



Prediction of soil depth using environmental variables in an anthropogenic landscape, a case study in the Western Ghats of Kerala, India

Sekhar L. Kuriakose^{a,*}, Sanjaya Devkota^b, D.G. Rossiter^a, V.G. Jetten^a

^a United Nations University-ITC School for Disaster Geo-information Management, Enschede, The Netherlands

^b Forum for Energy and Environmental Development (FEED) Pvt. Ltd., GPO Box: 12756, Kathmandu, Nepal

ARTICLE INFO

Article history:

Received 9 February 2009

Received in revised form 5 May 2009

Accepted 10 May 2009

Keywords:

Soil depth

Regolith

Predictive mapping

Interpolation

Regression kriging

The Western Ghats

Kerala, India

ABSTRACT

Soil (regolith) depth is a crucial input for modeling earth surface phenomena. However, most studies ignore its spatial variability. Techniques that map the spatial variability of soil depth are of three types: (1) physically-based; (2) empirico-statistical from environmental correlates; and (3) interpolation from point observations. In an anthropogenic landscape, soil depth does not depend primarily on natural processes, making it difficult to apply a physically-based approach. The present study compares empirico-statistical methods with geostatistical methods for predicting soil depth in such a landscape: Aruvikkal catchment (9.5 km²) in the Western Ghats of Kerala, India. Regression kriging applied on blocks of 20 m by 20 m using the environmental covariates elevation, slope, aspect, curvature, wetness index, land use and distance from streams, proved to be the best predictor of soil depth. This model explains 52% of the variability of soil depth in the catchment; with a prediction variance of 0.05 to 0.19. A Gaussian simulation was attempted for a more realistic visualization of the depth, as opposed to the smooth kriging prediction. The most important explanatory variable of soil depth in this landscape is land use, as expected from the strong human intervention.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Regolith depth, often referred to by geomorphologists and engineers as soil depth, is defined as the depth from the surface to more-or-less consolidated material. Soil depth in many other landscapes influences vegetation growth (Fuhlendorf and Smeins, 1998; Meyer et al., 2007) and hydro-mechanical responses of the slopes (Bertoldi et al., 2004; DeRose et al., 1991; Wang et al., 2006). Despite also being a major determinant of the rate of earth surface processes such as landslides and soil erosion, most studies have ignored its spatial variability by using constant values over generalized land units in their analysis (Bakker et al., 2005; Bathurst et al., 2007; Talebi et al., 2008). This is due in part, to the fact that it is impossible to observe and map soil depth directly; by definition the bottom of the regolith is hidden from view.

Soil depth varies as a function of many different factors, including slope, land use, curvature, parent material, weathering rate, climate, vegetation cover, upslope contributing area, and lithology (Dietrich et al., 1995; Minasny and McBratney, 1999). Methods to map soil depth over an

area can be classified as (1) physically-based, (2) empirico-statistical from environmental correlates, and (3) interpolation from point samples.

Physically-based methods predict depth from physical rates of weathering, denudation and accumulation, and from physical properties of the regolith or underlying consolidated rock. Dietrich et al. (1995) proposed a method for spatio-temporal prediction of colluvial soil depth based on the mass balance between soil production from underlying bedrock and soil transport by erosion. This method was successfully validated in a sub-catchment in California (Heimsath et al., 1999). D'Odorico (2000) and Minasny and McBratney (1999) also proposed similar soil production functions based on weathering rates and soil transport by erosive processes.

Empirico-statistical methods predict the spatial continuum of soil depth as a deterministic phenomenon (with random, spatially-uncorrelated error) based on empirical relationships established in feature (attribute) space with a set of independent variables which can be measured or estimated at each point in geographic space. They are of two types: inferential and environmental correlation.

Inferential methods use indicators such as plant species to estimate the mean soil depth in an area covered by the indicator (Treiber and Krusinger, 1975, 1979). This method is applicable only in areas with a dominant species and where the soil depth-indicator species relation is known. Inference can also be based on soil map units, using within-class statistical summaries of observed depth for each class. These methods necessarily ignore spatial continuity. Many researchers in land surface processes have resorted to using soil depth

* Corresponding author. International Institute for Geo-information Science and Earth Observation (ITC), Hengelosestraat 99, P.O. Box 6, 7500AA, Enschede, The Netherlands. Tel.: +31 534874504; fax: +31 534874336.

