Environmental correlation of three-dimensional soil spatial variability: a comparison of three adaptive techniques

S. J. Park and P. L. G. Vlek
Center for Development Research (ZEF), University of Bonn, Walter-Flex-Strasse 3, 53113, Bonn, Germany
Received 4 September 2001; revised 26 February 2002; accepted 25 April 2002.
Available online 28 May 2002.

Abstract

An appropriate inclusion of spatial variation of soils is becoming increasingly important for spatially distributed ecological modelling approaches. Even though soils are anisotropic vertically and laterally, most soil spatial variability studies have focused on the lateral variation of soil attributes over the landscape. This study characterizes the complexity of three-dimensional variations of individual soil attributes and examines the possibility of predicting soil property distribution using three different regression approaches: artificial neural networks (ANN), regression trees (RT) and general linear models (GLM). Thirty-two physiochemical attributes of 502 soil samples were collected from 64 soil profiles on a slope at Bicknoller Combe, Somerset, UK. After a principal component analysis, five soil attributes were selected to test for environmental correlation, assuming they reflect dominant pedological processes at the hillslope. Vegetation occurrence, soil types, terrain parameters and soil sample depth were used as predictors. Prediction using environmental variables was most successful for soil attributes whose spatial distribution is strongly influenced by lateral hydrological and slope processes with relatively simple depth functions (e.g. total exchangeable bases, Mn oxides and soil pH). These soil attributes also showed a high mobility, which implies that their spatial distribution quickly reaches an equilibrium with current slope processes. Soil taxonomic information only marginally improved the performance of models constructed from surface information such as vegetation and terrain parameters. On the other hand, soil attributes whose vertical distribution is strongly governed by vertical pedogenesis or unknown factors were poorly modelled by environmental variables due to stronger nonlinearity in their vertical distribution. Soil taxonomic information becomes more important for predicting these soil attributes. As an empirical modelling tool, GLM with interaction terms outperformed the other two methods tested, ANN and RT, in terms of both the simplicity of the model structure and the performance of derived empirical functions.

Author Keywords: Artificial neural networks; Regression trees; Generalised linear model; Soil spatial variability; Soil–landscape analysis

Article Outline

1. Introduction

An appropriate understanding and inclusion of spatial variation of soils is essential for ecological and environmental process modelling on the landscape scale. The lack of soil information at a detailed spatial resolution greatly increases the uncertainty of model outputs and is also one of the main limitations for further development of spatially distributed models (Jury; Beven; Moore; Burrough and McSweeney). During the last decade, many studies attempted to characterize and predict the spatial distribution of soils using more readily available environmental variables, a technique now called soil–landscape analyses (Hewitt, 1993) or environmental correlation (McBratney et al., 2000). In more recent years, there has been increasing scientific interest in how to combine the results of soil–landscape analyses and spatially distributed models (Heuvelink and Zhu). The underlying motivation is to improve model outputs and reduce the time and cost of collecting information on the spatial heterogeneity of soils by developing a framework to identify the spatial distribution of soil attributes over the landscape (Heuvelink and McBratney).

Great advances have been made in both the theoretical and methodological aspects of soil–landscape analyses. McBratney et al. (2000) categorised previous soil–landscape analyses into three approaches: (1) to find the spatial association of individual soil attributes with various environmental factors with the aim of identifying the driving forces for spatial distribution of soil attributes; (2) to characterise the statistical structures of soil attributes by means of geostatistical interpolation of stochastic components of soil distribution; and (3) hybrid approaches combining these two approaches. One should, however, note that most of those studies were limited to topsoil or to one arbitrarily chosen soil depth, which will be referred to in this paper as two-dimensional (x, y coordination on the surface) soil–landscape analyses.

Soils are strongly anisotropic vertically and laterally (Beckett and McSweeney). Vertical anisotropy of soil attributes is frequently much stronger than lateral anisotropy (Wilding, 1984). One of the unique characteristics of soil is its vertical eluviation–illuviation relationship. Soil constituents transported downward and materials leached out by infiltrated water are deposited in subsurface horizons forming illuvial B horizons. This becomes more complex when considering the fact that individual soil attributes show totally different eluvial–illuvial relationships within the same profile (Park and Burt, in press). Furthermore, the effect of environmental factors on the vertical distribution of soil attributes may not be linear in nature. Even though most previous soil–landscape analyses assumed a linearity between soil attributes and environmental factors (Odeh et al., 1994), it is not clear whether we are able to extend that assumption to the three-dimensional distribution of soil attributes. Considering the fact that almost all earth surface processes occur on, in and through the soil body, two-dimensional soil landscape analyses ignore 'in' and 'through' components of soil processes by concentrating on the 'on' components of soil spatial variations. Soil research should pay more attention to the three-dimensional arrangement of soil variation since most current deterministic models require vertical arrangement of soil information (horizonation) as parameters.

Massive sampling for accurate spatial interpretation is possibly one of the main practical constraints discouraging the examination of three-dimensional variability. The unavailability of statistical tools may be another limitation. In soil–landscape analyses, most frequently used statistical techniques are linear regressions and their derivatives (McKenzie; Gessler and McBratney) and geostatistical methods (summarised by McBratney et al., 2000). In soil science, geostatistical methods have been used for the investigation of spatial dependency of soil samples collected in a two-dimensional space. The application of many linear regression models may create problems for three-dimensional application, because the independence among measured soil attributes is inevitably violated when soil samples from different soil depths are taken at the same profile location.

Soil distribution at the catena and landscape scale is the least quantified and understood ecosystem (Hoosbeek and Bryant, 1992). Current soil–landscape studies are predominantly empirical, based on an assumption that a constructed model in a representative area will be applicable to unsurveyed areas where similar environmental conditions occur (Gessler and Lagacherie). Such empirical models might be the easiest available methodology for the prediction of three-dimensional soil variations. Some adaptive models have been recently introduced in environmental science for prediction purposes. Artificial neural networks (ANN) and regression trees (RT) are the most widely used numerical techniques. An important characteristic of these two techniques is their adaptive nature with regard to learning by examples to solve problems. They also allow illustration of complex and nonlinear relationships without rigorous assumption of the distribution of samples (Bishop and Breiman). These methods are gaining popularity for research areas where there is little or incomplete understanding of the problem to be solved, but where training data is available.

This study investigated the possibility of predicting soil attributes in a three-dimensional space by applying three different adaptive techniques: general linear model (GLM), artificial neural networks (ANN) and regression trees (RT). GLM has previously been used in soil–landscape analyses by McKenzie and Austin (1993), Gessler et al. (1995) and McBratney et al. (2000). McBratney et al. (2000) applied artificial neural networks (ANN) to predict soil distribution and conclude that ANN outperformed many other linear statistical methods. Zhu (2000) used ANN techniques to aid soil-mapping procedures at the regional scale. McKenzie and Ryan (1999) applied RT to elucidate soil and environmental relationships in Australia. McBratney et al. (2000) also applied RT and compared the results with other environmental correlation techniques. They conclude that RT shows the poorest performance among many other techniques compared in their research, mainly due to the discrete partitioning of data. All these applications were limited to two-dimensional environmental correlation of soil attributes at a single soil depth or to soil type mapping.

