Modeling the spatial distribution of soil properties by Generalized Least Squares regression: Towards a general theory of spatial variates

Santiago Beguería [santiago.begueria@csic.es], Valentina Spanu, Ana Navas, Javier Machín, Marta Angulo-Martínez [1]

Abstract: Assessing the spatial distribution of soil properties has achieved considerable interest among soil scientists, both for testing hypotheses about the soil formation processes and for predicting the properties of soils at non-sampled locations (mapping). In this paper we provide a discussion of the various approaches to the modeling of spatial variates, and we propose a modeling framework that is able to incorporate the most important effects usually found in spatial variates, including fixed and random spatial effects, spatial trends and heteroscedasticity. We provide a case study of the analysis of eight soil properties in a mountain catchment in the Spanish Pyrenees. As explanatory covariates we use several topography parameters, which can be related to the pedogenetic processes active in the area. Several of them proved useful for explaining the variability of soil properties, explaining up to 77% of their variance. We focus on the importance of model selection in order to determine which effects are relevant for modeling each soil parameter. We find that the full model is not necessarily optimum for all the variables tested, and that the model should be adapted to the complexity of each individual case. This paper is a contribution to the discussion on the modeling of spatial variates, and to the eventual development of a general theory of spatial variates.

Keywords: Soil properties, Soil mapping, Soil prediction, Spatial interpolation, Mixed-effects model, Generalized Linear Model, Geostatistics, Regression Kriging

1. Introduction

Assessing the spatial variation of soil properties has received considerable attention from soil scientists for many years. This is due to a great interest in the distribution of soil properties as environmental resources or as soil quality indicators (i.e.prediction); but also as a means of testing hypothesis regarding the influence of external parameters on soil formation processes (i.e.inference). More recently, there has been renewed interest in the topic because of the need for producing reliable maps of soil properties for spatially distributed ecological modeling. There is a general lack of detailed soil information at the appropriate spatial resolution and this is seen by many as one of the most important limitations for developing process-based simulation models. In fact, it is one of the main reasons for the development of spatial interpolation techniques (Burrough, 1993). Examples of soil properties which have been the subject of analysis include: genetic features such as the soil depth, soil type, or the presence of diagnostic horizons (Bourennane et al., 2000; Zhu et al., 2004); physical properties such as soil color, texture, total porosity or water content

[1] Santiago Beguería is a tenured scientist at Estación Experimental de Aula Dei—Agencia Consejo Superior de Investigaciones Científicas (CSIC) in Zaragoza, Spain. Valentina Spanu is a postdoctoral researcher at the Dipartimento di Scienze della Terra—Universita degli Studi di Cagliari, Cagliari, Italy. Ana Navas is a researcher scientist at Estación Experimental de Aula Dei—Agencia Consejo Superior de Investigaciones Científicas (CSIC) in Zaragoza, Spain. Javier Machín is a tenured scientist at Estación Experimental de Aula Dei—Agencia Consejo Superior de Investigaciones Científicas (CSIC) in Zaragoza, Spain. Marta Angulo-Martínez is a post-doctoral researcher at Technical University of Liberec, Czech Republic.

(Pachepsky et al., 2001; Bragato, 2004; Selige et al., 2006); physical-chemical properties such as electrical conductivity, cation exchange capacity or pH (Bishop et al., 2001; Goovaerts et al., 2002; Emery, 2006); chemical composition (Gessler et al., 1996 ; McKenzie and Ryan, 1999; Bishop et al., 2001; McBratney et al., 2003; Selige et al., 2006; Simbahan et al., 2006; Navas et al., 2008); organic matter (Lark, 2000); soil salinity (Corwin et al., 2003; Douaik et al., 2005; Taylor and Odeh, 2007); soil erosion (Ziadat, 2007); or contaminant concentrations (Goovaerts, 1999; Van Meirvenne and Goovaerts, 2001; Hang et al., 2002; Amini et al., 2005).

Studies on the spatial variability of soil properties are typically based on known values of these properties at a number of spatial locations obtained by field sampling and the relationships between these and other—auxiliary—covariates. Several mathematical models have been proposed for estimating the values of soil properties at unvisited locations, and the plethora of literature on this topic is reflective of this (see McBratney et al., 2003, for a comprehensive review). In general terms, two main methodological currents or lines of thought have prevailed: (1) standard statistical approaches; and (2) geostatistical approaches. The standard statistical approach includes tools from the family of univariate and multivariate regression methods as well as adaptive data-mining methods such as generalized additive models, splines, regression trees, artificial neural networks and support vector machines. These methods are especially suited to exploring the relationships among soil properties and other environmental factors and testing hypothesis concerning the variability of soil formation processes, and this is reflected in several seminal works (Yaalon, 1975; Webster, 1977) as well as in modern applications (Park and Vlek, 2002). However, these methods have also been extensively used with emphasis on their prediction abilities for generating maps of soil properties (Moore et al., 1993; Gessler et al., 1996; McKenzie and Ryan, 1999; Selige et al., 2006; Ben-Dor et al., 2006; Ziadat, 2007; Behrens and Scholten, 2007; Mackenzie and Gallant, 2007).

On the other hand geostatistical techniques exploit the self-correlation properties of spatial variates (Cressie, 1993). They are specifically suited for spatial prediction (spatial interpolation), and were developed in close relationship with the Earth Sciences. Geostatistical methods were applied to soil properties data since the 1960s (Davies and Gamm, 1969) but became more extensively from the 1980s onwards due to the popularity of kriging and their implementation in several commercial geographic information systems (GIS) packages (Burgess and Webster, 1980a, 1980b; Webster and Burgess, 1980; Vauclin et al., 1983; Goovaerts, 1994; Sinowski and Auerswald, 1999; Goovaerts et al., 2002; Pebesma, 2004; Douaik et al., 2005).

Extensive research has been devoted to comparing between different statistical and geostatistical techniques (Laslett et al., 1987; Leenaers et al., 1990; Weber and Englund, 1992; Knotters et al., 1995; Gotway et al., 1996; Goovaerts, 1999a, 1999b; Bourennane et al., 2000; Bishop and McBratney, 2001; Vicente-Serrano et al., 2003; Amini et al., 2005; Emery, 2006; Simbahan et al., 2006; Behrens and Scholten, 2007; Taylor and Odeh, 2007; Beguería and Pueyo, 2009). Both statistical and geostatistical approaches have proven to give good results, although some authors have warned against too simplistic formulations, such as linear regression with uncorrelated errors or simple kriging, which do not adapt well to the complexity usually found in spatial variates (Opsomer et al., 1999; Lark, 2000; Hengl et al., 2004). This has motivated the development of advanced methods sharing characteristics of both standard regression and geostatistics. Examples of such approaches are two-step formulations using ordinary regression followed by simple kriging on the residuals (Knotters et al., 1995; Carré and Girard, 2002; Inakwu et al., 2007); kriging with external drift (Bourennane et al., 1996; Hengl et al., 2003); the IRF-k theory (Matheron, 1982); universal and regression kriging (Knotters et al., 1995; Odeh et al., 1995; Hengl et al., 2004; Herbst et al., 2006; Heuvelink et al., 2006; Hengl et al., 2007); and mixed effects models including spatially-correlated errors (Lark, 2000; Pachepsky et al., 2001; Beguería and Pueyo, 2009). These techniques allow integrating all the information available to the researcher, from the purely spatial information from the field survey of dependent variables to background pedological knowledge in the form of relationships with environmental covariates. There is still, however, a lack of a general framework capable of integrating all these approaches into a common theory of spatial variables.

In practice, the selection of a given statistical model continues to be highly subjective and depends, among others things, on personal factors such as training and software availability. Ideally though, these factors should be of secondary importance to other considerations such as the purpose of the study (e.g. whether the stress is on hypothesis testing or just on spatial prediction); the sampling design; the previous knowledge of the variates being modeled; or the expected correlation with covariates. In this paper we propose a binary decision tree with criteria for helping decide which method is best suited for a given data set configuration. We use generalized least squares (GLS) regression to fit a mixed-effects model with spatially-correlated errors and heteroscedasticity (different variances) to assess the spatial variability of eight soil properties in a small mountain area in the Spanish Pyrenees. A model selection procedure is designed to help find the minimal adequate model for each soil property, from which best linear unbiased predictor maps are produced. This method is equivalent to what has been usually referred to as *regression kriging* (Schabenberger and Gotway, 2005; Hengl, 2007).

The study area corresponds to an experimental catchment for which a good level of knowledge exists (García-Ruiz et al., 1995; González et al., 1997; García-Ruiz et al., 2005; Navas et al., 2005a, 2005b; Lasanta et al., 2006; Navas et al., 2008). Like many other areas in the Spanish Pyrenees, the catchment supported intensive human use during the past centuries (what has been termed 'the traditional land use system'), but the land was rapidly abandoned during the first half of the 20th Century. Improving our knowledge about the spatial distribution of soil properties in areas that experienced changes in land use / land cover may help determining: i) how past land use arrangements affect soil quality, and ii) how the spatial variation of soil properties affects the regeneration of the natural vegetation after land abandonment.

2. Data and methods

Study area

The study was carried out in the Arnás River catchment, a small first-order stream in the Spanish Central Pyrenees (Figure 1). The catchment occupies an area of 2.84 km^2 and it corresponds to a middle mountain area with elevation ranging between 910 and 1341 m a.s.l. Average annual temperature is 10ºC, and mean annual precipitation is 930 mm. The Arnás River divides the catchment in two sides with contrasting topographical characteristics. To the left of the river (sunny side) the slopes are shorter and steeper, contrasting with the more gentle slopes found on the right (shady side).

Vegetation on the sunny side is composed of a dense scrubland with sparse trees or small patches of pine forest, whereas on the shady side there are large patches of pine and mixed forests, especially in the upper part of the slopes. Most of the catchment was cultivated until the mid 20th Century and was abandoned thereafter. Since then, vegetation recovery occurred naturally, especially on the shady side.

The geology of the catchment corresponds to the Eocene Flysch, with thin—centrimetric to decimetric— alternating layers of marls and sandstones. There are six soil types with the most abundant being Haplic Kastanozems, Calcaric Regosols and Rendsic Leptosols (Figure 2). In general, the soil texture is clay loam and the soils tend to be alkaline (Navas et al., 2005a).