E-mail address: sekhar.lk@gmail.com (S.L. Kuriakose).

maps with crisp class boundaries based on inferential methods for the simulation of soil erosion and plant growth (Bakker et al., 2005; Barbiéro et al., 2007; Fuhlendorf and Smeins, 1998; Jasrotia and Singh, 2006).

Environmental correlation methods use non-soil landscape attributes as predictors of soil depth, usually in multivariate linear or logistic regression models. Predictors are usually terrain attributes, e.g. digital terrain model (DTM), slope and curvature (DeRose et al., 1991; Okimura and Kawatani, 1987; Tsai et al., 2001), compound terrain index (wetness index; WI or CTI) (Boer et al., 1996; Gessler et al., 1995; Saulnier et al., 1997), and relative position of the sampling location on the hillslope (Catani et al., 2007). Canonical correspondence analyses (Odeh et al., 1991), expert knowledge and fuzzy logic (Zhu et al., 2001), principal component analysis and maximum likelihood classification (Boer et al., 1996), and multiple linear regression and maximum likelihood classification (Ziadat, 2005) have also been applied to predict soil depth.

Interpolation based on point samples can be non-geostatistical or geostatistical. Non-geostatistical interpolation methods (e.g. Thiessen polygons, inverse distance) do not need any theory of random fields; thus they can not provide prediction variances. Thampi et al. (1998) used inverse distance interpolation to derive isolines of soil depth and grouped the results to class intervals within terrain mapping units that were adjusted to geomorphic units. Mendonça Santos et al. (2000) used Triangular Irregular Network (TIN) interpolation and a quadratic finite element method to derive the spatial variation of soil horizons in Fribourg, Switzerland. These methods produce soil depth maps that either have abrupt boundaries or else are smooth but not based on explicit models of spatial structure. Thus for spatially continuous modelling of earth surface phenomena they are theoretically inadequate, although in specific cases models based on them may give adequate results.

Geostatistical methods use the spatial autocorrelation of a phenomenon based on the theory of random fields to interpolate it over an area (Goovaerts, 1997). There are three kinds of geostatistical methods that have been used to map soil depth: purely geostatistical, geostatistical with geographic trend, and geostatistical plus environ-

mental correlates. Pure geostatistical methods interpolate the target variable relying solely on point observations of the variable itself. A geographic trend in the coordinates can be incorporated. Finally, the same environmental correlation mentioned above may be integrated with the geostatistical methods.

Odeh et al. (1995) used heterotopic cokriging and regression kriging to predict soil depth in a Murray Darling sub basin. van Beek (2002) used ordinary kriging to spatially interpolate soil resistivity which was further categorized and merged with measured soil depth information using an iterative procedure. Penížek and Borůvka (2006) compared several geostatistical methods for predicting soil depth and concluded that co-kriging with slope yielded better results than ordinary kriging, regression kriging and linear regression. Hengl et al. (2004) compared the spatial prediction quality of various soil properties using six topographical variables and nine soil mapping units. They concluded that soil depth is best predicted using regression kriging with the principal components of the logit-transformed variables altitude, slope, wetness index, mean curvature, stream power index, and watershed and soil types, when compared to ordinary kriging and feature-space linear modelling.

Geostatistical techniques are thus known to be capable of producing spatially continuous soil depth maps (in fact, discretized on a regular grid). The success depends on the study area and the independent variables used. In the present research we attempted to identify a successful geostatistical technique for predicting soil depth over the continuous space of a small catchment heavily modified by anthropogenic activity. The success of prediction was assessed by comparing various indices which quantify the proportion of variability in the data set that is accounted for by the statistical models used. The natural landscape of the study area, the Aruvikkal catchment, a 9.5 km² sub-basin of Tikovil River basin, is part of the Western Ghats of Kerala, India (Fig. 1). It has been extensively modified with terraces, predominately for rubber (*Hevea brasiliensis*) plantations. The catchment is typical of the landslide-prone Western Ghats scarp lands (Kuriakose et al., 2009a).