It is widely believed that each soil attribute has a unique spatial distribution over the landscape due to differences in geochemical and geomorphological mobility, and also to differential responses to given pedological and ecological processes (Huggett and Park). Differences in the spatial distribution of individual soil attributes was considered throughout this study. Specific research questions were (1) to identify dominant pedogeomorphological processes that result in differences in spatial distribution of individual soil attributes; (2) to investigate potentials of GLM, ANN and RT as empirical modelling tools to predict three-dimensional variation of soil attributes over the landscape; and (3) to identify the most useful environmental variables to predict soil attributes.

2. Study area and field methods

2.1. Study area

The study area is a south-facing slope (approximately 100×300 m) at Bicknoller Combe on the Quantock Hills in Somerset, UK. Fig. 1 shows the spatial distribution of vegetation, soil type and slope configuration. The slope profile shows the progression: flat interfluve→convex slope→steep straight slope→(weak) concave foot slope. A hollow extends upslope from the base of the slope, almost reaching the flat interfluve. The three-dimensional arrangement of the hollow and spurs governs subsurface drainage (Anderson and Burt, 1978) and, in turn, strongly influences catenary soil development (Park et al., 1996) and slope denudation (Park and Burt, 2000). Estimated average annual rainfall is 1030 mm and mean air temperature is 9.2 °C.

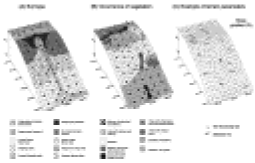 Display Full Size version of this image (58K)

Fig. 1. Environmental variables used in this research in Bicknoller Combe, Somerset, UK (3°15′26″W, 51°09′02″N). Soil type and vegetation species were recorded as presence (1) and absence (0). Only the slope gradient is presented out of eight terrain parameters used in this research.

The slope is used for sheep grazing. Bent-fescue grass (*Agrostis-Festuca*) and bracken (*Pteridium aquilinum*) are the dominant vegetation, with local occurrence of gorse (*Ulex europaeus*), cross-leaved heath (*Erica tetralix*), bristle bent (*Agrostis setacea*), bilberry (*Vaccinium myrtillus*) and sedges (*Carex* spp.) (see Fig. 1). The slope is covered by the deep (>2 m) regolith derived from the Hangman Grits of the Middle (or Late) Devonian, which consist of fine- to medium-grained thickly bedded or massive sandstones. Park et al. (1996) identified a well-developed podzolic catena on the slope (see Fig. 1). Stagnoluvic gley soils have developed on the flat interfluve, and this soil zone is succeeded by stagnogleyic podzolic brown soils on the upper convex shoulder slopes. On the steep main slope, podzols upslope connect to orthic brown soils downslope via podzolic brown soils. Stagno-gley podzols have developed along the convergence line of upslope hollow and upslope flanks, and merge into a type of podzolic soil where sesquioxides and fine soil materials accumulate in surface horizons. This soil zone merges into the seepage soils of the lower hollow. This catenary sequence is a result of active throughflow processes that are in turn governed by the slope configuration at the study sites (Park et al., 1996).

2.2. Field sampling, laboratory analyses and environmental variable calculation

Fifty-four soil pits were dug on a 25-m square sampling grid (Fig. 1). A further 10 soil profiles were investigated in the middle of the grid along the flank slopes where the soil changes rapidly. Pits ranged in depth from 40 to 110 cm depending on the stoniness of the subsoil. Approximately 2 kg of soil were collected at each 10 cm sampling interval. A total of 502 samples were carefully ground to pass through a 2-mm sieve and analysed. Table 1 lists the 32 soil attributes measured.

Table 1. The descriptive statistics of soil attributes analysed for this study

Group	Variable[a]	Unit	Mean	CV[b] (%)	PC 1	PC 2	PC 3	PC 4	PC 5	Communality
PC 1	TO_d	mg kg^{-1}	17623.89	27.01 ∗ ∗	0.94	−0.19	−0.22	−0.06	0.14	0.897
	OXIDE	%	1.84	27.15∗	0.94	−0.16	−0.20	−0.07	0.16	0.896
	Al_d	mg kg^{-1}	2739.21	10.78 ∗ ∗	0.94	−0.28	−0.05	−0.27	0.11	0.924
	TO_o	mg kg^{-1}	5434.67	10.54∗	0.93	−0.19	−0.27	−0.39	0.15	0.942
	Al_o	mg kg^{-1}	2631.04	12.02 ∗ ∗	0.91	−0.35	−0.09	−0.28	0.09	0.899
	Fe_d	mg kg^{-1}	14301.61	25.88∗	0.88	−0.21	−0.16	0.02	0.16	0.823
	Fe_o	mg kg^{-1}	2383.11	42.18∗	0.83	−0.09	−0.12	−0.50	0.32	0.846
	Si_o	mg kg^{-1}	160.45	21.05 ∗ ∗	0.73	−0.05	−0.31	−0.15	−0.13	0.606
PC 2	TEB	mg kg^{-1}	7.96	52.39 ∗ ∗	−0.22	0.98	−0.18	−0.24	0.27	0.978
	Ca_e	mg kg^{-1}	87.02	29.35 ∗ ∗	−0.21	0.96	−0.20	−0.23	0.23	0.932
	Mg_e	mg kg^{-1}	23.54	42.63 ∗ ∗	−0.27	0.94	−0.23	−0.16	0.24	0.931
	CMS	%	21.48	35.12 ∗ ∗	−0.31	0.90	−0.33	−0.04	−0.21	0.947
	BS	$mmol_c$/100 g	19.01	36.16 ∗ ∗	−0.31	0.90	−0.32	−0.04	−0.25	0.969
	AS	%	79.62	24.46∗	0.27	−0.87	0.32	0.11	0.18	0.874
	Na_e	mg kg^{-1}	8.78	25.59 ∗ ∗	0.11	0.79	−0.07	−0.17	0.31	0.745
	K_e	mg kg^{-1}	50.70	22.65 ∗ ∗	−0.20	0.72	0.02	−0.34	0.37	0.639
PC 3	Mn_d	mg kg^{-1}	676.59	38.39	0.37	0.19	−0.93	−0.09	−0.11	0.922
	Mn_d/Fe_d	%	4.61	61.32	0.10	0.27	−0.93	−0.10	−0.17	0.896
	Mn_o	mg kg^{-1}	420.59	59.73	0.33	0.18	−0.93	−0.12	−0.16	0.900
	pHc	−log $[H^+]$	3.94	9.78∗	0.53	−0.01	−0.69	0.09	−0.40	0.733
	pHh	−log $[H^+]$	3.23	8.11 ∗ ∗∗	0.21	0.16	−0.61	0.05	−0.29	0.422
	Fe_e	mg kg^{-1}	5.03	26.62	−0.17	0.03	0.60	−0.11	0.41	0.446
	Mn_e	mg kg^{-1}	8.21	23.68	−0.14	0.58	−0.59	−0.22	0.19	0.701
PC 4	SILT	%	45.18	16.83	0.12	0.19	−0.07	−0.92	0.18	0.854
	SAND	%	44.67	10.39∗	−0.25	−0.22	0.13	0.91	−0.25	0.865
	CLAY	%	8.30	13.20∗	0.20	0.32	−0.19	−0.69	0.47	0.651
	Fe_o/Fe_o	%	16.39	30.46∗	0.36	0.02	−0.05	−0.66	0.27	0.520
	FF	%	34.80	35.19	−0.37	0.08	0.38	−0.41	0.18	0.443
PC 5	ECEC	$mmol_c$/100 g	39.31	14.67	0.16	0.18	0.23	−0.36	0.93	0.879
	ETC	mg kg^{-1}	461.17	29.97	0.04	0.49	0.13	−0.41	0.86	0.895
	Al_e	mg kg^{-1}	277.88	32.34∗	0.26	−0.42	0.42	−0.23	0.78	0.915
	LOI	%	5.51	32.46∗	0.18	0.43	0.44	−0.48	0.56	0.782
Eigenvalue					8.94	7.73	5.45	2.05	1.48	
Variance explained					27.9	24.1	17.0	6.4	4.6	
Cumulative percentage					27.9	52.1	69.1	75.5	80.1	