Data

A regular sampling scheme was devised in order to obtain an evenly spaced sample (Figure 1). The sample consisted of 74 points, separated at distances of 100 m from each other. The sampling was carried out during one intensive field campaign in order to restrain variation in the climatic and hydrological conditions. The samples were stored at 4ºC until they were analysed. Samples were air-dried, ground, homogenized and quartered, to pass through a 2 mm sieve. Eight soil properties were determined for each sampling point: i) clay content of the fraction below 2 mm (%); ii) bulk density (g cm^{-3}); iii) carbonates content (% $CaCO_3$); iv) pH; v) field capacity (% of the soil volume); vi) organic matter content (%); vii) nitrogen content (%); and viii) cation exchange capacity (meq g^{-1}). These properties were measured following standard techniques. Grain size analysis was performed using Coulter laser equipment. To eliminate the organic matter, samples were chemically disaggregated with 10% H_2O_2 heated at 80ºC, then stirred while ultrasound was also used to facilitate particle dispersion. Carbonates were measured using a pressure calcimeter. The pH (1:2.5 soil:water) was measured using a pH-meter. Water retention at field capacity was determined at -33 kPa using a Richards Membrane. Organic matter was determined by titration. Total nitrogen was measured using the Kjeldhal Method. To determine the exchangeable capacity by sodium displacement, a Mg $(NO_3)_2$ solution was used followed by ICP-OES analysis. Basic statistics of the eight soil properties are shown in Table 1.

A map of soil types was prepared based on field survey and soil profiles. Six soil types were identified and classified according to FAO (1989), and their spatial extension was mapped. A digital elevation model (DEM) with a 5 m spatial resolution, generated from photogrammetric restitution, was employed for deriving a set of topographic parameters. Elevation was used as the primary covariate, due to its influence in several climatic variables relevant for pedological processes such as rainfall amounts or temperature. Wang and Liu's (2006) algorithm was then applied to the original DEM for filling small depressions and obtaining a hydrologically continuous model. The first derivatives of the elevation (slope gradient and aspect) were computed following the method of Zevenbergen and Thorne (1996). The slope gradient is related with erosion and deposition processes, and usually is correlated with soil depth and other soil properties. The aspect, on the other hand, is related to the amount of solar energy received by the slope. A cosine transformation was applied to the aspect for transforming from an angular to a linear scale, thus stressing the variability in the north-south axis. A multiple flow direction algorithm was used for computing the catchment area, with a concentration exponent of 1.1 and a threshold of 1500 cells for the initiation of concentrated flow in line with the approach suggested by Freeman (1991) and Quinn (1991). The catchment area is related to the accumulation of water and sediment flows in the landscape, thus having an influence on pedogenic processes. The topographic wetness index (Moore et al., 1991) was finally derived from the previous variables. It is related to pedogenic processes such the redistribution of soil moisture in the landscape or soil erosion and accumulation. All topographic analyses were performed using the following modules of SAGA GIS (Böhner et al., 2006): *fill sinks*, *local morphometry*, *parallel processing*, and *topographic indices*. Basic statistics of the topographical covariates are shown in Table 1, and their spatial distribution is shown in Figure 2.

Exploratory analysis

We performed a preliminary analysis in order to determine the main factors of variation for each soil property. Only four soil types (calcaric Regosols, rendsic Leptosols, haplic Kastanozems and haplic Phaeozems) were considered, since the other two soil types found in the area occupied a small surface and were not represented in the soil sampling. We performed ANOVA tests in order to check the explanatory capacity of the soil classification. A pairwise t-test using the Holm (1979) method for adjusting the p-values in multi-contrast analyses was then applied on those soil properties yielding positive results in order to determine the pairs of soil types that were different. The Levene test for homogeneity of variances was used for checking against heteroscedasticity among soil types. Pearson's correlation was used for exploring pairwise relationships between the soil properties and the covariates. Bivariate plots were used for exploring differences in these relationships between soil types. Finally, the Moran's I test was used for checking against spatial autocorrelation of the soil properties.

Exploratory analysis can only give suggestions about the effects on the variability of the dependent variables and help defining the structure of the models: i) whether heteroscedasticity and spatially correlation must be included in the error term, and ii) which covariates and interactions among them to include. Moreover, some of the analysis such as ANOVA assume independence, normality and homoscedastic errors. The significance and sign of these effects can only be determined by a mixed-effects model analysis.

Mixed-effects model analysis

It is possible to formulate a universal model for a spatial variate as a sum of deterministic (fixed) and stochastic (random) components in the following form:

$$Z(s) = m(s) + \varepsilon'(s) + \varepsilon''(s), \qquad \text{(eq. 1)}$$

where Z is the value of a spatially-explicit variate depending on the spatial coordinates s, m is a deterministic function of spatial variation containing the relationship with covariates and whose parameters may or may not vary spatially, ε' is a stochastic component of spatial variation (a spatially correlated error), and ε'' is a random (uncorrelated) error. This model is often termed *regression kriging* (Hengl, 2007) and is virtually identical to a linear mixed-effects model with spatially-correlated errors (Pinheiro and Bates, 2000). Several methods such as ordinary least squares (OLS) regression and simple and ordinary kriging are in fact special (incomplete) cases of this model. In OLS regression the fixed effects (relationship with covariates) is captured but the spatial random variation is ignored, with negative effects on the confidence intervals of both parameter estimates and predicted values (Beguería and Pueyo, 2009). In simple and ordinary kriging the spatial random variation is captured but the relationship with covariates is not considered.

A binomial decision tree such as the one in Figure 3 can help deciding which technique is most suitable for assessing the spatial variability of a soil parameter. If a physical (deterministic) model is available it is preferable to any statistical method because it usually allows for a deeper insight on the processes responsible for the variability of the soil parameter of interest. For example, mass balance models have been used for assessing profiles of ^{137}Cs activity (Soto and Navas, 2004 and 2008) or soil erosion and deposition (Alatorre et al., 2011). If a physical model is not available a statistical approach is the only option. If the soil parameter is expected to be correlated with other environmental variables this relationship can be used for modeling its variability. OLS regression has been extensively used for that, but it is not recommended for spatial variables for the reasons stated above, so methods based on the Generalized Least Squares (GLS) algorithm such as a mixed-effects model or regression kriging are to be preferred. If no correlation exists with other environmental variables or if data on other variables is not available, the stochastic random variation of the soil property of interest can be modeled by geostatistical methods such as ordinary or simple kriging, inverse distance weighted local regression, splines, etc. If the soil property shows no spatial correlation at the sampling scale it is not possible to go beyond the null model (the sample mean) unless additional data are obtained.

For our case study, with four continuous covariates and one factor (four soil types) and presumed spatial correlation and heteroscedasticity (unequal variances) we defined a full model in the following form:

$$z(j,s) = \beta(j)\,x(s) + \varepsilon(j,s), \qquad \varepsilon(j) \sim N\left(0,\sigma_j^2\Lambda\right), \qquad \text{(eq. 2)}$$

where $z(j,s)$ is a realization of Z associated with location s and soil type j; $\beta(j)$ is a vector of regression coefficients that includes an intercept and vary among soil types; x is a vector of covariates at location s; and $\varepsilon(j,s)$ is a spatially-dependent, heteroscedastic error term. Unlike the standard linear model in which the errors are independent and identically distributed, the model in equation 2 allows different variances according to the levels of the factor (heteroscedasticity) and autocorrelation through the variance-covariance matrix $\sigma_j^2 \Lambda$. Such a model can be fit by GLS (Pinheiro and Bates, 2000) using maximum likelihood (ML) or, best, restricted maximum likelihood (REML) methods. Best linear unbiased predictions (BLUPs) at a given location s can be obtained including the fixed effects (covariates) and the spatial random effects estimated from known measurements of the response variate at nearby locations. These are different from the best linear estimates (BLUEs), which include only the fixed effects.

For modeling the spatial dependence of the errors it is common to assume that the correlation between two error values ε_s and $\varepsilon_{s'}$ depends on the Euclidean distance d between their location vectors, (s,s'), and a correlation parameter, ρ. As it is classical in the geostatistical literature, we express the error correlation structure through the semivariogram, i.e. a model of the variance of the difference between two values of ε at different spatial locations depending on the distance between them (Cressie, 1993). Several options exist for modeling the empirical semivariogram arising from the data (variogram models): linear, Gaussian, spherical, exponential, etc. Moreover, the characteristics of the semivariogram may vary in space (non-stationarity) or depend on the direction (anisotropy).

With complex models such as the one described above a model selection procedure is needed to help determining which effects are really relevant for the soil parameter under study: i.e. find which covariates and covariate interactions are significant and decide whether or not a random effect (spatially-correlated errors) and heteroscedasticity must be included in the model. The objective is not only to simplify the model as much as possible by removing unnecessary parameters, but also to achieve appropriate *p*-values for the covariates since the power of GLS analysis increases when only the significant effects are left (Crawley, 2007). Model selection consists on finding an optimum model configuration in which only the significant effects are included. Here we followed a top-down strategy (Diggle et al., 2002; Zuur et al., 2009):

1. Beyond optimal model. A full model including all the covariates and meaningful interactions, as well as heteroscedasticity and spatial correlation of the errors was fit for each dependent variable. The models were fit by GLS using the REML method. Alternative models with different semivariogram models were fit, and a likelihood-ratio test was used for choosing the most appropriate one. These models contained presumably a higher number of effects than really needed, i.e. they included non-significant effects, so the p-values obtained for the covariates are not to be trusted. In the following steps all non-significant effects were removed until an optimal model was achieved.
2. Covariates selection. The t-statistic was used for determining the significance of covariates, and those that did not achieve significance at a confidence level α=0.01 were removed. The covariate with the highest p-value was removed each time, and the process was iterated until only significant covariates were left.
3. Residual model. Alternative models to the best one obtained from step 2 were fit with no heteroscedasticity and no spatially correlated errors. A likelihood-ratio test was then used to compare between the models and determine which was the best configuration of the residual model.
4. Optimum model. The model arising from step 3 was considered the best one and used for further analysis (cross-validation, prediction maps, etc).

For best clarifying this process, the R code used for performing model selection and needed data are provided as online supplementary material to this article.

Leave-one-out cross-validation was used to check the ability of the fitted models to predict the values of soil properties at non-sampled locations. This procedure involved fitting the model as many times as samples are in the data set, but each time keeping one sample out of the training sample. This allows computing independent validation statistics such as the mean absolute error (MAE), the mean bias error (MBE) and the root mean square error (RMSE) from the left-out samples.

Maps of best linear unbiased estimations (BLUEs) were produced from the optimum fitted models using the functions for regression kriging in the `gstat R` library (Pebesma, 2004). Sample data and code for replicating the analysis are provided as online supplementary material to this article.

