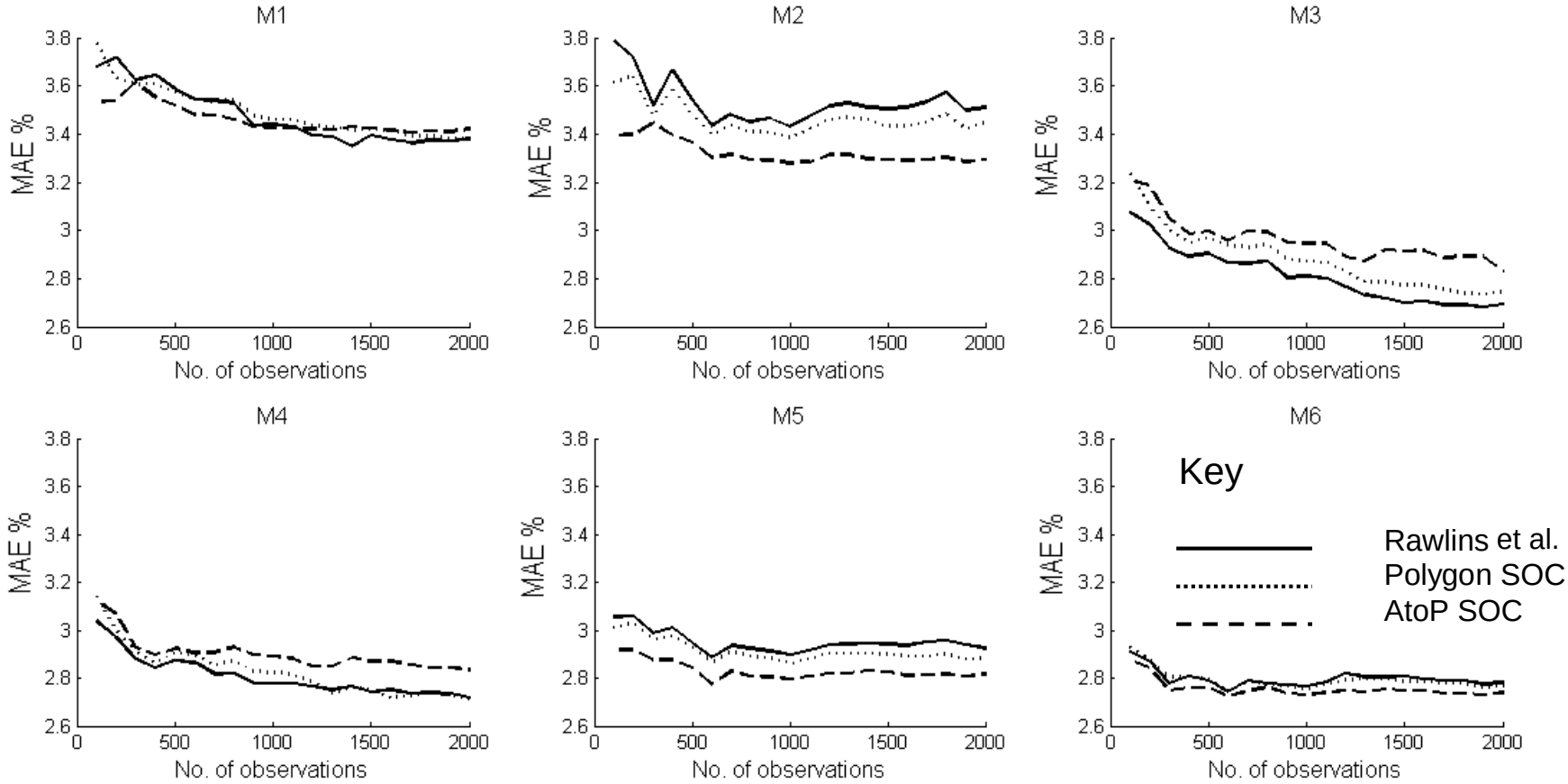

Figure 6. Jack-knife results for organo-mineral soil types for models 1-6 (M1-M6)