Correlation matrix among soil attributes and five principal components (PC) were estimated after a direct oblimin rotation of the principal components.

Soil types, vegetation and terrain parameters were used as independent environmental variables. Eight different soil types were surveyed at the study slope (Fig. 1A). Eleven plant species showing a strong catenary distribution over the slope were mapped over the soil survey site (Fig. 1). A laser distance meter (LDM) and differential global positioning system (DGPS) were used to construct a 10-m DEM of the slope. Eight terrain attributes were calculated: elevation, slope gradient, aspect, plan curvature, profile curvature, upslope area, curvature and wetness index. The primary topographic indices—slope gradient, aspect, plan curvature, profile curvature, curvature and upslope area—were calculated using the grid-based algorithm developed by Zevenbergen and Thorne (1987), while the wetness index was calculated using the methods given in Moore et al. (1993). The definition of each terrain parameter and its physical meaning in terms of slope and pedogenic processes are described in Zevenbergen and Thorne (1987) and Moore et al. (1993).

3. Statistical models

3.1. Identification of dominant pedogeomorphological processes

In order to identify dominant pedogeomorphological processes, a principal component analysis (PCA) was first applied using all soil attributes measured. Principle component analysis is frequently used as a data reduction technique, but also as an indirect ordination technique in ecological research. The assumption for PCA as an ordination technique is that when a broad enough suite of soil attributes is included in PCA analysis, the pedological and geomorphological processes that determine the spatial distribution of soil properties may be revealed (Jongman et al., 1995). More detailed statistical procedures for the ordination technique and the descriptions of the spatial characteristics of each component and soil attributes are available in Park and Burt (in press).

Five principal components explaining 80% of the total variation were interpreted as the effects of 'podzolisation', 'nutrient dynamics', 'solute leaching', 'weathering-erosion' and 'soil acidification' after examining spatial distribution and correlation with terrain parameters (Table 1). Park and Burt (in press) constructed separate regression models at different soil depths for each individual component in order to characterise the their spatial patterns. Soil attributes belonging to each principle component show similar spatial distribution at the study slope. Fig. 2 and Fig. 3 show two examples: citrate–bicarbonate–dithionite (CBD) extractable Mn and silt content, which show a contrasting spatial distribution.

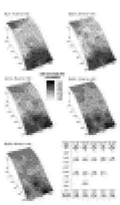 Display Full Size version of this image (53K)

Fig. 2. Spatial distribution of CBD extractable Mn over the study slope at five different soil depths. The table in this figure shows the stepwise multiple regression function for each soil layer. The values in parentheses are the change of R^2 for each independent variable for the stepwise selection procedure.

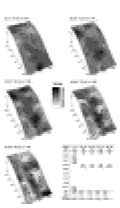 Display Full Size version of this image (55K)

Fig. 3. Spatial distribution of silt content (%) over the study slope at five different soil depths. See Fig. 2 for details of the table in this figure.

The five soil attributes showing the highest component score for each principal component and soil pH were subjected to environmental correlation analyses. These soil attributes include total exchangeable bases (TEB) for 'nutrient dynamics effect', CBD extractable Mn (DMN) for 'solute leaching effect' (Fig. 2), effective cation exchange capacity (ECEC) for 'soil acidification effect', silt content (SILT) for 'weathering–erosion effect' (Fig. 3) and CBD extractable oxides for 'podzolisation effect'. Using the soil attributes as proxies would yield results similar to those using the component score, since the selected soil attributes have the highest component score within each principal component (Table 1). Soil pH is included because of its importance in general soil-forming processes and also for comparison with other soil properties, especially with DMN, which belongs to the same group 'solute effect'.

3.2. Environmental correlation models

The three adaptive techniques, GLM, ANN and RT, have different theoretical assumption and development pathways. Detailed reviews of these techniques are beyond the scope of this paper, and this section gives only a brief summary and procedure of each technique. One common advantage of these modelling techniques is their ability to include both categorical and continuous variables as predictors. Because the predictor variables in this research are a mixture of categorical (soil type and vegetation) and continuous variables (terrain parameters), the multivariate normality assumption is not upheld. This problem is solved in the proposed adaptive methods via either specific data coding (in the case of GLM) or unique data fitting algorithms (in the cases of RT and ANN). The latter two models are able to model linear as well as nonlinear relationships between dependant variables and predictors and, thus, can be applied to complex soil and landscape relationships.

3.2.1. Artificial neural networks (ANN)

The artificial neural network method originated in artificial intelligence research that attempts to mimic the capacity to learn through biological neural systems. Many different types of neural nets are available, whose structure is described in Bishop (1995), Ripley (1996) and Principe et al. (2000). The ANN structure used in this paper is a multilayer perceptron (MLP), which is the most popular neural network structure in ecological modelling and soil science (Schultz and Dawson). In order to introduce nonlinearity during the ANN training, the hyperbolic tangent function, which is again the most popular activation function among many others, is used. MLP also requires determination of the number of hidden layers and numbers of neurons. The difference in some internal parameters of MLP showed relatively small changes in the overall NMSE and *r* (Table 2). Even though the best performance (lowest NMSE and highest *r*) among compared models was achieved by a MLP with two hidden layers and normalized input values, one hidden layer network without normalization was chosen for this research. This decision was made in order to reduce the complexity of the models and the difficulty in interpreting the results in the normalized variables. There is no theory to date for determining the optimum numbers of hidden layers to approximate any given function (Bishop and Maier), and the results between one and two hidden layers are quite comparable (Table 2). The optimum number of hidden nodes in the one hidden layer MLP was estimated by a sensitivity analysis of the relations between the numbers of hidden nodes and number of soil samples (Fig. 4). A simple generalization was impossible from the sensitivity analysis, but 27 hidden nodes from 23 input nodes (environmental variables) and six output nodes (soil attribute used in this research) frequently showed the best network performance.