3. Results

Exploratory analysis

The exploratory analysis provided evidence of differences in mean and variance, as well as spatial autocorrelation, for the eight soil properties and for the topographic covariates (Table 2 and Figure 4). Haplic Phaeozems were the most different soils, with higher values of clay content, organic matter, field capacity, nitrogen content and cation exchange and lower bulk density, carbonates content and pH. They appeared in medium to high areas of the shady aspect slopes of the catchment, with moderately high values of the topographic wetness index. Haplic Kastanozems also appeared in the shady aspect of the catchment, but on the lower and gentler parts of the slopes, with high values of the topographic wetness index, and their soil properties were average. Rendsic leptosols were characteristic of the higher parts of the sunny aspect slopes, with low values of the topographic wetness index. They could have high values of field capacity, organic matter and cation exchange capacity and relatively low carbonate content and bulk density, but they also showed a very large variance. Calcaric Regosols appeared in preference in the lower slopes of the sunny side of the catchment, with relatively low values of the topographic wetness index. They had the highest bulk density and pH and low field capacity, nitrogen and organic matter content. Globally, the soil classification seemed to have a relatively good capacity for predicting the soil properties, although it was well correlated to the topographic covariates too.

Good correlations were found among soil properties, which were lower in the case of the clay content (Figure 5). Significant correlations were also found between the soil properties and several topographic covariates, especially with the altitude and the slope aspect, suggesting a good predictive capacity. Significant correlations were also found between the slope gradient and the aspect, and between the wetness index and the rest of topographic covariates.

It must be noted that all of these factors of variability were checked in the exploratory analysis *as being independent from each other,* although some of them could be interrelated. For example, spatial autocorrelation in the dependent variables could be caused–and be totally explained–by spatial autocorrelation of the covariates. The methodological approach followed guaranteed that only the significant effects would be retained in the final models, allowing for a better interpretation of the results. As suggested by the exploratory analysis, the whole set of covariates were included in the analysis–soil type, altitude, slope, aspect, wetness index–, as well as the interactions between aspect and altitude and altitude and slope. Other possible interactions among topographical covariates were rejected due to the correlations between them. Interactions between the soil types and the topographic covariates were also not included for the same reason.

Mixed-effects analysis: selection of covariates and residual model

The process of model selection allowed determining the optimum model configuration for each soil parameter (Table 3). The soil classification (soil type) was only retained for clay and carbonates content, which had different intercepts for the haplic Phaeozems. For the remaining soil properties the optimum models consisted only on combinations of the topographic covariates. The combination of slope aspect (asp) and the interaction between aspect and altitude (asp:alt) determined the optimum model for the bulk density, field capacity, nitrogen content and cation exchange capacity. The slope gradient (slope) or the slope gradient and its interaction with the slope aspect were selected for the carbonates content, pH and organic matter content. The topographic wetness index was not included in any model.

The standard errors (standard deviation of the residuals) of were relatively high when compared to the variance of the dependent variables, for example as expressed by the interquantile range (Table 4 and Table 1, respectively). With respect to the residual models, heteroscedasticity was required for all dependent variables except the clay content and the cation exchange capacity, as demonstrated by a likelihood-ratio test between the models with and without heteroscedasticity, at the confidence level α=0.05. As compared to the calcaric Regosols, the haplic Phaeozems had larger residual variances for the pH (almost four times higher), field capacity, organic matter and nitrogen content, while it was lower for the bulk density and the carbonates content. Rendsic Leptosols had larger residual variances for pH, field capacity and organic matter, and lower for the remaining properties. Haplic Phaeozems had lower residual variances for all soil properties except nitrogen content. Spatial autocorrelation was required for four soil parameters: bulk density, pH, field capacity and organic matter. The best semivariogram models varied between the spherical, gaussian and rational models, and the range parameters varied between 125 and 480 m.

Validation and BLUPs

Validation statistics based on leave-one-out cross-validation (Table 5) yielded low to moderate R^2 values ranging between 0.118 and 0.513, indicating that the prediction ability of the models in unvisited locations was not high. The predictions showed very little bias compared with the observed values (MBE close to zero), and the absolute errors (MAE) and RMSE were lower than the range of variation of the dependent variables. Prediction plots showed relatively good fit between predicted and measured values was higher for some variables such as the pH and the carbonates content, while others such as the clay content or the organic matter were poorly predicted (Figure 6).

Maps of the predicted soil properties were produced based on the optimum models fitted (Figure 7). Maps of the random variabilidy (standard error) were also produced based on the residual models (Figure 8). The spatial distribution of predictions and errors reflected the effects included on each models. The effect of the soil classification was clearly visible for the clay and carbonates content, while in the remaining cases only the topographical effects were present. The contribution of the spatially-correlated random effect was especially noticeable for the field capacity and organic matter.

6. Discussion

A review of the methodological developments for modeling spatial variables reveals a convergence of regression and geostatistical techniques towards mixed approaches that are able to account for fixed and random sources of spatial variation. Such approaches, despite the varying terminology (mixed-effects models, regression kriging), are based on the generalized least squares algorithm (GLS) and facilitate taking advantage of all the information available for estimating the values of spatial variates at non-sampled locations (BLUEs). Compared to ordinary least squares (OLS) regression, the mixed-effects approach allows incorporating spatial autocorrelation on the error term and reduce inference errors (bias in the estimation of parameter confidence levels), and allows obtaining unbiased estimations of the dependent variables. Compared to traditional geostatical procedures such as ordinary or simple kriging, regression-kriging allows incorporating the researcher's pedological knowledge in the form of relationships with other environmental covariates. The main difference between the mixed-effects regression approach and regression kriging is that, while kriging was traditionally focused on predicting the spatial distribution of a variate, the mixed-effects approach was more focused on statistical inference, i.e. on drawing conclusions about the effects explaining the variation of a given spatial propery. This explains that the mixed-effects approach has developed very precise tools for determining which effects (covariates, factor levels, etc) are significant, while these are much less developed in the kriging environment.

In our case we used a mixed-effects analysis approach since we were interested in determining the significance of different sources of variation in our data, including the significance of incorporating heteroscedasticity and spatial correlation in the error term. Starting from a complete model formulation, a model selection procedure allowed removing all non–significant effects to finally achieve an optimum model for the each dependent variate, given the available information.

Despite sharing a homogeneous parent material, the soils of the study area had significant differences with respect to eight soil properties analyzed. Significant relationships were found with the topography except for the clay content, and one soil type (haplic Phaeozems) had significantly different intercepts for two soil properties (clay and carbonates content). In the remaining cases, the differences in soil properties found among soil types during the exploratory analysis were explained by the topography. The error term also had a complex structure, since different standard errors by soil type were needed in most models, and spatial correlation was present in four out of eight cases.

As several authors pointed out, topographic covariates obtained from digital terrain models have a good ability for predicting soil properties (McKenzie and Ryan, 1999; Farenhorst et al., 2003; Leij et al., 2004). The organic matter and the nitrogen content were the soil properties most poorly predicted, a characteristic of soil attributes whose variability is mostly governed by vertical pedogenetic processes and by local variation in ecological properties such as the plant cover (Park and Vlek, 2002). On the other hand soil properties such as the carbonates content were best explained by the topography, as it could be expected from soil attributes which are influenced not only by vertical processes within the soil but also by the lateral movement of surface and subsurface water and soil particles. The slope aspect and its interaction with the altitude had a significant effect on the bulk density, field capacity, nitrogen content and cation exchange capacity, while the slope and its interaction with the altitude were significant for the carbonates content, pH and organic matter. The presence of the altitude in most of the models could be related to the general pattern of precipitation in the catchment that increases with the elevation. Also, a more dense vegetation cover is found in the upper parts of the catchment and is likely related to a higher input of organic matter. Apart from the altitude, either the aspect or the slope was included in most models. Slope is related to the intensity of erosion processes and the accumulation of organic matter and other soil properties, while the aspect controls the energy balance and hence it is related to the hydrology of the soils and the vegetation activity. Given the correlation between them we are probably facing a similar topographic effect here, so it is difficult determining which of the two covariates (or even both) is really affecting the soil properties. Interestingly, the topographic wetness index was rejected consistently in all the models. This could be an indication that topographic properties of cumulative nature are not so important in the area so only short-range pedogenic processes prevail. But, given the correlation between the wetness index and the remaining topographic covariates it could be also that the combined index did not incorporate any new information.

The residual models revealed that the variances varied between soil types. The haplic Kastanozems, found in medium and low slopes on either slope aspects, had a lower variance for almost all soil properties, while the rendsic Leptosols and the haplic Phaeozems, corresponding to forestal soils located near the topographic divides, had larger variances for several soil parameters.

This can be related to a longer and more complex pedogenic evolution of the latter, resulting in increasing spatial variability. The presence of spatial autocorrelation in the residuals in four out of eight models indicates that other sources of spatial variation not included in the analysis might be significant. For example, variations in the parent material, vegetation composition and even in the land use history could help improving the models.

7. Conclusions

While new model formulations have currently been proposed and tested, the question of how to best model the spatial variation of soil properties with the purposes of inference and prediction still remains. Significant advances have been made on the topic by soil scientists, which have been traditionally at the cutting edge of the discipline. In this paper we provided a brief discussion of the various approaches to the modeling of spatial variates and proposed a theoretical framework that is able to incorporate the most important effects usually found in spatial variates, including fixed and random spatial effects, spatial trends and heteroscedasticity. Here we used a mixed effects regression approach fitted by the generalized least squares (GLS) algorithm. We discussed the nature of the different effects and provided a practical example through a case study. We found that the full model is not necessarily optimum for all the variables tested and that the model should be adapted to the complexity found on each particular case. As such, this paper intends to be a useful contribution to the discussion on the modeling of spatial variates and to the development of a general theory of spatial variates.

Acknowledgments

This work has been supported by the following research projects: EROMED (CGL2011-25486) and DISDROSPEC (CGL2011–24185) financed by the Spanish Commission of Science and Technology (CICYT) and FEDER, ChangingRISKS (OPE00446/PIM2010ECR-00726) financed by EU ERA-NET CIRCLE Programme, and Grupo de Excelencia E68 financed by the Aragón Government. Research of M. A.-M. is supported by a research grant from the Spanish National Research Council (JAE-Predoc, CSIC).

References

Akaike, H., 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19(6), 716–723.

Alatorre, L.C., Beguería, S., Lana-Renault, N., Navas, A., García-Ruiz J.M., 2011. Soil erosion and sediment delivery in a mountain catchment under scenarios of land use change using a spatially distributed numerical model. Hydrology and Earth System Sciences 16, 1321–1334, 2012.

Amini, M., Afyuni, M., Fathianpour, N., Khademi, H., Flühler, H., 2005. Continuous soil pollution mapping using fuzzy logic and spatial interpolation. Geoderma 124, 223–233.

Beguería, S., Pueyo, Y., 2009. A comparison of simultaneous autoregressive and generalized least squares models for dealing with spatial autocorrelation. Global Ecology and Biogeography 18(3), 273–279.

Behrens, T., Scholten, T., 2007. A comparison of data-mining techniques in predictive soil mapping. In: P. Lagacherie, A.B. McBratney and M. Voltz (Editors), Digital soil mapping. An introductory perspective. Elsevier, Amsterdam, pp. 327– 349.