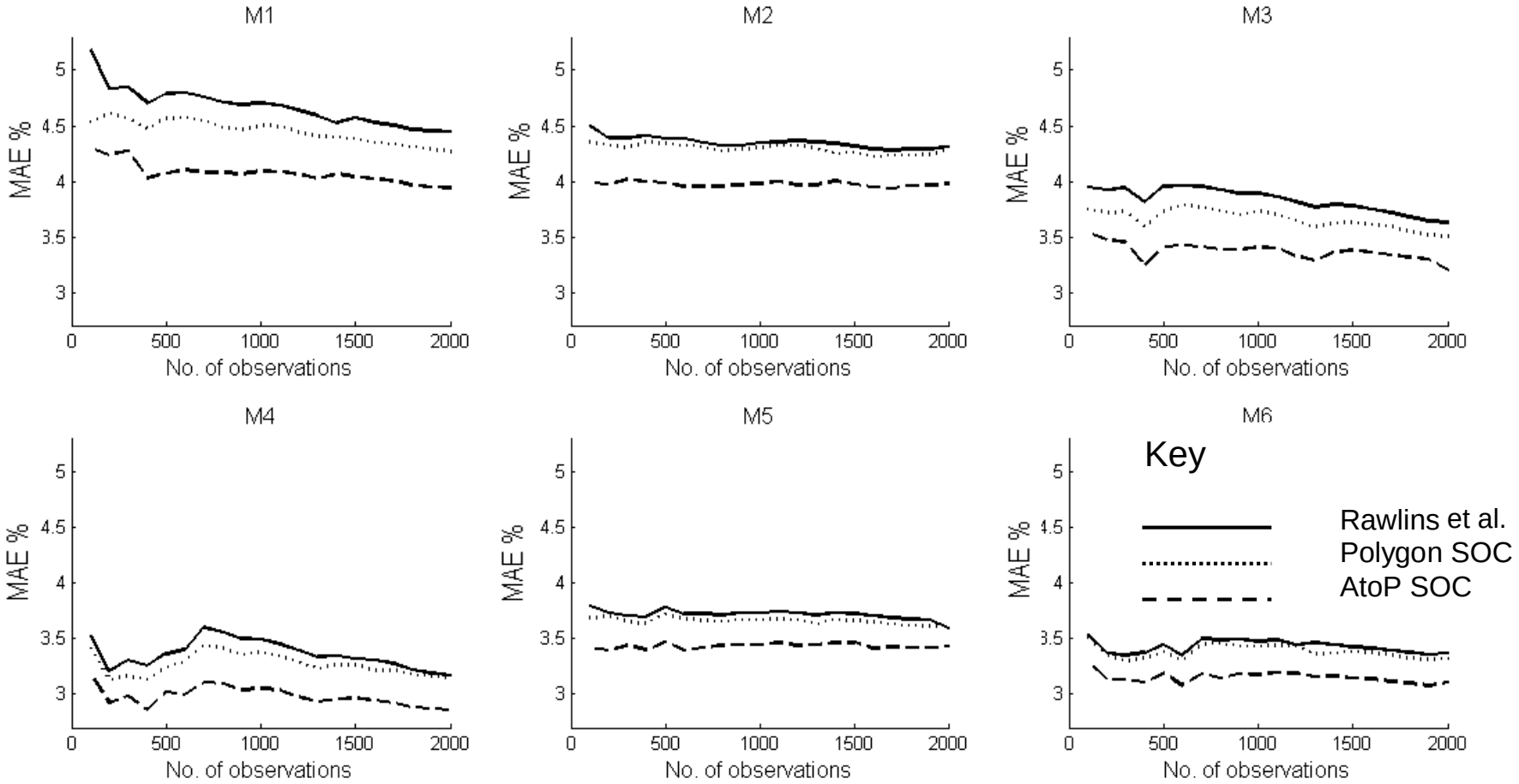

Figure 7. Jack-knife results for peat soils for models 1-6 (M1-M6)