Table 2. Comparison of different internal parameters for MLP used in this research

Hidden layers	Normalized	Activation function	No. of PE	Normalised mean square error					
				TEB	DMN	ECEC	SILT	CBD	PHC
1	No	Tan*h*	27	0.23	0.33	0.48	0.65	0.54	0.19
1	No	Tan*h*	20	0.28	0.29	0.54	0.58	0.51	0.19
1	No	Tan*h*	15	0.27	0.27	0.53	0.62	0.52	0.18
1	No	Tan*h*	10	0.28	0.29	0.63	0.71	0.51	0.19
1	No	Tan*h*	5	0.40	0.44	0.58	0.87	0.57	0.23
1	No	Sigmoid	27	0.33	0.44	0.71	0.76	0.70	0.30
2	No	Tan*h*	27	0.29	0.35	0.54	0.58	0.57	0.20
2	No	Sigmoid	27	0.69	0.53	0.87	0.92	0.93	0.66
3	No	Tan*h*	27	0.30	0.28	0.61	0.60	0.47	0.22
4	No	Tan*h*	27	0.34	0.50	0.62	0.74	0.53	0.21
1	Yes	Tan*h*	27	0.28	0.26	0.51	0.61	0.48	0.19
1	Yes	Tan*h*	20	0.27	0.30	0.50	0.55	0.47	0.20
1	Yes	Tan*h*	15	0.27	0.31	0.54	0.70	0.54	0.17
1	Yes	Tan*h*	10	0.30	0.28	0.61	0.70	0.55	0.17
1	Yes	Tan*h*	5	0.28	0.39	0.69	0.78	0.53	0.26
1	Yes	Sigmoid	27	0.31	0.43	0.70	0.79	0.68	0.30
2	Yes	Tan*h*	27	0.22	0.31	0.53	0.56	0.52	0.19
2	Yes	Sigmoid	27	0.74	0.53	0.84	0.90	0.92	0.81
3	Yes	Tan*h*	27	0.23	0.33	0.65	0.52	0.48	0.17
4	Yes	Tan*h*	27	0.42	0.42	0.81	0.68	0.49	0.41

For all networks, 23 input nodes and six output nodes were used. The used learning rule is momentum with its value 0.7. Maximum epoch was set at 1000, but an early stop of learning was used for each training by a cross-validation procedure using 10% of total training data set.

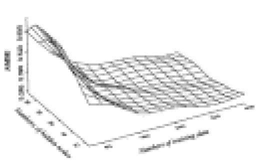
Display Full Size version of this image (9K)

Fig. 4. An example of error surface of CBD extractable Mn oxides in the MLP. NMSE is plotted as a function of sample numbers and numbers of hidden nodes in one hidden layer MLP. Hidden node 27 is used because NMSE is relatively small despite different numbers of soil samples.

For training neural networks, a cross-validation procedure is a highly recommended method to stop network training for ANN to avoid overfitting. Cross-validation during the ANN was not performed in the case of the models tested against the testing data set, because pilot analyses showed that overfitting is relatively minor, and comparison with other techniques is the first priority in this study. Early stopping based on cross-validation, however, was used in the case of the total soil sample models (see Fig. 5).

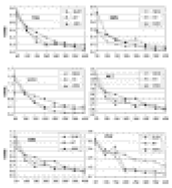 Display Full Size version of this image (28K)

Fig. 5. The comparison of goodness-of-fit of three regression techniques using different training data sets. The value is averaged NMSE after 10 runs of each model after randomisation.

3.2.2. General linear model (GLM)

The general linear model (GLM) may be considered as a special form of ANN that does not have any hidden layers generating nonlinearity during the training (McCullagh and Nelder, 1989). GLM differs from the well-known multiple regressions in two main respects. First, the distribution of the dependent or response variable does not have to be continuous. GLM allows categorical or nominal variables as predictors by recoding them into a number of dichotomous variables. Second, unlike multiple regressions, which are intrinsically univariate methods, GLM allows linear combinations of multiple dependent variables. This is a great advantage for this study, because it can take into account not only the relationships of the predictor variables with the dependent variables, but also the relationships among the multiple dependent variables. Since we could anticipate that the environmental variables are highly correlated, the addition of interaction terms may give insight into which aspects of the response variables are, and are not, related to the predictor variables. On the other hand, it may be a disadvantage because an identification of the 'best set' of independent variables may be less meaningful due to the possible increase of multicollinearity.

3.2.3. Regression trees (RT)

A regression tree is a nonparametric type of regression model that successively splits the response value using the predictor variables until the error reaches a predefined criterion. The continuous splitting of data is called the binary recursive partitioning (RP) algorithm (Breiman et al., 1984). Tree-based models also require a selection of internal parameters, particularly for the number of terminal nodes. If a large overgrown tree model is used, the model is frequently overfitted to the training data; it provides a close approximation of the collected samples but a biased description of the sampled population. The trees grown with the RT algorithm are usually postpruned to ensure a better compromise between comprehensibility and predictive accuracy. While there are some rules on the decision on the numbers of terminal nodes, a trial-and-error approach using a cross-validation procedure is widely used (Breiman et al., 1984). In our study, the optimum number of terminal nodes for each RT model was determined by a minimal deviance complexity cross-validation pruning procedure (Chambers and Hastie, 1992). A 10-fold cross-validation procedure was used to decide the number of optimum terminal nodes (see Table 3).

Table 3. Comparison of the standard deviation of NMSE and r of the goodness-of-fit of testing data set for general linear model (GLM), regression tree (RT) and artificial neural networks (ANN) (see Fig. 5)

Prediction error	Model	Standard deviation					
		TEB	DMN	ECEC	SILT	CBD	PHC
NMSE	GLM	0.040	0.015	0.039	0.029	0.016	0.024
	RT	0.062	0.055	0.071	0.035	0.078	0.051
	ANN	0.015	0.031	0.021	0.034	0.026	0.041
r	GLM	0.024	0.005	0.020	0.027	0.013	0.012
	RT	0.038	0.029	0.066	0.023	0.105	0.060
	ANN	0.009	0.015	0.016	0.029	0.011	0.021

3.2.4. Statistical procedures

The measured values of ECEC, TEB and DMN show a positive skewness. They were transformed to a base-10 logarithmic scale. All soil attributes investigated showed an approximately normal distribution, but the Kolmogorov–Smirnov test indicates that only ECEC shows a normal distribution at the given sample mean and standard deviation. Plant species and soil types were coded as presence (1)/absence (0) for each species and soil type. GLM and RT were conducted using S-PLUS (Mathsoft, 1999), whereas ANN analyses were conducted using NeuroDimension 3.0 (Principe et al., 2000).

The results of model testing are presented as the normalized mean square error (NMSE) and Pearson's correlation coefficient (r). NMSE is calculated by the following equation

$$\mathrm{NMSE} = \frac{\frac{1}{n}\sum_{i=1}^{n}(z_i - \hat{z}_i)^2}{S^2}$$

where z_i is the observed value of sample i, $\hat{z}_i$ is the predicted value of sample i, n is the total number of samples and S^2 is the total variance of observed samples. The normalization of MSE allows a direct comparison of model performance for different soil attributes. In order to test the uncertainty of each model using a single data set, all analyses were repeated 10 times after rerandomising the data set. The uncertainty of model prediction is presented as the standard deviation of NMSE or r.