Ben-Dor, E., Levin, N., Singer, A., Karnieli, A., Braun, O., Kidron G. J., 2006. Quantitative mapping of the soil rubification process on sand dunes using an airborne hyperspectral sensor.

Geoderma 131, 1–21.
Beven, K.J., Kirkby, M.J., 1979. A physically based, variable contributing area model of basin hydrology. Hydrol. Sci. Bull. 24, 43–69.
Bishop, T.F.A., McBratney, A.B., 2001. A comparison of prediction methods for the creation of field-extent soil property maps. Geoderma 103, 149–160.
Böhner, J., McCloy, K.R., Strobl, J. (Eds.), 2006. *SAGA – Analysis and Modelling Applications*. Göttinger Geographische Abhandlungen, Vol. 115, 130 pp.
Bourennane, H., King, D., Chery, P., Bruand, A., 1996. Improving the kriging of a soil variable using slope gradient as external drift. Eur. J. Soil Sci. 47, 473–483.
Bourennane, H., King, D., Couturier, A., 2000. Comparison of kriging with external drift and simple linear regression for predicting soil horizon thickness with different sample densities. Geoderma 97, 255– 271.
Bragato G., 2004. Fuzzy continuous classification and spatial interpolation in conventional soil survey for soil mapping of the lower Piave plain. Geoderma 118, 1–16.
Burgess, T.M., Webster, R., 1980a. Optimal interpolation and isarithmic mapping of soil properties, I. The semivariogram and punctual kriging. J. Soil Sci. 31, 315–331.
Burgess, T.M., Webster, R., 1980b. Optimal interpolation and isarithmic mapping of soil properties, II. Block kriging. J. Soil Sci. 31, 331–341.
Burrough, P.A., 1993. Soil variability: a late 20th century view. Soils and Fertilizers 56, 529–562.
Burrough P.A, McDonnell R.A., 1998. Principals of Geographical Information Systems. Oxford University Press, 333 p.
Carré, F., Girard, M.C., 2002. Quantitative mapping of soil types based on regression kriging of taxonomic distances with landform and land cover attributes. Geoderma 110, 241–263.
CSIC, 1976. Comisión de métodos analíticos. Anales Edafol. Agrobiol. 35, 813–814.
Corwin, D.L., Kafka, S.R., Hopmans, J.W., Mori, Y., van Groenigen, J.W., van Kessel, Lesch, S.M., Oster, J.D., 2003. Assessment and field-scale mapping of soil quality properties of a saline-sodic soil. Geoderma 114, 231–259.
Crawley, M.J., 2007. *The R Book.* John Wiley and Sons, Chichester (UK), 942 p.
Cressie, N.A.C., 1993. Statistics for spatial data. Wiley, New York.
Davies, B.E., Gamm, S.A., 1969. Trend surface analysis applied to soil reaction values from Kent, England. Geoderma 3, 223–231.
Desmet, P.J.J., Govers, G., 1996. A GIS Procedure for Automatically Calculating the USLE LS Factor on Topographically Complex Landscape Units, Journal of Soil and Water Conservation 51(5), 427–433.
Douaik, A., Van Meirvenne, M., Tóth, T., 2005. Soil salinity mapping using spatio-temporal kriging and Bayesian maximum entropy with interval soft data. Geoderma 128, 234–248.
Emery, X., 2006. Ordinary multigaussian kriging for mapping conditional probabilities of soil properties. Geoderma 132, 75–88.
FAO, 1989. Soil map of the World. Revised legend. FAO, Rome.
Farenhorst, A., Florinsky, I. V., Monreal, C. M., Muc, D., 2003. Evaluating the use of digital terrain modelling for quantifying the spatial variability of 2,4-D sorption by soil within agricultural landscapes. Canadian Journal of Soil Science 83(5), 557–564.
Freeman, G.T., 1991. Calculating catchment area with divergent flow based on a regular grid, Computers and Geosciences 17, 413–422.
Fotheringham, A.S., Brundson C., Charlton, M., 2002. Geographically weighted regression. John Wiley and Sons, New York.
García-Ruiz, J.M., Arnáez, J., Beguería, S., Seeger, M., Martí-Bono, C., White, S., 2005. Runoff generation in an intensively disturbed, abandoned farmland catchment, Central Spanish Pyrenees, Catena 59(1), 79–92.
García-Ruiz, J.M., Lasanta, T., Martí, C., González, C., White, S., Ortigosa, L., Ruiz-Flaño, P., 1995. Changes in runoff and erosion as a consequence of land-use changes in the Central

Spanish Pyrenees, Phy. Chem. Earth 20(3-4), 301–308.
Gessler, P.E., McKenzie N.J., Hutchinson, M.F., 1996. Progress in soil-landscape modeling and spatial prediction of soil attributes for environmental models. In Proceedings, Third International Conference/Workshop on Integrating GIS and Environmental Modeling, Santa Fe, NM, January 21-26, 1996. Santa Barbara, CA: National Center for Geographic Information and Analysis. http://www.ncgia.ucsb.edu/conf/SANTA_FE_CD-ROM/main.html.
González, C., García-Ruiz, J.M., Martí, C., White, S., Errea, M.P., Arnáez, J., 1997. Sediment sources in a small, abandoned farmland catchment, Central Spanish Pyrenees, Phy. Earth 22 (3-4), 291–293.
Goovaerts, P., 1994. Study of spatial relationships between two sets of variables using multivariate geostatistics. Geoderma 62, 93–107.
Goovaerts, P., 1999a. Geostatistics in soil science: state-of-the-art and perspectives. Geoderma 89, 1–45.
Goovaerts, P., 1999b. Using elevation to aid the geostatistical mapping of rainfall erosivity. Catena 34(3–4), 227–242.
Goovaerts, P., Viscarra Rossel, R.A., McBratney, A.B., 2002. Geostatistical modeling and propagation of uncertainty: application to the management of agricultural fields. Proceedings of IAMG02, Berlin, September 15-20, 2002.
Gotway, C.A., Ferguson, R.B., Hergert, G.W., Peterson, T.A., 1996. Comparison of kriging and inverse-distance methods for mapping soil parameters. Soil Sci. Soc. Am. J. 60, 1237–1247.
Hang, S., Sereno, R., 2002. Adsorción de atrazina y su relación con las características sedimentológicas y el desarrollo del perfil de dos suelos de la provincia de Córdoba. Revista de Investigaciones Agropecuarias 31 (3), 73–87.
Hengl, T., 2007. A practical guide to geostatistical mapping of environmental variables. Office for Official Publications of the European Communities, Luxembourg, 143 pp, ISBN 1018-5593.
Hengl, T. Heuvelink, G.B.M., Stein, A., 2003. Comparison of kriging with external drift and regression-kriging. Technical note, ITC, available on-line at http://www.itc.nl/library/Academic_output/.
Hengl, T., Heuvelink G. B. M., Stein, A., 2004. A generic framework for spatial prediction of soil variables based on regression- kriging. Geoderma 120, 75–93.
Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, **6**, 65–70.
Herbst, M., Diekkru B., Vereecken, H., 2006. Geostatistical co-regionalization of soil hydraulic properties in a micro-scale catchment using terrain attributes. Geoderma 132, 206–221.
Heuvelink, G.B.M., Dick, J.B., De Gruijter, J.J., 2006. Optimization of sample configurations for digital mapping of soil properties with universal kriging. Dev. Soil Sci. 31, 139–153.
Inakwu, O.A., Crawford, M., McBratney A.B., 2007. Digital mapping of soil attributes for regional and catchment modelling, using ancillary covariates, statistical and geostatistical techniques. In: P. Lagacherie, A.B. McBratney and M. Voltz (Editors), Digital soil mapping. An introductory perspective. Elsevier, Amsterdam, pp. 437–453.
Knotters, M., Brus, D.J., Oude Voshaar, J.H., 1995. A comparison of kriging, co-kriging and kriging combined with regression for spatial interpolation of horizon depth with censored observations. Geoderma 67, 227– 246.
Lark, R.M., 2000. Regression analysis with spatially autocorrelated error: simulation studies and application to mapping of soil organic matter. Int. J. Geogr. Inf. Sci. 14, 247–264.
Lasanta, T., Beguería, S., García-Ruiz, J.M., 2006. Geomorphic and hydrological effects of traditional shifting agriculture in a Mediterranean mountain, Central Spanish Pyrenees, Mt. Res. Dev. 26(2), 146–152.
Laslett, G.M., McBratney, A.B., Pahl, P.J., Hutchinson, M.F., 1987. Comparison of several spatial prediction methods for soil pH. J. Soil Sci. 38, 325–341.

Leenaers, H., Okx, J.P., Burrough, P.A., 1990. Comparison of spatial prediction methods for mapping floodplain soil pollution. Catena 17, 535–550.
Leij, F.J., Romano, N., Palladino, M., Schaap, M.G., Coppola, A., 2004. Topographical attributes to predict soil hydraulic properties along a hillslope transect. Water Resources Research 40: 1–15.
Matheron, G., 1969. *Le krigeage universel*. Vol. 1. Cahiers du Centre de Morphologie Mathematique, École des Mines de Paris, Fontainebleau, p. NA.
Matheron, G., 1982. *Pour une analyse krigeante de données régionalisées*. Centre de Géostatistique, Ecole des Mines de Paris, Report N-732, Fontainebleau.
McBratney, A.B., Mendonça Santos, M. L., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3–52.
McKenzie, N.J., Gallant, J.C., 2007. Digital soil mapping with improved environmental predictors and models of pedogenesis. In: P. Lagacherie, A.B. McBratney and M. Voltz (Editors), Digital soil mapping. An introductory perspective. Elsevier, Amsterdam, pp. 327– 349.
Mckenzie, N.J., Ryan, P.J., 1999. Spatial prediction of soil properties using environmental correlation. Geoderma 89, 67– 94.
Moore, I., Gessler, P., Nielsen, G., Peterson, G., 1993. Soil attribute prediction using terrain analysis. Soil Sci. Soc. Am. J. 57(2), 443–452.
Moore, I.D., Grayson, R.B., Ladson, A.R., 1991. Digital terrain modelling: a review of hydrogical, geomorphological, and biological applications, Hydrological Processes 5(1), 3–30.
Navas, A., Machín, J., Beguería, S., López-Vicente, M., Gaspar, L., 2008. Soil properties and physiographic factors controlling the natural vegetation re-growth in a disturbed catchment of the Central Spanish Pyrenees, Agroforestry Systems 72(3), 173–185.
Navas, A., Machín, J., Soto, J., 2005a. Assessing soil erosion in a Pyrenean mountain catchment using GIS and fallout ^{137}Cs. Agr. Ecosyst. Environ. 105, 493–506.
Navas, A., Soto, J., Machín, J. 2005b. Mobility of natural radionuclides and selected major and trace elements along a soil toposequence in the central Spanish Pyrenees. Soil Sci. 170, 743–757.
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.
Opsomer, J.D., Ruppert, D., Wand, M.P., Holst, U., Hossjer, O., 1999. Kriging with nonparametric variance function estimation. Biometrics 55, 704–710.
Pachepsky, Y.A., Timlin, D.J., Rawls, W.J., 2001. Soil water retention as related to topographic variables. Soil Sci. Soc. Am. J. 65, 1787–1795.
Park, S.J., Vlek, P.L.G., 2002. Environmental correlation of three-dimensional soil spatial variability: a comparison of three adaptive techniques. Geoderma 109, 117–140.
Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. Computers & Geosciences 30, 683–691.
Pinheiro, J.C., Bates, D.M., 2000. Mixed-effects models in S and S-Plus. New York, Springer Verlag.
Quinn, P.F., Beven, K.J., Chevallier, P., Planchon, O., 1991. The prediction of hillslope flow paths for distributed hydrological modelling using digital terrain models, Hydrological Processes 5, 59–79.
Schabenberger, O., Gotway , C. A., 2005. *Statistical methods for spatial data analysis*. Chapman & Hall/CRC.
Selige, T., Böher, 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., Goovaerts, P., Ping, J., Haddix, M.L., 2006. Fine-resolution mapping of soil organic carbon based on multivariate secondary data. Geoderma 132, 471–