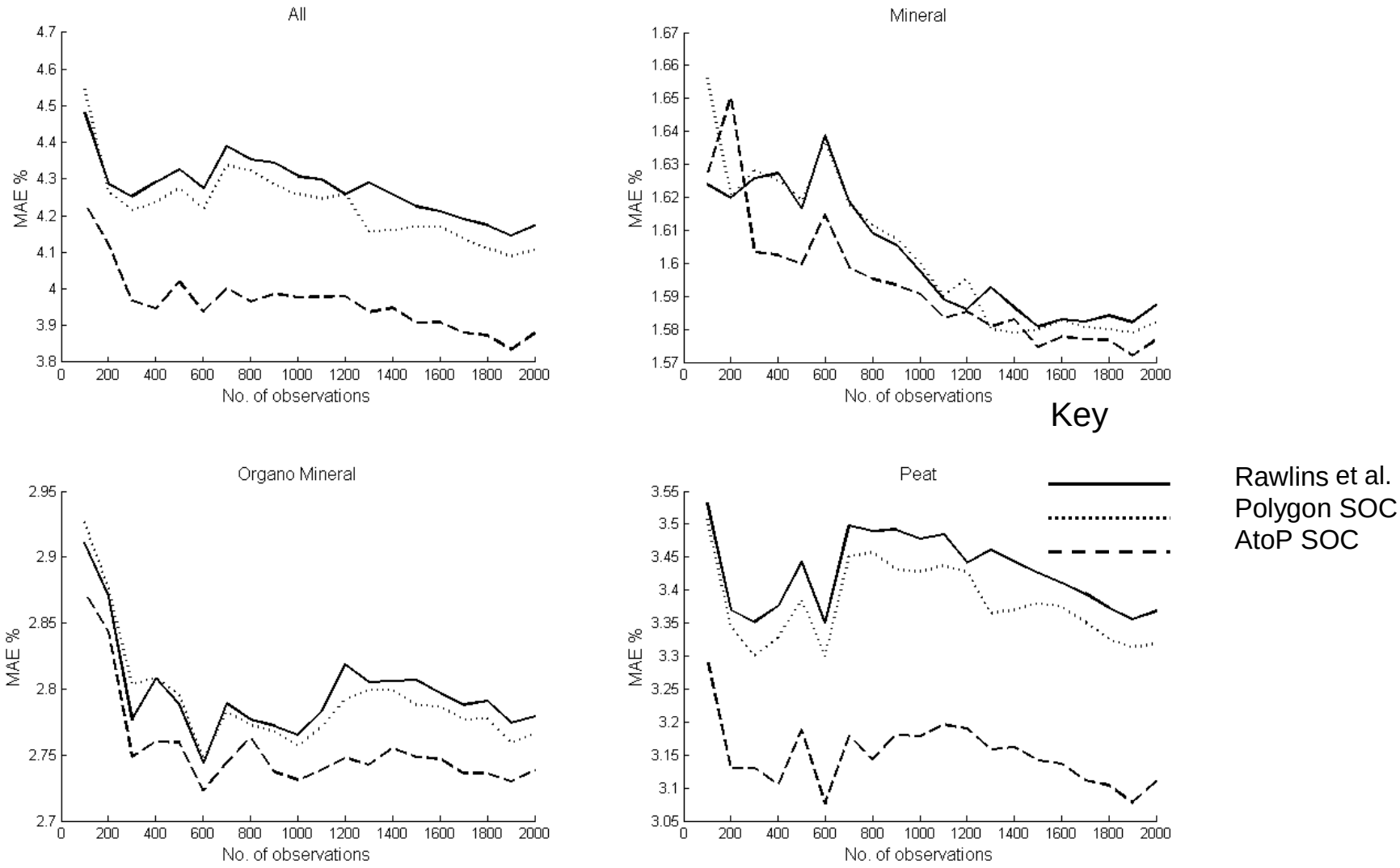

Figure 8. Jack-knife results for model 6 for all soil types