The bedrock of the catchment consists of Precambrian crystalline rocks predominated by charnockites that weather slowly leading to a

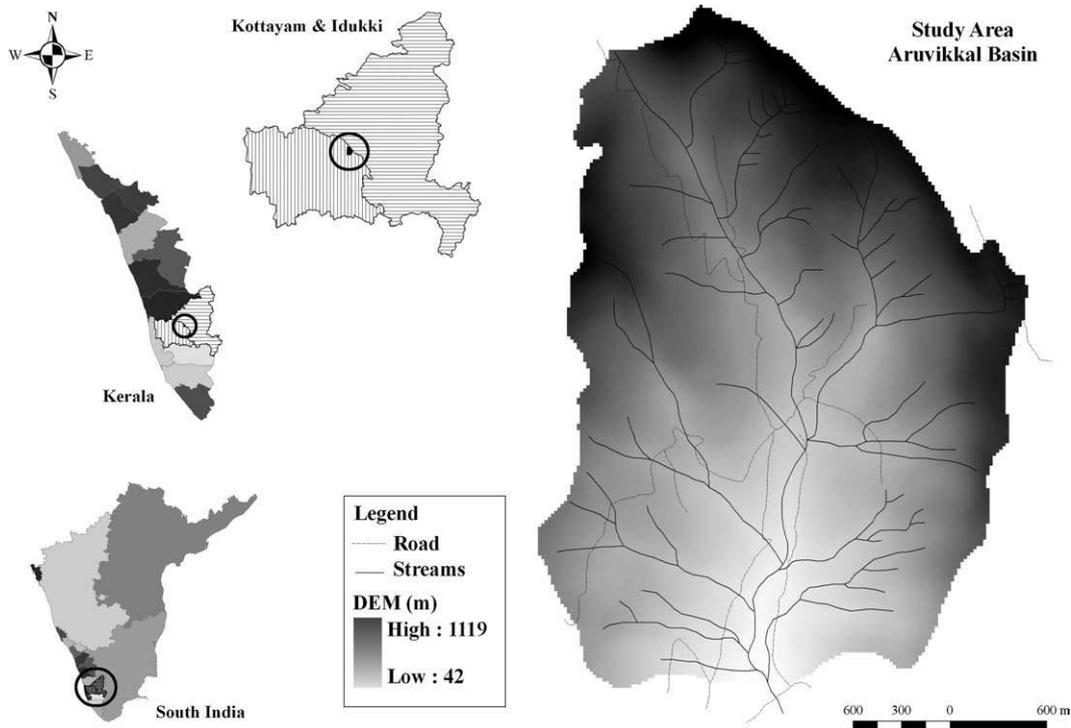
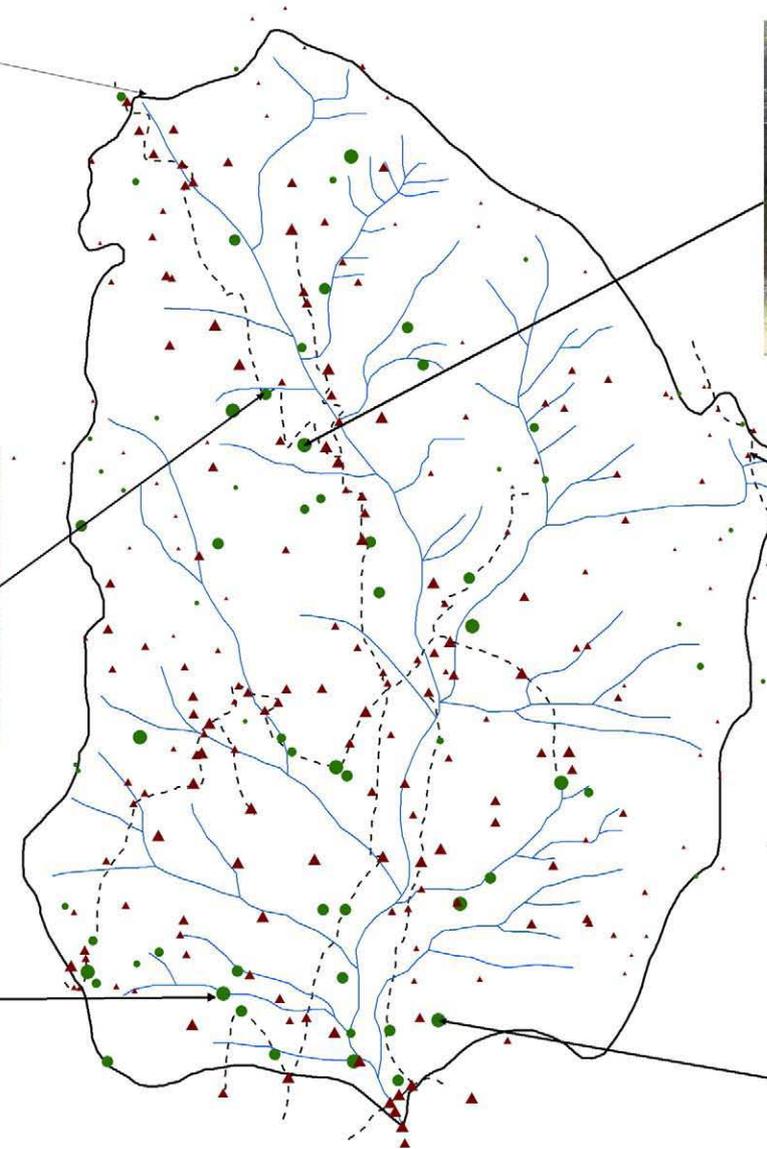