489.
Sinowski, W., Auerswald, K., 1999. Using relief parameters in a discriminant analysis to stratify geological areas with different spatial variability of soil properties. Geoderma 89, 113–128.
Soto, J. and Navas, A., 2004. A model of 137Cs activity profile for soil erosion studies in uncultivated soils of Mediterranean environments. Journal of Arid Environments 59, 719–730.
Soto, J. and Navas, A., 2008. A simple model of Cs-137 profile to estimate soil redistribution in cultivated stony soils. Radiation measurements 43, 1285–1293.
Taylor, J.A., Odeh, I.O.A., 2007. Comparing discriminant analysis with binomial logistic regression, regression kriging and multi-indicator kriging for mapping salinity risk in Northwest New South Wales, Australia. In: P. Lagacherie, A.B. McBratney and M. Voltz (Editors), Digital soil mapping. An introductory perspective. Elsevier, Amsterdam, pp. 455–464.
Van Meirvenne, M., Goovaerts, P, 2001. Evaluating the probability of exceeding a site-specific soil cadmium contamination threshold. Geoderma 102, 75–100.
Vauclin, M, Vieira, S.R., Vachaud, G., Nielsen, D.R., 1983. The use of cokriging with limited field soil observations. Soil Sci. Soc. Am. J. 47, 175–184.
Vicente Serrano, S.M., Saz, M.A., Cuadrat, J.M., 2003. Comparative analysis of interpolation methods in the middle Ebro valley (Spain): application to annual precipitation and temperature, Climate Res. 24, 161–180.
Wang, L. & H. Liu (2006): An efficient method for identifying and filling surface depressions in digital elevation models for hydrologic analysis and modelling. International Journal of Geographical Information Science 20(2), 193–213.
Weber, D.D., Englund, E.J., 1992. Evaluation and comparison of spatial interpolators. Math. Geol. 24, 381–391.
Webster, R., 1977. Canonical correlation in pedology: how useful? J. Soil Sci. 28, 196–221.
Webster, R., Burgess, T.M.,1980. Optimal interpolation and isarithmic mapping of soil properties, III. Changing drift and universal kriging. J. Soil Sci. 31, 505–524.
Yaalon, D.H., 1975. Conceptual models in pedogenesis: can soil-forming functions be solved? Geoderma 14, 189–205.
Zhu, J., Morgan, C.L.S., Norman, J.M., Yue, W., Lowery, B., 2004. Combined mapping of soil properties using a multi-scale tree-structured spatial model. Geoderma 118, 321–334.
Zevenbergen, L.W., Thorne, C.R., 1987. Quantitative analysis of land surface topography, Earth Surface Processes and Landforms 12, 47–56.
Ziadat, F.M., 2007. Land suitability classification using different sources of information: Soil maps and predicted soil attributes in Jordan. Geoderma 140, 73–80.

Tables

Table 1. Basic statistics of soil properties (dependent variables, in uppercase) and covariates (in lowercase): minimum and maximum values, median and interquantile range (Q3-Q1). Sample size N = 74.

Variable	Definition and units	Min	Max	Median	Int. range
CLAY	Clay content (%)	12.01	27.23	20.12	4.07
BULK	Bulk density (g cm^{-3})	0.63	1.57	1.16	0.27
CO3	Carbonates content (%)	0.11	51.03	25.62	22.14
PH	pH	6.36	8.55	8.12	2.40
FIELD	Field capacity (% volume)	22.09	46.81	32.01	8.37
MO	Organic matter content (%)	1.53	14.99	5.10	2.29
N	Nitrogen content (%)	0.12	0.66	0.28	0.11
CEC	Cation exchange capacity (meq g^{-1})	153.3	232.6	191.5	20.9
alt	Altitude (m above sea level)	926	1304	1108	136
slope	Slope gradient (m m^{-1})	0.05	0.68	0.33	0.22
asp	Slope aspect, cosine (-)	0.08	4.79	1.74	2.41
wet	Topographic wetness index ()	3.88	9.27	5.90	2.30

Table 2. Analysis of variance, Levene's test for homogeneity of variance and Moran's test for spatial autocorrelation (p-value) of soil properties (dependent variables, in uppercase) and covariates (in lowercase).

Variable	Analysis of variance		Levene's test		Moran's test	
	$F_{3,70}$	p-value	$F_{3,70}$	p-value	I	p-value
CLAY	6.762	<0.001	0.930	0.431	0.033	<0.001
BULK	5.093	0.003	2.689	0.053	0.038	<0.001
CO3	20.051	<0.001	5.602	0.002	0.109	<0.001
PH	16.840	<0.001	9.578	<0.001	0.094	<0.001
FIELD	9.058	<0.001	0.830	0.482	0.053	<0.001
OM	6.533	0.001	2.653	0.055	0.023	0.009
N	4.635	0.005	0.974	0.410	0.020	0.015
CEC	4.360	0.007	1.192	0.319	0.060	<0.001
alt	8.173	<0.001	3.550	0.019	0.210	<0.001
slope	11.02	<0.001	4.310	0.008	0.057	<0.001
asp	22.364	<0.001	3.533	0.019	0.253	<0.001
wet	21.04	<0.001	1.929	0.133	0.073	<0.001

Table 3. Model selection: covariates included in the optimum models for each variable. Interactions between covariates are indicated by ':'.

Variable	Covariates	Value	Std. Error	t-value	p-value
CLAY	(Intercept)	19.6	0.375	52.3	<0.001
	soilHP	4.049	0.932	4.34	<0.001
BULK	(Intercept)	0.996	0.0348	28.6	<0.001
	asp	0.402	0.118	3.40	0.001
	asp:alt	-3.12E-04	1.08E-04	-2.90	0.005
CO3	(Intercept)	21.2	3.117	6.80	<0.001
	soilHP	-25.8	1.880	-13.7	<0.001
	slope	21.4	7.824	2.74	0.008
PH	(Intercept)	8.09	6.20E-02	130.4	<0.001
	slope	2.70	0.959	2.816	<0.001
	slope:alt	-2.32E-03	8.69E-04	-2.669	<0.001
FIELD	(Intercept)	35.3	1.45	24.3	<0.001
	asp	-11.2	3.61	-3.09	0.003
	asp:alt	8.94E-03	3.36E-03	2.66	0.010
OM	(Intercept)	5.69	0.681	8.36	<0.001
	slope	-26.9	10.1	-2.67	0.009
	slope:alt	2.38E-02	9.06E-03	2.63	0.011
N	(Intercept)	0.345	1.79E-02	19.3	<0.001
	asp	-0.164	5.01E-02	-3.26	0.002
	asp:alt	1.30E-04	4.69E-05	2.76	0.007
CEC	(Intercept)	203	3.12	65.0	<0.001
	asp	-41.5	7.38	-5.63	<0.001
	asp:alt	3.35E-02	6.64E-03	5.04	<0.001

Table 4. Residual model: residual standard error, variance structure and spatial correlation structure for each dependent variable.

Variable	Standard error	Variance structure				Correlation structure		
		soilCR	soilRL	soilHK	soilHP	model	range	nugget
CLAY	2.96	–	–	–	–	–	–	–
BULK	0.212	1	0.841	0.731	0.382	spherical	405	8.92E-9
CO3	11.4	1	0.987	0.829	0.298	–	–	–
PH	0.172	1	1.79	0.720	3.90	gaussian	480	0.481
FIELD	5.27	1	1.33	0.720	1.189	spherical	125	1.40E-6
OM	2.02	1	1.98	0.780	1.39	spherical	387	5.46E-8
N	0.0914	1	0.677	1.54	1.05	–	–	–
CEC	13.4	–	–	–	–	–	–	–

Table 5. Cross-validation statistics: leave-one-out BLUPs (estimations considering both the fixed effects and the spatial random effects) were compared with the measured values at left-out locations and used for computing several statistics (mean bias error, mean absolute error, root mean square error).

Variable	R^2	MBE	MAE	RMSE
CLAY	0.169	-3.60E-16	2.40	2.99
BULK	0.212	-0.0177	0.134	0.173
CO3	0.513	-6.80E-03	7.65	9.85
PH	0.145	0.089	0.208	0.361
FIELD	0.197	-0.144	4.37	5.32
OM	0.030	-0.164	1.84	2.49
N	0.118	-1.80E-03	0.0699	0.0953
CEC	0.313	-0.0142	10.5	13.8

Figures

Figure 1. Location of the study area within the Iberian Peninsula, aerial photograph of the catchment and sampling scheme.

Figure 2. Maps of soil types and topographic covariates.

Figure 3. Decision chart for the analysis of spatial variables.

Figure 4. Boxplots of soil properties and topographic covariates according to soil type: a, calcaric Regosols; b, rendsic Leptosols; c, haplic Kastanozems; d, haplic Phaeozems. The horizontal line represents the global mean for each variable. Letters above the box plots indicate significant differences between soil type pairs.