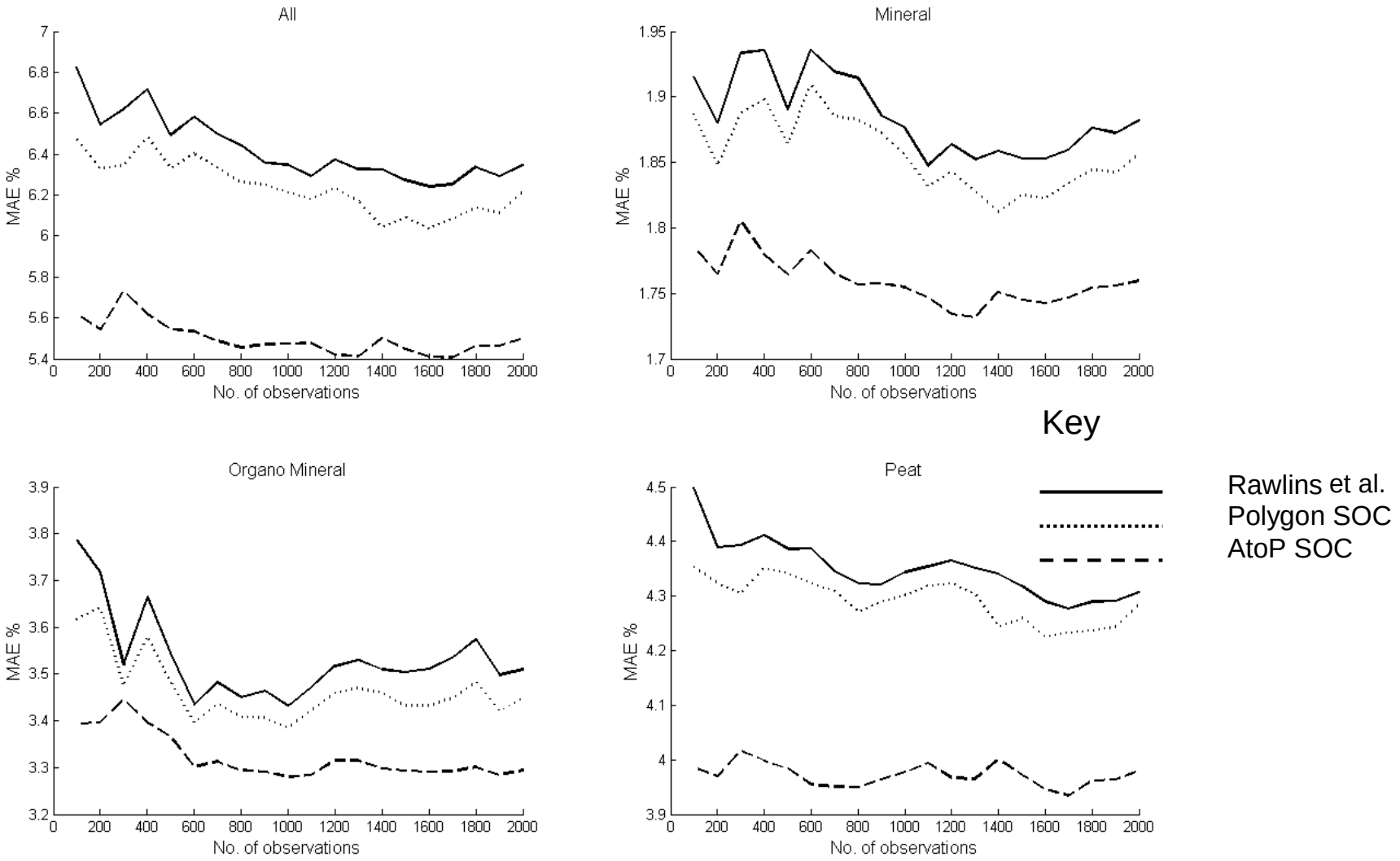

Figure 9. Jack-knife results for model 2 for all soil types