Fig. 1. Study area – Upper Tikovil River Basin, Kerala, India.

Soil Depth Measurements Aruvikkal Basin



Legend	
Training data	
◦	0.00 - 0.25
▲	0.25 - 0.75
▲	0.75 - 1.30
▲	1.30 - 2.00
▲	2.00 - 3.00
Test data	
◦	0.00 - 0.25
◦	0.25 - 0.75
◦	0.75 - 1.30
◦	1.30 - 2.00
◦	2.00 - 3.32
—	Streams
- - -	Road

0 155310 620 930 1,240
Meters

Fig. 2. Locations of soil depth measurements – training and test data.

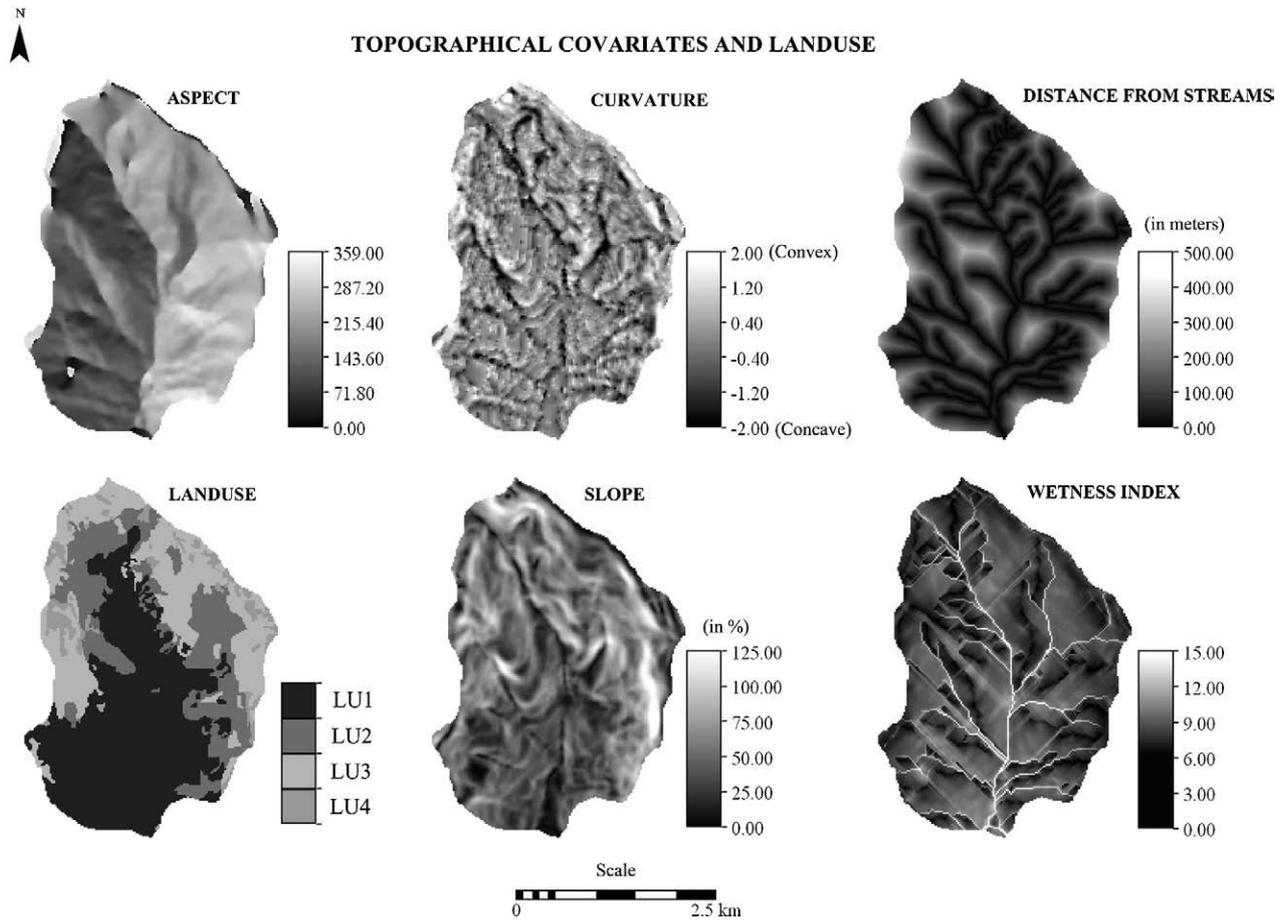


Fig. 3. Topographical covariates and land use map used for predicting soil depth.

shallow regolith (Fig. 2) so that most of the area has a regolith depth of less than 2 m (Thampi et al., 1998). The major current weathering process in the area is hydrolysis under high precipitation (Bourgeon, 2001; Deepthy and Balakrishnan, 2005). Average annual rainfall (1952–2006) at the nearest station is 5182 mm. The denudation rate in the region has been estimated at 20 to 75 m Ma⁻¹ (Gunnell, 1998). Soil production and erosion rates appear to be in equilibrium (Gunnell, 2000).

Anthropogenic land disturbances in the area started in the late 1880s (Victor, 1962). The major land uses are rubber plantations (38% of area) and mixed hill crops (22%) with cassava as an intercrop. Cassava has been implicated in accelerated soil erosion (Putthacharoen et al., 1998). In the plantations, the land is contour terraced and bounded by stone-packed earthen walls made at least every two meters. Further upslope they are often built across hollows and rills in order to gain more cultivable land and trap topsoil washed from upslope. Hollows are