One of the major limitations of empirical model building exercises is that they require large amounts of data to obtain reliable training results and to validate trained networks (Schultz et al., 2000). This is particularly relevant in this study considering the cost and time necessary for collecting soil samples at varying depths. Theoretical discussion on the optimum amount of sampling is beyond the scope of this paper, but the influence of different numbers of soil samples on the performance of each regression technique was assessed by varying the numbers of training sets (from 50 to 400) against fixed numbers of testing sets (102) (Fig. 5). When more samples are collected in the training phase of the model, the prediction error becomes smaller (Fig. 5). In general, the model performance rapidly improves up to 200 soil samples, then further improvement of model prediction becomes small. The predictions for DMN and SILT still gradually increase towards 400 in the training set, but the increment is rather marginal. No systematic relationship with total variance of soil attributes was recognized. For successive analysis, all 502 samples were first randomised and then split into training (300) and testing sets (202).

Direct interpretation of constructed models for GLM and ANN is difficult. ANN is not able to identify the influence of individual predictors in the model results. In the case of GLM, the possible multicollinearity and also the use of interaction terms prohibit the direct interpretation of the regression model. RT produces a plot for the interpretation of model results, but it is not easy when many environmental predictors are included as predictors. Therefore, a hierarchical approach was applied in order to estimate the importance of individual environmental variables following Schaap et al. (1998). The four groups of environmental parameters were successively brought in at a training phase to test the trained model against the testing set. Park and Burt (in press) show that soil sampling depth is the greatest source of total soil variance except for some soil attributes (e.g. secondary Mn). Therefore, soil sampling depth was brought into the model first, and other environmental variables were added. In total, eight different combinations of environmental variables were presented in this research and their model performance was compared.

4. Results

4.1. Environmental correlation of individual soil attributes

This section first interprets the success of the techniques used for the individual soil attributes, while more detailed discussions on the comparison of the techniques themselves will follow in two subsequent sections (Table 4). In terms of goodness-of-fit, DMN, TEB and PHC all show a high *r* value (>0.80, $p<0.01$) and low NMSE (Fig. 6). Even though the soil samples were collected from varying depths, more than 70% of total variance of those attributes was predicted by environmental factors. Park and Vlek (in press) reviewed previous attempts of environmental correlation and found that only few regression functions show R^2 more than 0.75. Moore et al. (1993) explicitly argue that explaining more than 70% of total variance is difficult due to the intrinsic random variability of soils and other technical constraints such as errors in both measuring soil attributes and collecting environmental variables at relevant scales.

Table 4. Comparison of the standard deviation of NMSE of regression functions with different environmental factors

Soil attributes	Model	*D*	*D*+*V*	*D*+*S*	*D*+*T*	*D*+*S*+*V*	*D*+*T*+*V*	*D*+*T*+*S*	*D*+*T*+*S*+*V*
TEB	GLM	0.043	0.045	0.050	0.050	0.039	0.032	0.055	0.040
	RT	0.042	0.058	0.055	0.074	0.054	0.071	0.046	0.062
	ANN	0.028	0.068	0.028	0.009	0.040	0.038	0.025	0.015
DMN	GLM	0.025	0.187	0.035	0.030	0.034	0.025	0.023	0.015
	RT	0.068	0.040	0.066	0.053	0.081	0.065	0.046	0.055
	ANN	0.025	0.040	0.045	0.057	0.023	0.049	0.043	0.031
ECEC	GLM	0.051	0.070	0.064	0.062	0.037	0.071	0.059	0.039
	RT	0.030	0.035	0.033	0.091	0.012	0.053	0.082	0.071
	ANN	0.041	0.031	0.037	0.063	0.035	0.062	0.042	0.021
SILT	GLM	0.005	0.028	0.055	0.023	0.043	0.043	0.011	0.029
	RT	0.039	0.045	0.074	0.065	0.071	0.214	0.069	0.035
	ANN	0.004	0.035	0.079	0.026	0.025	0.038	0.061	0.034
CBD	GLM	0.036	0.102	0.150	0.022	0.006	0.048	0.004	0.016
	RT	0.051	0.049	0.061	0.070	0.084	0.100	0.089	0.078
	ANN	0.026	0.031	0.038	0.038	0.028	0.055	0.046	0.026
PHC	GLM	0.082	0.165	0.022	0.093	0.021	0.141	0.022	0.024
	RT	0.074	0.091	0.068	0.088	0.116	0.051	0.063	0.051
	ANN	0.085	0.148	0.019	0.080	0.038	0.129	0.034	0.041

See Fig. 8 for the comparison of NMSE of each soil attribute. *D*=depth, *V*=vegetation type, *S*=soil type, *T*=terrain indices.

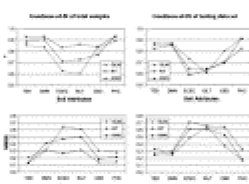 Display Full Size version of this image (16K)

Fig. 6. The comparison of goodness-of-fit between the model for total data set and models for testing data set (trained by 300 soil samples). Terminal nodes for total data set in the RT mode were 27 (TEB), 29 (DMN), 44 (ECEC), 46 (SILT), 44 (CBD) and 24 (PHC) after the 10-fold cross-validation procedures.

The high environmental correlation coefficient of DMN, TEB and PHC is already documented in a previous attempt to identify pedogeomorphological processes on the same slope (Park and Burt, in press). DMN and PHC both belong to the 'solute leaching effect' factor from the PCA analysis. The solute-leaching effect generally shows a clear downslope increase with gradual decrease or increase with depth. As an example, Fig. 2 shows a clear catenary distribution of Mn oxides. Mn oxides on soils upslope generally increase with depth, whereas soils downslope show a steady decrease with depth. Redistribution of Mn is closely tied to the oxidation–reduction dynamic of Mn. Secondary Mn, once released by chemical weathering, moves in a reduced form from upslope, and accumulates in the A horizon at lower slope positions through formation of relatively stable Mn^{2+}–organic matter complexes under higher Eh conditions (McDaniel et al., 1992). Unlike many other soil oxides in highly podzolised soils, the lateral difference along the slope extends to deeper soil horizons, even though the difference diminishes with depth.

Total exchangeable bases, which represents the 'nutrient dynamic effect' on the slope, shows a similar topsoil content across the slope despite the steep slope and catenary variations of soils and vegetation. This characteristic could be explained by the fact that the spatial distribution of major nutrient cations is tightly controlled by the soil–vegetation system (Trudgill, 1988). In subsurface soils, however, the cations freed from such soil–vegetation systems are subject to downslope redistribution according to hydrological flowpaths governing the downslope redistribution (Burt and Park, 1999).