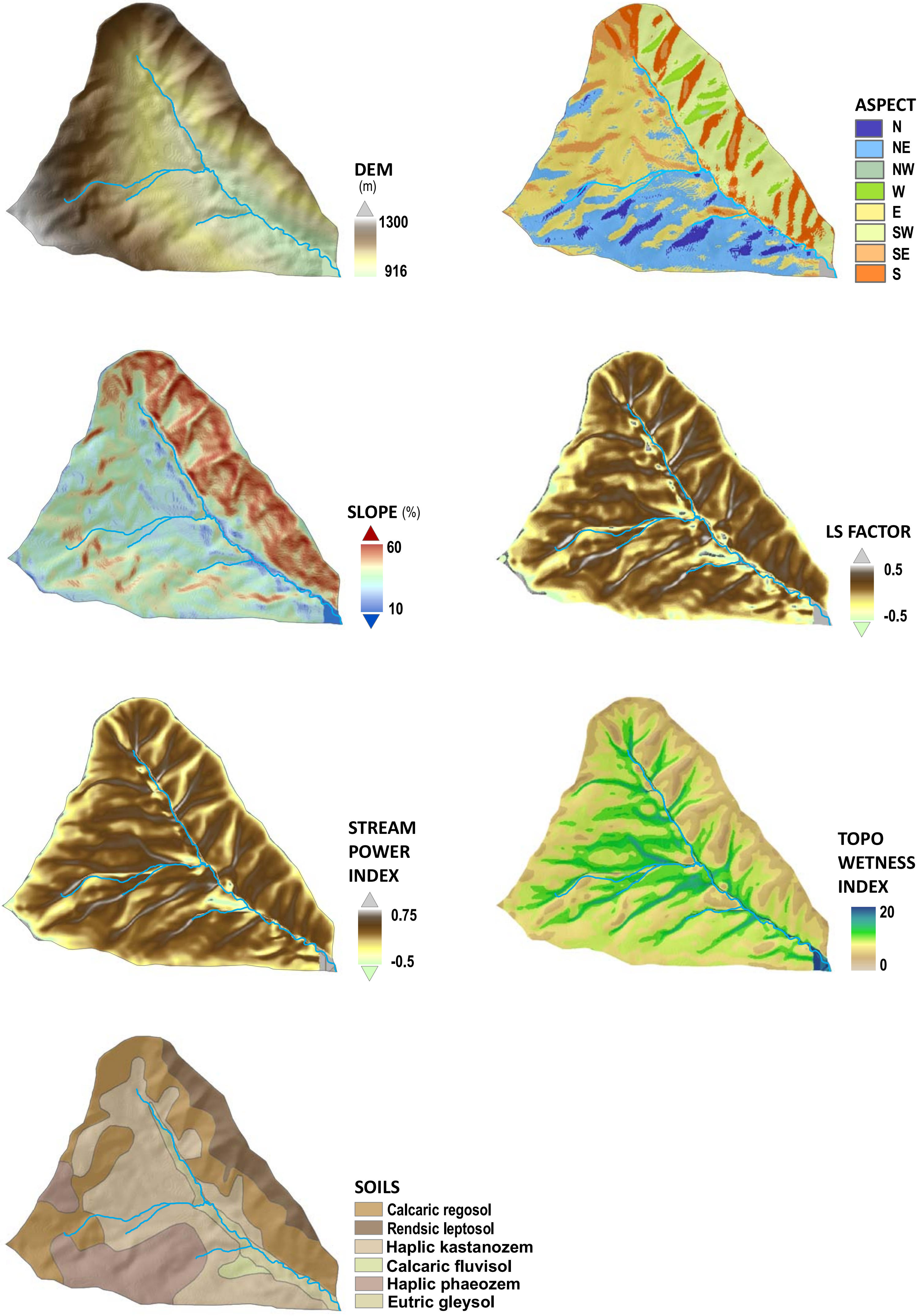
Figure 5. Scatterplot matrix: frequency distributions (diagonal panel), bivariate plots for each variable combination (lower panel, point pairs as circles and loess smoother as a bold line) and Pearson's correlation (upper panel, correlations significant at α=0.05 are marked with an asterisk and a bigger font).

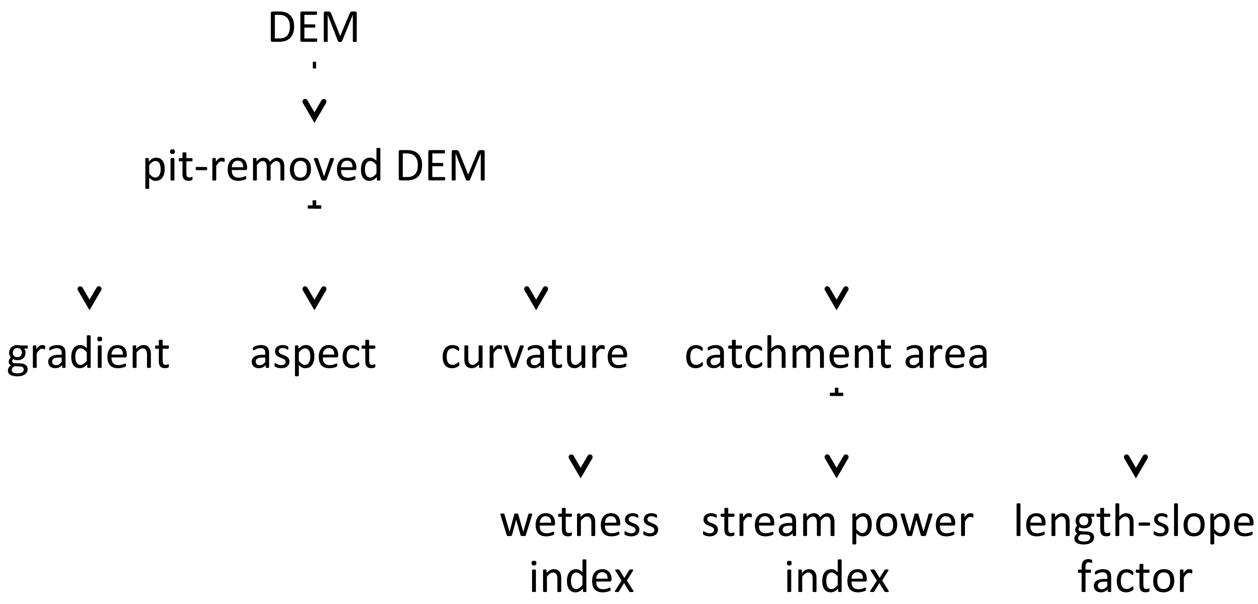
Figure 6. Prediction plots: leave-one-out jacknife predictions against measured values, and line of perfect fit (1:1).

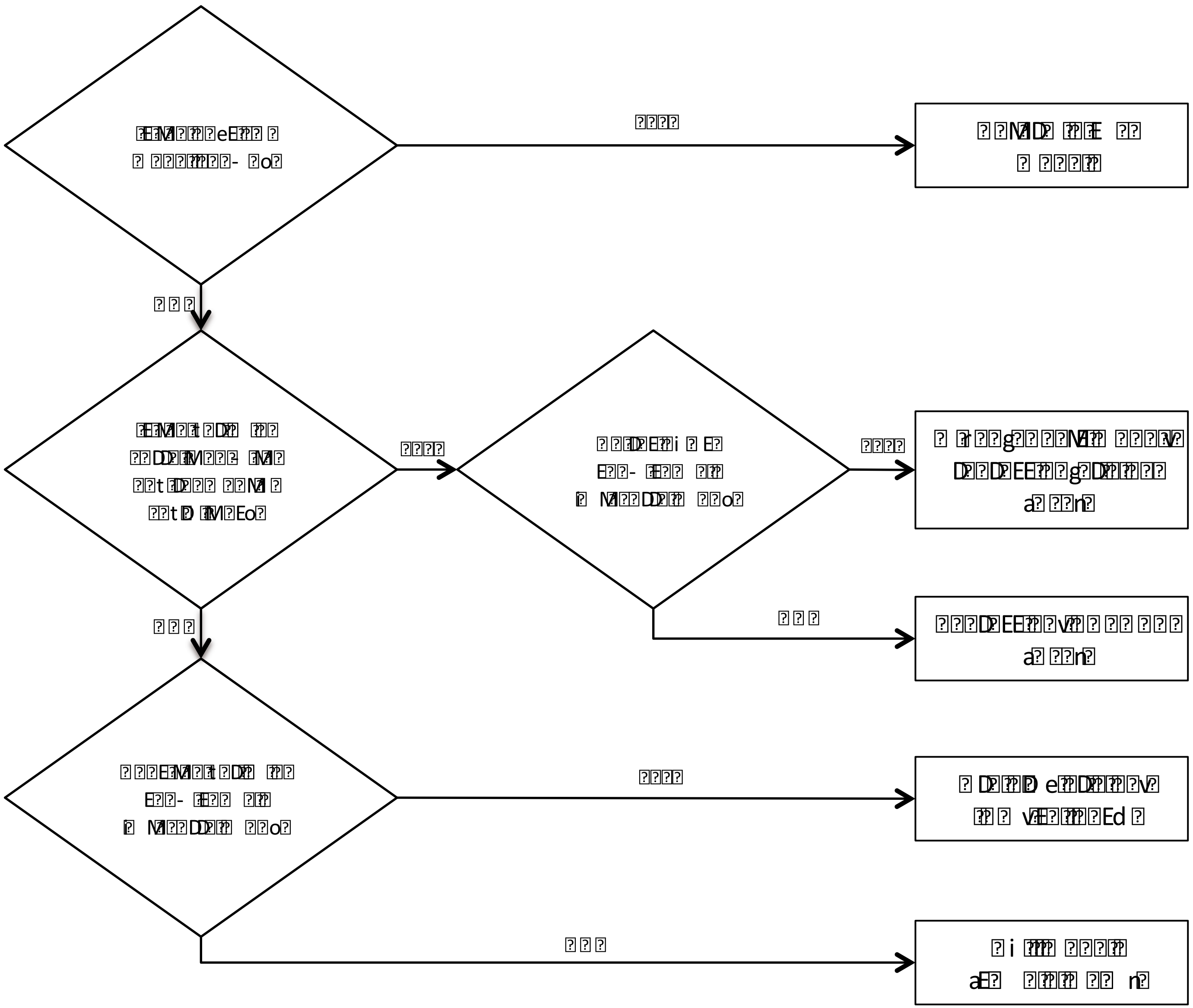
Figure 7. Prediction maps for eight soil properties, based on the optimum fitted models.