topographic depressions that are often the source areas of the first order ephemeral streams. As soil (colluvial material) accumulates in the hollows, farmers clear the natural vegetative cover, terrace the accumulated soil and plant tuber crops. This human-induced reorganization of soil determines soil depth in the cultivated slopes and thus also limits the applicability of any generic pedogenesis functions for soil depth predictions.

These changes in land use/land cover result in the loss of root reinforcement and the mitigating influence of the canopy on the generation of critical pore water pressure conditions through interception and evapotranspiration (Styczen and Morgan, 1995). Persistence of critical pore water pressure as a consequence of prolonged and high intensity rainfall is often the trigger of shallow landslides in the region (Kuriakose et al., 2008). An earlier attempt (Kuriakose et al., 2009b) to quantify shallow landslide initiation probabilities in the region with physically based models showed that the process is highly

Table 1
Descriptive statistics of soil depth measurements.

Number of points collected using each measurement methods								
Road cut	Slope cut and foundation trenches	Knocking pole	Exposed rock	Pits and open wells	Proxy			
69	49	76	44	21	17			
Distribution of soil depth measurements in each land use class								
Land use	Area (%)	Training data	Testing data	Soil depth (m)				
				Mean	SD	Min	Max	Median
Rubber, rubber fallow settlements	47.05	103	32	1.72	0.79	0.30	3.00	1.50
Mixed crops, mixed crops fallow	23.15	38	12	1.64	0.77	0.25	3.00	1.50
Grass and rock, natural vegetation	23.58	32	9	0.73	0.59	0.00	2.00	0.50
Rock	6.22	40	10	0.00	0.00	0.00	0.00	0.00

Table 2
Descriptive statistics of the topographical covariates.