In contrast to the previously discussed soil attributes, silt content ('weathering–erosion effect', Fig. 3), ECEC ('soil-acidification effect') and CBD ('podzolisation effect') show relatively low environmental correlation (Fig. 6). In the case of silt content, clear spatial patterns that are closely associated with geomorphological processes do exist in the surface layers, but rapidly disappear with depth (Fig. 3). On the other hand, the convergent throughflow along the hollow produces a fine soil texture via active hydrolysis processes and possible translocation of fine soil particles along the hollow (Park and Burt, 2000). The dominant slope process on the study slope is throughflow, and its intensity shows a good correspondence with the three-dimensional arrangement of slope forms (Anderson and Park). On the other hand, the limited influence of throughflow and the heterogeneity of the parent material in subsurface soils greatly reduce the systematic catenary variation of soil attributes, which is difficult to model by any numerical methods. The nature of the parent materials, stratified periglacial colluviums, indicates that random components of soil variance might become dominant with depth, while active erosion–sedimentation processes at the surface and in shallow surface soils result in a clear lateral soil differentiation.

ECEC shows even stronger unexplained variation over the slope due to more complex geochemical processes that are involved in the spatial distribution of ECEC. In the case of CBD, however, a much more systematic variation of vertical and also lateral distribution has been recognized by Park et al. (1996). In general, the increase of podzolisation intensity is associated with stronger eluviation in topsoils and illuviation in the podzolic B horizon. The depth of the illuvial B horizon tends to be deeper in better developed podzolic soils. Furthermore, the influence of hydromorphic processes also adds systematic depletion and accumulation patterns, especially for Fe oxides, over the slope. All of these spatial patterns could be clearly presented by a verbal and conceptual model, but detailed modelling by numerical tools is difficult.

4.2. Comparison among statistical techniques

Fig. 6 compares (1) regression functions fitted to the total samples and (2) regression functions fitted to the testing data (202) after training with training data (300). Fig. 7 illustrates the scatter plots for measured versus predicted variables using the total data set. In this figure, RT shows a 'step'-shape distribution among measured and predicted values, which may be explained by the consecutive partitioning algorithm of RT. On the other hand, ANN and GLM show 'smoother' linear patterns between measured and predicted values.

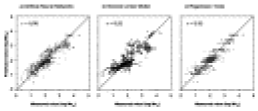 Display Full Size version of this image (10K)

Fig. 7. The comparison of goodness-of-fit among three regression techniques for Mn oxides. Total samples (502 in total) were used in this diagram.

RT shows the best performance for the case of total soil samples; it is followed by ANN and then GLM. The prediction against the testing data set, however, shows the opposite result: GLM and ANN show a better performance than RT. This is a consistent result throughout the hierarchical examination of different environmental variables (Fig. 6) and also the preliminary examination of sample numbers (Fig. 5). This discrepancy may be explained by the possible overfitting of RT function to the training data set. Overfitting of the tree model by excessive partitioning is a well-known disadvantage of the RT model (Breiman et al., 1984), even with the cross-validation procedure used. Overfitting of the ANN model is also recognizable in Fig. 6, even though an early stopping was applied during the training phase for the total sample cases. This indicates that the generalization potential of ANN and RT models is lower than that of GLM.

The direct comparison of GLM and ANN is rather dubious. ANN is a better predictor for PHC, TEB and SILT, but GLM yields better results for DMN, ECEC and CBD. The standard two-sample *t*-test, however, shows that only the mean value of NMSE of CBD and ECEC are statistically different from one another ($p<0.01$). Table 3 shows a comparison of standard deviation of r and NMSE for each technique. ANN and GLM clearly show lower uncertainty than RT, but a clear difference between ANN and GLM is not evident. These observations indicate that ANN and GLM both produce compatible results for empirical prediction of soil variation on the slope. However, considering the complexity involved in the preanalysis of ANN for a decision on internal parameters, GLM is a more straightforward procedure than ANN, with slightly better model performance in the case of ECEC and CBD.

4.3. Best environmental predictors

In the hierarchical analysis of environmental correlation of most soil attributes, GLM and ANN appeared the better techniques, while RT shows a rather erratic result (Fig. 8). ANN and GLM produce fairly comparable results, but GLM clearly outperforms ANN in the case of ECEC and CBD. Considering the observation that these two soil attributes are strongly influenced by vertical pedogenesis, GLM may be considered as the more suitable model-building tool for such soil attributes.

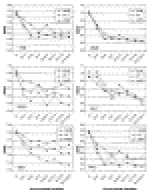 Display Full Size version of this image (30K)

Fig. 8. Hierarchical regression of environmental variables for six soil attributes analysed in this research. *D* stands for sampling depth; *V* stands for vegetation type; *S* stands for soil type; and *T* stands for terrain indices.

In general, the prediction was improved when more predictors were used. In the case of TEB, MND and PHC, there are relatively small differences between prediction errors by landform–surface vegetation and landform–surface vegetation–soil types. These soil attributes (or different pedogeomorphological processes) show a clear lateral differentiation with minor differences in the vertical distribution function ('solute effect' and 'nutrient-dynamic effect'). The prediction by surface information alone, such as landform, surface and vegetation, is attractive for empirical model building because it does not require detailed soil surveys over the landscape. The strong influence of terrain parameters on soil spatial variation is now a well-known principle, which is supported in this study. Vegetation information, in addition to the terrain parameters, further improves the model performance for most of the soil attributes.

Soil attributes that show relatively simple linear distribution patterns in both the vertical and lateral context can be more easily modelled by empirical functions. The successfully modelled three soil attributes in this research are all relatively mobile attributes, which are under the strong influence of solute processes over the slope. In an early discussion of soil redistribution processes, Huggett (1975) argues that more mobile soil attributes reach an equilibrium state much more quickly than less mobile soil attributes. The equilibrium state of soil attributes and environmental variables is also a key assumption of many soil–landscape analyses (Gerrard, 1992). Inevitably, the randomness of soil attributes becomes stronger for the less mobile soil attributes. A good example in this research is the spatial distribution of the silt content. The parent material at the study site is stratified periglacial colluviums originating from quartzitic sandstones (Park et al., 1996). There is a clear spatial gradient visible in the topsoil where active contemporary processes occur, but such spatial patterns become obscure with depth due to the poorly sorted parent materials in the subsurface horizons. This observation could also be applicable to ECEC and CBD. In addition to the parent material factor, their spatial distribution is further complicated by the complex depth functions of these soil attributes, which depend on the intensity of pedogenesis, podzolisation for CBD and soil acidification for ECEC. The complex vertical depth function of these attributes may be far beyond any mathematical and statistical descriptions.

In the case of other soil attributes (ECEC, SILT, CBD), more input variables result in a better model fit. It is especially well demonstrated that soil type (pedogenesis) is important for the prediction of these soil attributes. For example, ECEC shows a much lower prediction using only surface information, such as terrain parameters and vegetation, in comparison with other input variables with soil types. The same argument is valid for SILT and CBD whose spatial distribution is closely controlled by pedogenesis (such as total oxide content on this slope). The higher performance of GLM compared with the other two techniques is a rather unexpected result, especially for soil attributes that show strong nonlinearity in their vertical depth function (ECEC and CBD). This might be caused by the inclusion of interaction terms in the training phase of GLM, while ANN does not explicitly consider that interaction effect during its model training. There is some sort of interaction between environmental factors and the vertical depth function of soil attributes. This might be modelled by interaction of environmental variables rather than statistical approximation of individual predictors.