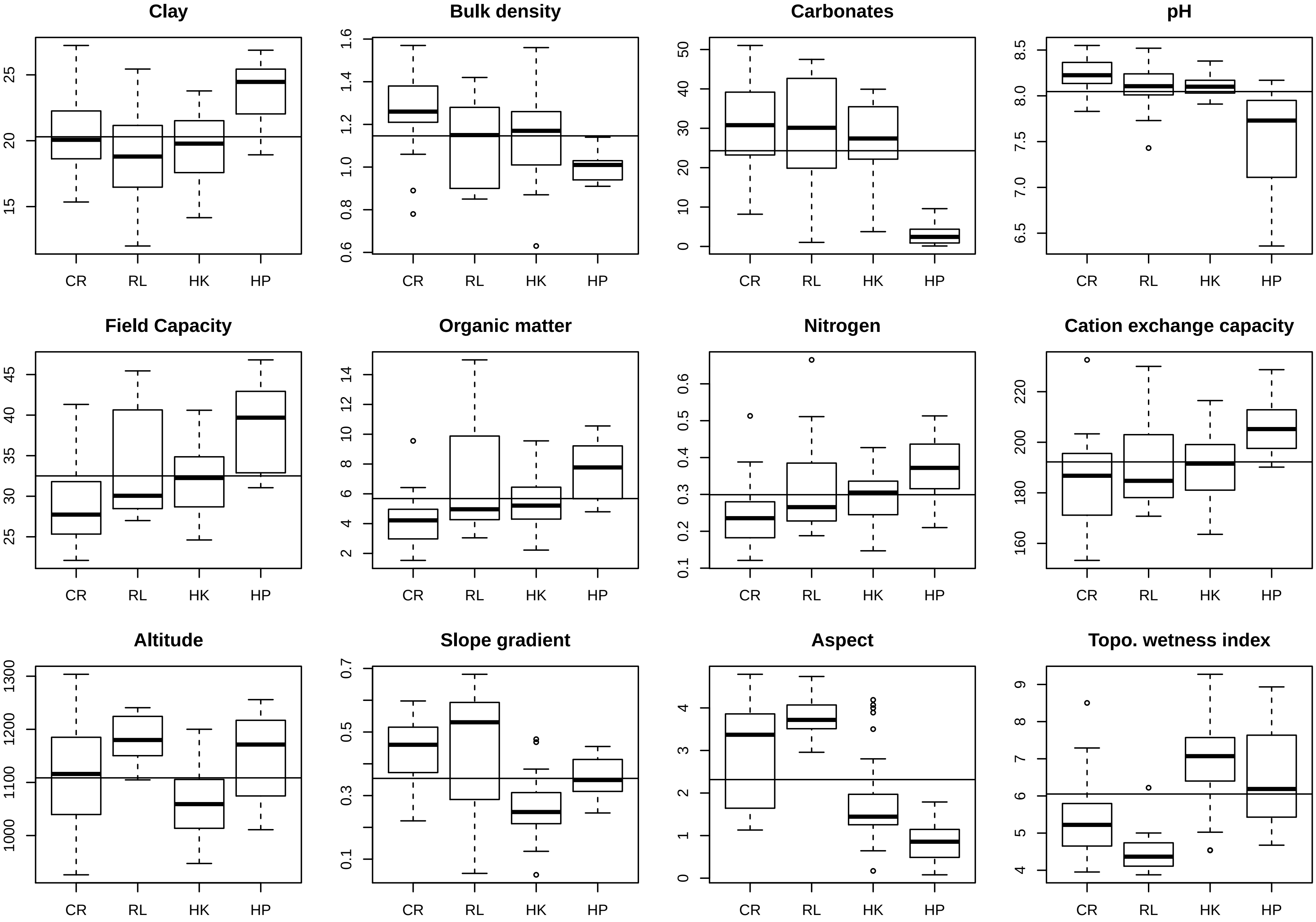
Figure 8. Standard error maps for eight soil properties.


N
42º 38' 20'' N
00º 35' 00'' W
250 125 0 250 500 m.

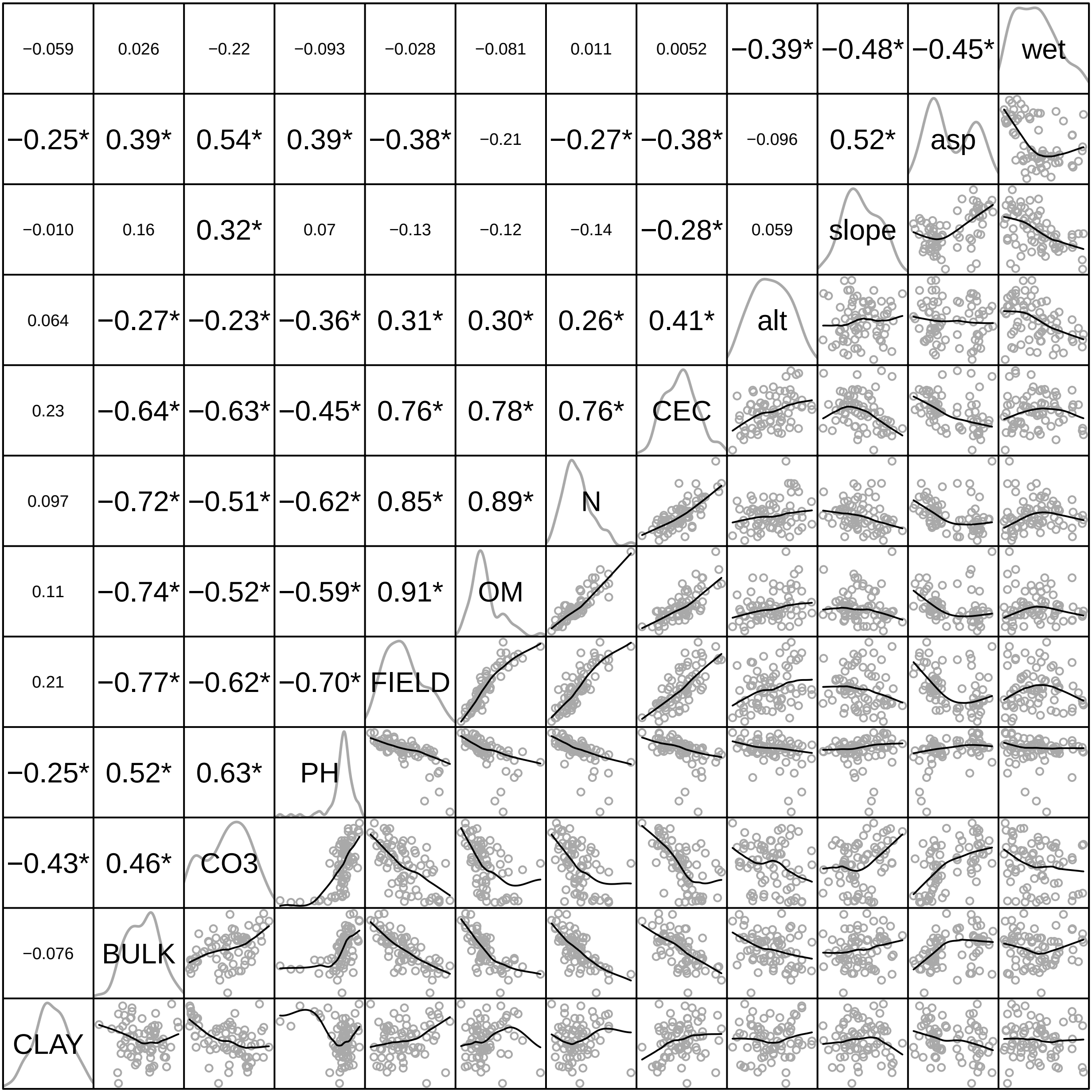

DEM
(m)
1300
916
ASPECT
N
NE
NW
W
E
SW
SE
S
SLOPE (%)
60
10
LS FACTOR
0.5
-0.5
STREAM POWER INDEX
0.75
-0.5
TOPO WETNESS INDEX
20
0
SOILS
Calcaric regosol
Rendsic leptosol
Haplic kastanozem
Calcaric fluvisol
Haplic phaeozem
Eutric gleysol


DEM
pit-removed DEM
gradient
aspect
curvature
catchment area
wetness index
stream power index
length-slope factor