	Altitude (m from MSL)	Slope%	WI	Curvature	Distance from streams (DS) (m)	Aspect
<i>a) In the catchment</i>						
Mean	596.5	50.1	8.7	0	264.9	173.7
Range	44.3–1113.5	0.3–184.7	5.6–20.7	–3.9–3.7	0–1554	0–360
Standard deviation	271.7	23.0	1.5	0.7	278.0	84.4
<i>b) At the sampling points</i>						
Mean	535.6	47.8	8.8	0.04	105.3	176.5
Range	48.9–1094.1	2.8–183.2	6.1–19.3	–2.1–2.2	0.7–463.2	17.8–331.0
Standard deviation	271.7	24.3	1.7	0.7	90.4	75.6

sensitive to soil depth. Thus the objectives of this paper are, to 1) evaluate pragmatic in-situ measurement techniques for regolith thickness (soil depth) measurement, 2) model the spatial variability of soil depth using topographic attributes, land use and spatial auto-correlation and 3) determine the most successful interpolation technique for soil depth mapping in the study area.

2. Materials and methods

A 20 m horizontal-resolution DEM was prepared using data from Thampi et al. (1998) and used to generate terrain attributes of slope, aspect, curvature and wetness index (WI) (Fig. 3). A steady state

wetness index reflects the spatial distribution of water flow and thus the accumulation process in a closed catchment (Beven and Kirkby, 1979) with higher WI values for those pixels where the local slope is lower. WI was calculated using ILWIS®. Both plan and profile curvatures were calculated in ArcGIS®; negative values represent concave slopes and positive values represent convex slopes. Maps of drainage network and road network were also available from Thampi et al. (1998); these were updated by manual interpretation of satellite images and by field GPS survey. The distance of every pixel from the nearest stream segment (DS) was computed using the Proximity function of ArcGIS 9.2®. A land use map was prepared by visual interpretation of an Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) image from January 31st 2007 based on detailed field knowledge (Fig. 3). Land use was generalized and grouped into four classes thought to be related to soil depth in order to have sufficient observations in each class. The classes were a) Rubber, rubber fallow, and settlements (LU1), b) Mixed crops, mixed fallow (LU2), c) Grass and rock, and natural vegetation (LU3) and d) Exposed rock (LU4) (Fig. 3 and Table 1).

Field work was conducted in August and September 2007. Because of the rough terrain, purposive sampling of representative sites was used. Measurements were made along the 15 km road sections (Fig. 2) and in open pits where the regolith down to bedrock was exposed. In areas where no profile sections were exposed, 76 measurements were made by the knocking pole method (Uchida et al., 2008): an iron spear was pierced into the soil until it reached bed rock. Three situations can arise with this method, 1) a representative surface to bedrock measurement, 2) an underestimate, measured only to a rubble layer and 3) a local overestimate, caused by cracks in the bedrock (Murray, 1999). Measuring soil depths with this method required minimal training, which allowed us to utilize the service of two local

Table 3
a) Intercept and slope, b) the AIC values of various multiple linear regression equations used for the regression kriging and c) the nugget, structural sill and range of the three residual variograms used for BRK_{lu}, BRK_{all} and BRK_{stall}.

a) Intercept, slope and significance						
Predictor	LM _{all}		LM _{stall}		LM _{lu}	
	Intercept	Slope	Intercept	Slope	Intercept	Slope
Altitude	1.7858 *****	–0.0006**	2.1650*****	–0.0006**	1.7592*****	–
Slope		–0.0009α		NA		–
Wetness index		0.0399α		NA		–
Distance from streams		0.0006α		NA		–
Aspect		–0.0012*		–0.0012*		–
Curvature		0.0915α		NA		–
Rubber, rubber fallow, settlement (LU1)		NA		NA		NA
Mixed crops, mixed crops fallow (LU2)		–0.0833α		–0.1197α		–0.3211**
Grass and rock, natural vegetation (LU3)		–0.7326*****		–0.7373*****		–0.9629*****
Rock (LU4)		–1.4066*****		–1.4154*****		–1.7592*****
b) AIC values						
Null regression	LM _{all}		LM _{stall}		LM _{lu}	
595.1	459.9		455.8		459.7	
Significance codes: (****) 0–0.001; (***) 0.001–0.01; (**) 0.01–0.05; (*) 0.05–0.1; (α) 0.1–1						
c) Nugget, structural sill and range of the residual variograms used for BRK _{lu} , BRK _{all} and BRK _{stall}						
Method	Nugget		Structural Sill		Range	
BRK _{lu}	0.32		0.19		390	
BRK _{all}	0.38		0.12		710	
BRK _{stall}	0.31		0.18		393	

Table 4
Statistics and accuracy assessment of the soil depth predictions by different methods.