5. Summary and discussion

This research reveals the importance of understanding soil-forming processes underlying the spatial distribution of individual soil attributes for any environmental correlation of soil attributes over the landscape. Differential involvement of individual attributes in pedological, biological and hydrological processes determines the applicability of environmental correlation of soil attributes. In this study, soil attributes (total exchangeable bases (TEB), Mn oxides (MND) and pH (PHC)) that were influenced by lateral slope processes can be more easily modelled by environmental parameters than soil attributes whose vertical distribution function is strongly governed by pedogenesis (total oxides (CBD) in this study) or strong intrinsic random soil variance (silt content) and effective cation exchange capacity (ECEC) in this study.

As an empirical prediction-modelling tool, GLM with interaction terms outperforms the other two nonlinear adaptive models: ANN and RT, with respect to both the simplicity of model structure and performance of derived empirical functions. Due to the stronger dependency of ANN and RT on training data sets than GLM, the generalizations of the former two techniques become less successful than that of GLM. The expected nonlinearity modelling of vertical soil functions by ANN and RT was not clearly seen in this study. On the contrary, the interaction terms among predictors in the GLM model seem a more effective means to describe the depth function of soil attributes as a combined effect of environmental predictors.

The strong influence of terrain parameters on the soil spatial variation is now a well-known principle, which is confirmed in this study. Vegetation information, in addition to the terrain parameters, further improves the model performance for most soil attributes. Surface information such as landform and vegetation has proved to be an effective environmental predictor in modelling soil distribution. Once again, however, this generalization does not apply to soil attributes whose spatial distribution is strongly controlled by pedogenesis, such as total oxides in podzolised soils and ECEC and silt content whose intrinsic random variability rapidly increase with depth.

Many ecological and environmental models call for more accurate spatial variability of soil attributes over the landscape. Our knowledge and methodological framework of soil prediction are still limited. This is especially the case for the characterization and prediction of three-dimensional spatial variability, which should jointly consider vertical and lateral variation of soils. A more deterministic and physically sound pedological modelling approach must be a future research priority, but empirical environmental correlation is one of the possible means to predict the spatial distribution of soils in a three-dimensional space. Many previous soil–landscape analyses, especially geostatistical approaches, focus on the stochastic interpretation of soil spatial variability rather than on a process-based interpretation. Further attention should be paid to the development of statistical tools to explain the three-dimensional variation with an adequate consideration of environmental gradients in the distribution of individual soil attributes.

Any attempts to empirically model or predict the variability of soil attributes on a landscape scale need to consider the spatial and temporal scale of soil and environmental processes. It is known that different processes and interactions are likely to become more important as we change the spatial scale from the plot level to a regional level (Kirkby et al., 1996). Terrain parameters are likely the most effective soil predictors at the hillslope scale as water and material flows are strongly governed by slope configuration (Huggett, 1975). With increasing spatial scale, however, the predominance of topography as the main predictors would gradually be replaced by other environmental factors, such as parent materials and climatic factors. In addition, the possibility to predict soil attributes using environmental predictors is complicated by our ability to record and analyse environmental factors (Moore et al., 1993). Future soil–landscape studies should address these problems in order to model interactions between spatial variation of soils and other environmental processes.

Acknowledgements

This paper benefited from comments of Kalen Homes (University of California, Santa Babara) and Thomas Oberthür (CIAT). The soil data used were collected while the first author was the member of the School of Geography, Oxford. We are grateful for the financial support for the field and laboratory analyses.

References

Anderson and Burt, 1978. M.G. Anderson and T.P. Burt , The role of topography in controlling throughflow generation. *Earth Surface Processes* **3** (1978), pp. 331–344. **Full Text** via CrossRef

Beckett and Webster, 1971. P. Beckett and R. Webster , Soil variability: a review. *Soils and Fertilizers* **34** (1971), pp. 1–15.

Beven, 1989. K. Beven , Changing ideas in hydrology—the case of physically based models. *Journal of Hydrology* **105** (1989), pp. 157–172. Abstract | **Abstract + References** | PDF (1825 K) | View Record in Scopus | Cited By in Scopus

Bishop, 1995. C.M. Bishop , Neural Networks for Pattern Recognition. , Oxford Univ. Press, Oxford (1995).

Breiman et al., 1984. L. Breiman, J.H. Friedman, R.A. Olshen and C.J. Stone , Classification and Regression Trees. , Wadsworth, Belmont, CA (1984).

Burrough, 1993. P.A. Burrough , Soil variability: a late 20th century view. *Soils and Fertilizers* **56** (1993), pp. 529–562.

Burt and Park, 1999. T.P. Burt and S.J. Park , The distribution of solute processes on an acid hillslope: I. Exchangeable bases. *Earth Surface Processes and Landforms* **24** (1999), pp. 851–865.

Chambers and Hastie, 1992. J.M. Chambers and T.J. Hastie , Statistical Models in S. , Wadsworth, New York (1992).

Dawson and Wilby, 2001. C.W. Dawson and R.L. Wilby , Hydrological modelling using artificial neural networks. *Progress in Physical Geography* **25** (2001), pp. 80–108. **Full Text** via CrossRef | View Record in Scopus | Cited By in Scopus

Gerrard, 1992. J. Gerrard , Soil Geomorphology. , Chapman & Hall, London (1992).

Gessler et al., 1995. P.E. Gessler, I.D. Moore, N.J. McKenzie and R.J. Ryan , Soil–landscape modelling and spatial prediction of soil attributes. *International Journal of Geographical Information Systems* **9** (1995), pp. 421–432. View Record in Scopus | Cited By in Scopus

Heuvelink and Pebesma, 1999. G.B.M. Heuvelink and E.J. Pebesma , Spatial aggregation and soil process modelling. *Geoderma* **89** (1999), pp. 47–65. Abstract | PDF (894 K) | View Record in Scopus | Cited By in Scopus

Hewitt, 1993. A. Hewitt , Predictive modelling in soil survey. *Soils and Fertilizers* **56** (1993), pp. 305–314.

Hoosbeek and Bryant, 1992. M.R. Hoosbeek and R.B. Bryant , Towards the quantitative modeling of pedogenesis: a review. *Geoderma* **55** (1992), pp. 183–210. Abstract | **Abstract + References** | PDF (1757 K) | View Record in Scopus | Cited By in Scopus

Huggett, 1975. R.J. Huggett , Soil landscape systems: a model of soil genesis. *Geoderma* **13** (1975), pp. 1–22. Abstract | **Abstract + References** | PDF (1309 K)

Jongman et al., 1995. R.H.G. Jongman, C.J.F. Ter Braak and O.F.R. van Tongeren , Data Analysis in Community and Landscape Ecology. , Cambridge Univ. Press, Cambridge (1995).