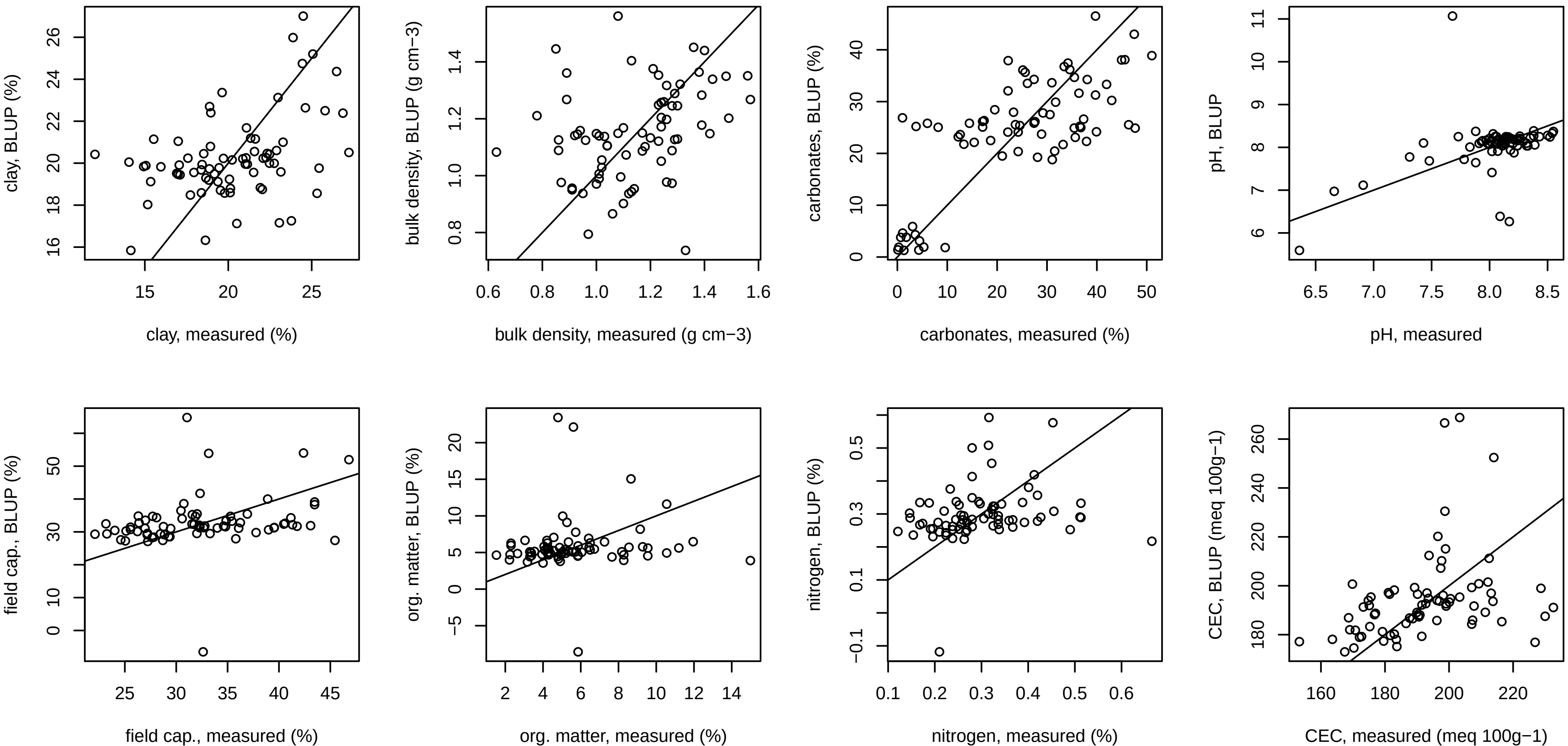


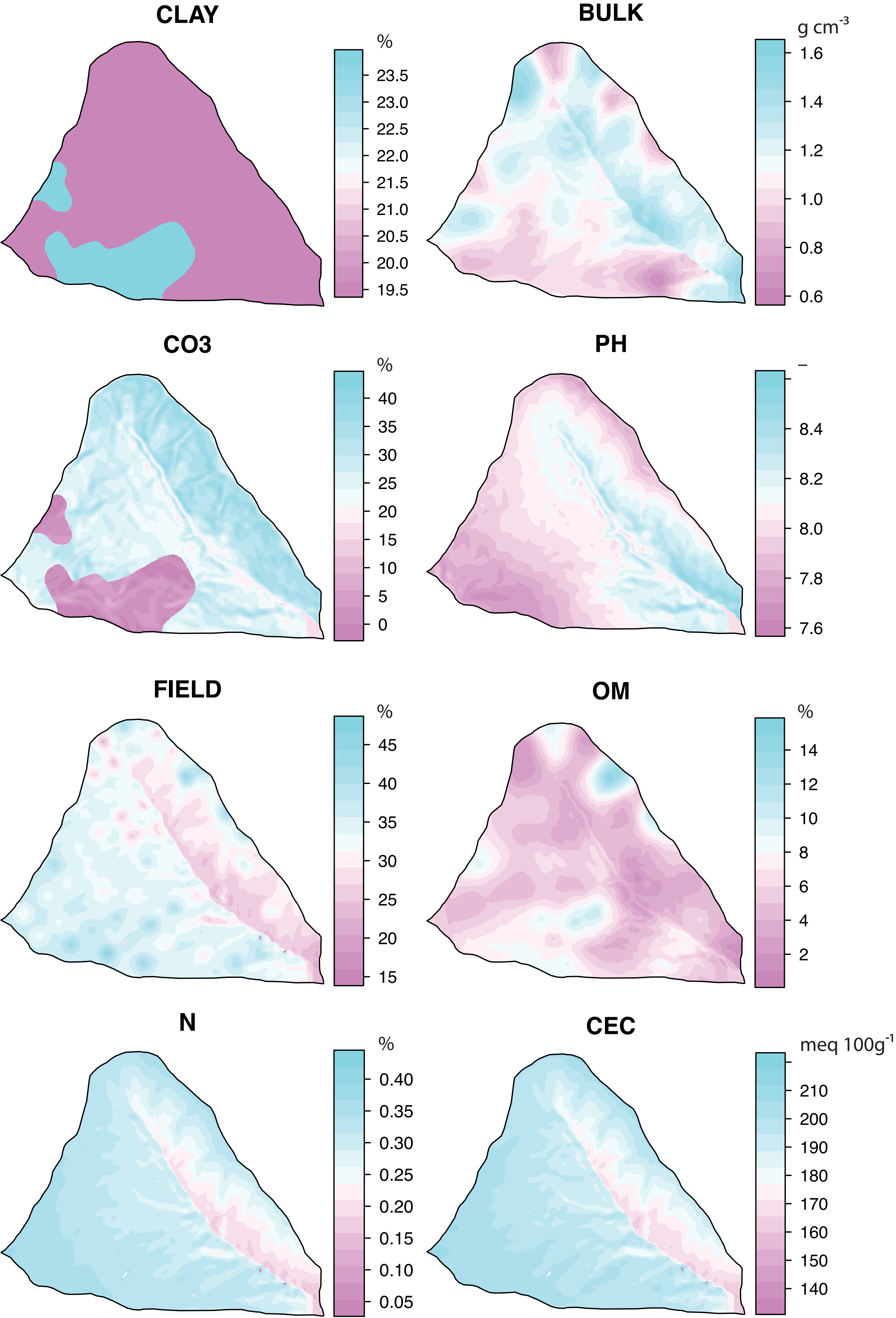


Clay
Bulk density
Carbonates
pH
Field Capacity
Organic matter
Nitrogen
Cation exchange capacity
Altitude
Slope gradient
Aspect
Topo. wetness index
CR
RL
HK
HP

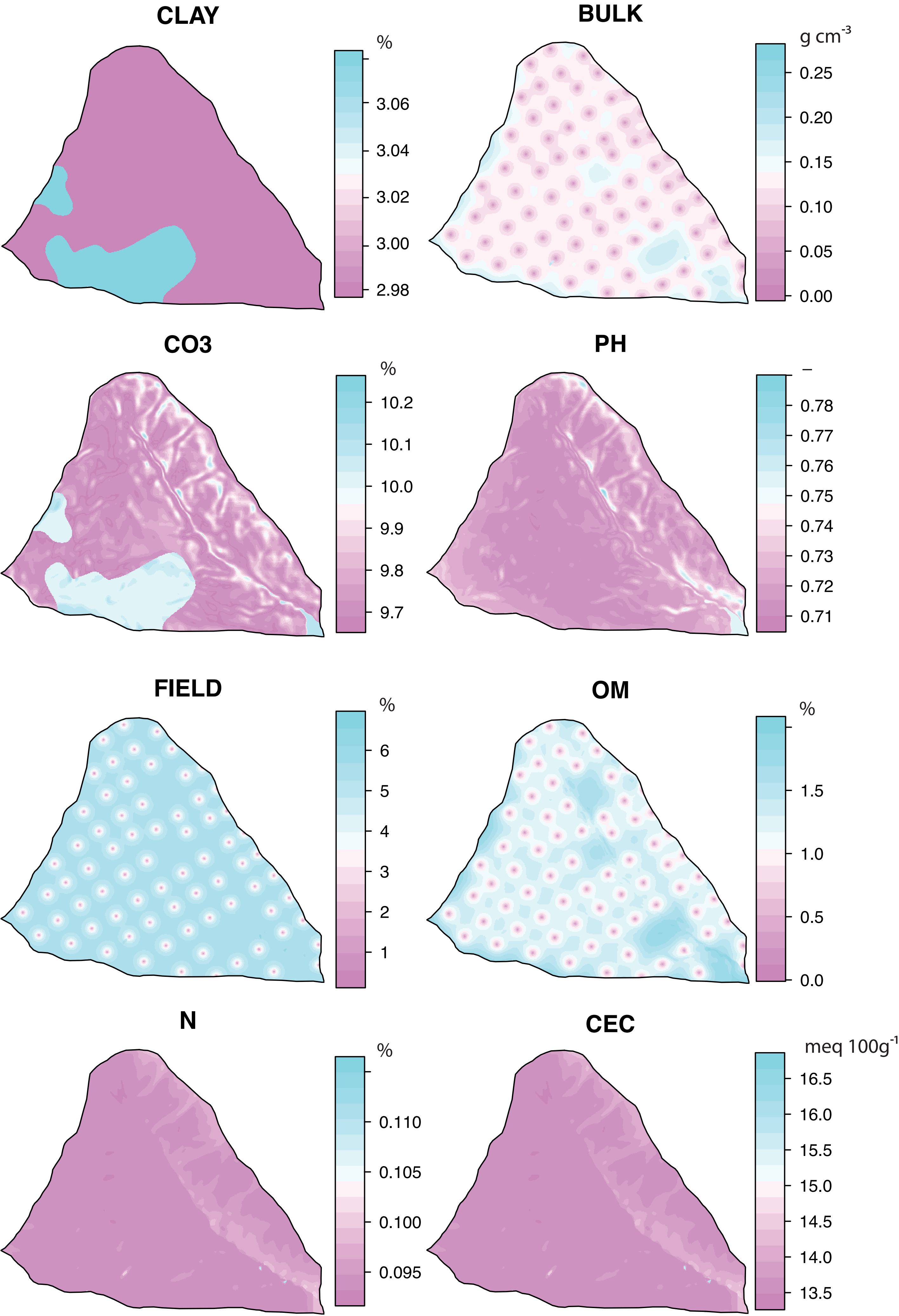


−0.059
0.026
−0.22
−0.093
−0.028
−0.081
0.011
0.0052
−0.39*
−0.48*
−0.45*
wet
−0.25*
0.39*
0.54*
0.39*
−0.38*
−0.21
−0.27*
−0.38*
−0.096
0.52*
asp
−0.010
0.16
0.32*
0.07
−0.13
−0.12
−0.14
−0.28*
0.059
slope
0.064
−0.27*
−0.23*
−0.36*
0.31*
0.30*
0.26*
0.41*
alt
0.23
−0.64*
−0.63*
−0.45*
0.76*
0.78*
0.76*
CEC
0.097
−0.72*
−0.51*
−0.62*
0.85*
0.89*
N
0.11
−0.74*
−0.52*
−0.59*
0.91*
OM
0.21
−0.77*
−0.62*
−0.70*
FIELD
−0.25*
0.52*
0.63*
PH
−0.43*
0.46*
CO3
−0.076
BULK
CLAY


clay, BLUP (%)
clay, measured (%)
bulk density, BLUP (g cm−3)
bulk density, measured (g cm−3)
carbonates, BLUP (%)
carbonates, measured (%)
pH, BLUP
pH, measured
field cap., BLUP (%)
field cap., measured (%)
org. matter, BLUP (%)
org. matter, measured (%)
nitrogen, BLUP (%)
nitrogen, measured (%)
CEC, BLUP (meq 100g−1)
CEC, measured (meq 100g−1)



CLAY
%
23.5
23.0
22.5
22.0
21.5
21.0
20.5
20.0
19.5
BULK
g cm-3
1.6
1.4
1.2
1.0
0.8
0.6
CO3
%
40
35
30
25
20
15
10
5
0
PH
–
8.4
8.2
8.0
7.8
7.6
FIELD
%
45
40
35
30
25
20
15
OM
%
14
12
10
8
6
4
2
N
%
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
CEC
meq 100g-1
210
200
190
180
170
160
150
140


CLAY
%
3.06
3.04
3.02
3.00
2.98
BULK
g cm-3
0.25
0.20
0.15
0.10
0.05
0.00
CO3
%
10.2
10.1
10.0
9.9
9.8
9.7
PH
–
0.78
0.77
0.76
0.75
0.74
0.73
0.72
0.71
FIELD
%
6
5
4
3
2
1
OM
%
1.5
1.0
0.5
0.0
N
%
0.110
0.105
0.100
0.095
CEC
meq 100g-1
16.5
16.0
15.5
15.0
14.5
14.0
13.5