	BOK	LM _{stall}	LM _{all}	BRK _{stall}	BRK _{lu}	BRK _{all}	SRK _{all1}	SRK _{all2}	SRK _{all3}	SRK _{all4}
a) Minimum, maximum, mean and standard deviation of predictions (m)										
Min	0.02	-0.30	-0.31	-0.30	-0.21	-0.37	-0.78	-0.97	-0.72	-1.10
Max	2.77	2.13	2.50	2.71	2.65	2.72	3.15	2.74	2.63	3.36
Mean	1.19	1.32	1.31	1.31	1.33	1.29	1.31	1.17	1.26	1.28
Std	0.64	0.54	0.54	0.59	0.57	0.59	0.63	0.64	0.56	0.72
b) Accuracy assessment										
R ²	0.31	0.44	0.47	0.49	0.51	0.52	0.47	0.41	0.43	0.42
RMSE	0.77	0.67	0.66	0.65	0.63	0.63	0.66	0.71	0.69	0.71
NRMSE	0.26	0.23	0.22	0.22	0.21	0.21	0.22	0.24	0.23	0.24

Min: minimum, Max: maximum, Std: standard deviation, R²: coefficient of determination, RMSE: root mean square error, NRMSE: normalized root mean square error.

inhabitants. The prerequisite for conducting the measurements was inexpensive devices such as iron spears and pounding hammers. Thus the method was a 'pragmatic in-situ technique' which could be reproduced by local farmers to identify deep soil pockets in steep slopes and hollows that are possible slope failure zones.

Care was taken to minimize these errors. Percussion was done until the bed rock was reached; it could be recognized by the sound of the knocking pole. The deepest measurement required ~15 min; about 20 working days were required to complete 259 measurements. In addition to the measured samples, 17 zero-depth points were marked so as to represent the bare rock areas (Table 1). The locations of these zero-depth points were determined to be bare rock areas both based on field work and the land use/land cover map. The sample set was randomly divided into two parts, 75% as training data and 25% as testing data (Fig. 2) using the *Create Subset* operation within the Geostatistical Wizard of ArcGIS 9.2®.

Four methods were compared for predicting soil depth: (1) feature-space linear models with environmental predictors (LM), (2) block ordinary kriging (BOK), (3) block regression kriging with the environmental predictors (BRK) and, (4) stochastic simulation (SRK) with the best BRK model. All kriging predictions were averages of 20 m by 20 m square blocks. Although soil depths were necessarily measured at near-points with minimal geostatistical support, most phenomena for which depth is relevant do not operate at this size blocks, but rather at some larger block size. For example, mass movement is neither observed nor modeled at points, but rather at blocks and so block kriging was preferred over punctual kriging. Block kriging has the additional advantage of substantially reducing kriging prediction variance, since spatial variability smaller than the block size is averaged out. A detailed description of these methods can be found in Hengl (2007).

LM was implemented (1) with land use categories, (2) with all predictors and (3) reverse stepwise regression beginning with all predictors and removing them one-by-one until the most parsimonious model is achieved, as evaluated by Akaike's information criterion (AIC). This latter method is an automatic procedure for statistical model selection in cases where there are a large number of predictors, and no underlying theory on which the model selection can be based (Venables and Ripley, 2002). The optimal model selection in the step method implementation was based on the minimum Akaike's information criterion (AIC). All statistical and geostatistical computations were carried out in the R environment for statistical computing (Ihaka and Gentleman, 1996) using the *gstat* package (Pebesma, 2004). LM maps were produced in ILWIS 3.3 using the equations derived from the linear modelling methods (*lm* and *step*) in R.