Jury, 1985. W.A. Jury , Spatial variability of soil properties. In: S.C. Hern and S.M. Melancon, Editors, *Vadose Zone Modeling of Organic Pollutants*, Lewis, Chelsea, MI (1985), pp. 245–269.

Kirkby et al., 1996. M.J. Kirkby, A.C. Imeson, G. Bergkamp and L.H. Cammeraat , Scaling up processes and models from the field plot to the watershed and regional scale. *Journal of Soil and Water Conservation* **51** (1996), pp. 391–396.

Lagacherie and Voltz, 2000. P. Lagacherie and M. Voltz , Predicting soil properties over a region using sample information from a mapped reference area and digital elevation data: a conditional probability approach. *Geoderma* **97** (2000), pp. 187–208. Abstract | PDF (1550 K) | View Record in Scopus | Cited By in Scopus

Maier and Dandy, 1998. H.R. Maier and G.C. Dandy , The effect of internal parameters and geometry on the performance of back-propagation neural networks: an empirical study. *Environmental Modelling and Software* **13** (1998), pp. 193–209. SummaryPlus | **Full Text + Links** | PDF (500 K) | View Record in Scopus | Cited By in Scopus

MathSoft, 1999. MathSoft, S-Plus 2000 Professional Release 2. , Insightful Corp., Seattle, WA (1999).

McBratney et al., 2000. A.B. McBratney, I.O.A. Odeh, T.F.A. Bishop, M.S. Dunbar and T.M. Shatar , An overview of pedometric techniques for use in soil survey. *Geoderma* **97** (2000), pp. 293–327. SummaryPlus | **Full Text + Links** | PDF (2700 K) | View Record in Scopus | Cited By in Scopus

McCullagh and Nelder, 1989. P. McCullagh and J.A. Nelder , General Linear Models. (2nd edn. ed.),, Chapman & Hall, London (1989).

McDaniel et al., 1992. P.A. McDaniel, G.R. Bathke, S.W. Buol, D.K. Cassel and A.L. Falen , Secondary manganese/iron ratios as pedochemical indicators of field-scale throughflow water movement. *Soil Science Society of America Journal* **56** (1992), pp. 1211–1217. View Record in Scopus | Cited By in Scopus

McKenzie and Austin, 1993. N.J. McKenzie and M.P. Austin , A quantitative Australian approach to medium and small scale surveys based on soil stratigraphy and environmental correlation. *Geoderma* **57** (1993), pp. 329–355. Abstract | **Abstract + References** | PDF (1487 K) | View Record in Scopus | Cited By in Scopus

McKenzie and Ryan, 1999. N.J. McKenzie and P.J. Ryan , Spatial prediction of soil properties using environmental correlation. *Geoderma* **89** (1999), pp. 67–94.

SummaryPlus | **Full Text + Links** | PDF (2413 K) | View Record in Scopus | Cited By in Scopus

McSweeney et al., 1994. K. McSweeney, P.E. Gessler, B. Slater, R.D. Hammer, J. Bell and G.W. Petersen , Towards a new framework for modelling the soil–landscape continuum. In: R.G. Amundson and J.W. Harden, Editors, *Factors of Soil Formation: A Fiftieth Anniversary Retrospective*, Soil Science Society of America, Madison (1994), pp. 1–13.

Moore et al., 1993. I.D. Moore, P.E. Gessler, G.A. Nielsen and G.W. Peterson , Soil attribute prediction using terrain analysis. *Soil Science Society of America Journal* **57** (1993), pp. 443–452. View Record in Scopus | Cited By in Scopus

Odeh et al., 1994. I.O.A. Odeh, A.B. McBratney and D.J. Chittleborough , Spatial prediction of soil properties from landform attributes derived from a digital elevation model. *Geoderma* **63** (1994), pp. 197–214. View Record in Scopus | Cited By in Scopus

Park and Burt, 1999. S.J. Park and T.P. Burt , Identification of throughflow intensity using the distribution of secondary iron oxides in soils. *Geoderma* **93** (1999), pp. 61–84. SummaryPlus | **Full Text + Links** | PDF (4600 K) | View Record in Scopus | Cited By in Scopus

Park and Burt, 2000. S.J. Park and T.P. Burt , Spatial distribution of chemical weathering intensity on an acid hillslope. *Zeitschrift fuer Geomorphologie* **44** (2000), pp. 379–402. View Record in Scopus | Cited By in Scopus

Park and Burt, 2002. S.J. Park and T.P. Burt , Identification and characterisation of pedogeomorphological processes on a hillslope using an indirect gradient analysis. *Soil Science Society of America Journal* (2002) (in press) .

Park and Vlek, 2002. S.J. Park and P.L.G. Vlek , Soil–landscape analysis as a tool for sustainable land resource management in developing countries. *Professional Geographers* (2002) (in press) .

Park et al., 1996. S.J. Park, T.P. Burt and P.A. Bull , A soil–landscape continuum on a three-dimensional hillslope, Quantock Hills, Somerset. In: M.G. Anderson and S.M. Brooks, Editors, *Advances in Hillslope Processes* **vol. 1**, Wiley, Chichester (1996), pp. 367–396.

Principe et al., 2000. J.C. Principe, N.R. Euliano and W.C. Lefebvre , Neural and Adaptive Systems: Fundamentals Through Simulations. , Wiley, New York (2000).

Ripley, 1996. B.D. Ripley , Pattern Recognition and Neural Networks. , Cambridge Univ. Press, Cambridge (1996).

Schaap et al., 1998. M.G. Schaap, J.L. Feike and M.Th. van Genuchten , Neural network analysis for hierarchical prediction of soil hydraulic properties. *Soil Science Society of America Journal* **62** (1998), pp. 847–855. View Record in Scopus | Cited By in Scopus

Schultz et al., 2000. A. Schultz, R. Wieland and G. Lutze , Neural networks in agroecological modelling—stylish application or helpful tool?. *Computers and Electronics in Agriculture* **29** (2000), pp. 73–97. SummaryPlus | **Full Text + Links** | PDF (526 K) | View Record in Scopus | Cited By in Scopus

Trudgill, 1988. S.T. Trudgill , Soil and Vegetation System. , Clarendon Press, Oxford (1988).

Wilding, 1984. L.P. Wilding , Spatial variability: its documentation, accommodation and implication to soil surveys. In: D.R. Nielsen and J. Bouma, Editors, *Soil Spatial Variability*, Pudoc, Wageningen (1984), pp. 166–193.

Zevenbergen and Thorne, 1987. L.W. Zevenbergen and C.R. Thorne , Quantitative analysis of land surface topography. *Earth Surface Processes and Landforms* **12** (1987), pp. 47–56. **Full Text** via CrossRef | View Record in Scopus | Cited By in Scopus

Zhu, 2000. A.X. Zhu , Mapping soil landscape as spatial continua: the neural network approaches. *Water Resources Research* **36** (2000), pp. 663–677. View Record in Scopus | Cited By in Scopus

Corresponding author. Tel.: +49-228-734926; email: spark@uni-bonn.de