Accuracy was assessed by comparing the values of the root mean square error (RMSE), RMSE normalized by the observation range (NRMSE) (Hengl, 2007) and coefficient of determination (R²) (Nash and Sutcliffe, 1970) between the observed and predicted values of soil depth at the validation data locations. A higher R² indicates a better fit between the observed and predicted, while lower values of RMSE and

NRMSE indicate better prediction accuracy. For BOK and BRK predictions median kriging prediction standard deviation over the catchment was also computed.

As the feature space linear models and kriging are not convex predictors, it is possible to obtain values outside the original data range; some may be realistic but others clearly are not. In particular, negative slope coefficients in the case of feature space linear models and negative weights assigned to deep residual depths can result in negative predictions of soil depth, especially in regions near zero-depth observations. These non-physical predictions were converted to zero-depth, i.e. bare rock, before the map was used for accuracy assessment and further applications. Another limitation was that the stream network in the catchment was not incorporated directly into the models and thereby into the kriging procedures; WI is only a partial substitute. This resulted in unrealistic prediction of soil depth along the downcutting perennial streams that are devoid of any soil depth in reality. However, as there were no sampling points within the stream network, the accuracy assessment was not influenced by the presence and absence of soil depth in those pixels that are part of the stream network.

In addition to modelling with the topographic variables, principal component analysis was carried out on standardized variables to

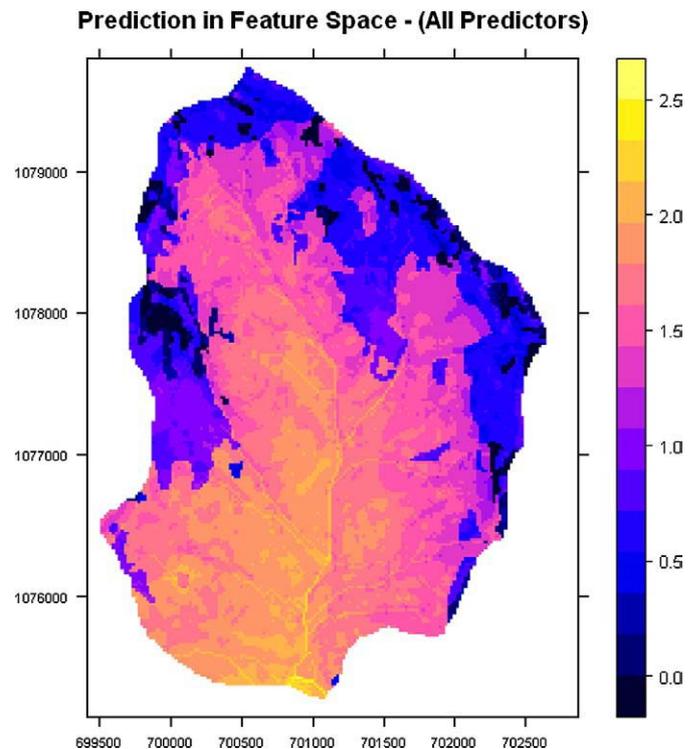


Fig. 4. Soil depth prediction in feature space with the multiple linear model using all variables.

account for co-linearity. The results in this case were not better than those using the original variables and so are not described herein.

3. Results and discussion

Table 1 shows the frequency of soil depth measurements made using each measurement method, the proportional area of each land use class, the frequency of training and testing samples in each land use class and the mean and standard deviation of soil depth in each land use class.

Table 2 (a and b) indicates that the sampling locations are representative of the topographical characteristics of the study area. Maximum measured soil depth in the region was 3 m near the catchment

outlet (Fig. 2); average soil depth of the basin is 1.24 m \pm 0.96. The large variation in soil depth is because of the presence of in situ residual soil pockets and accumulated deep alluvial soil near the outlet, as well as bare rock. The residual soil pockets are adjacent to the plateau margins and most often are topographical hollows.

3.1. Feature-space linear models

Three linear models were computed: one-way ANOVA on land use class (LM_{lu}), LM with all variables (LM_{all}) and backwards stepwise LM (LM_{stall}). Table 3 (a, b and c) shows the equations, the corresponding AIC values, the significance of the intercepts and slopes and the parameters of the residual variograms used for regression kriging.

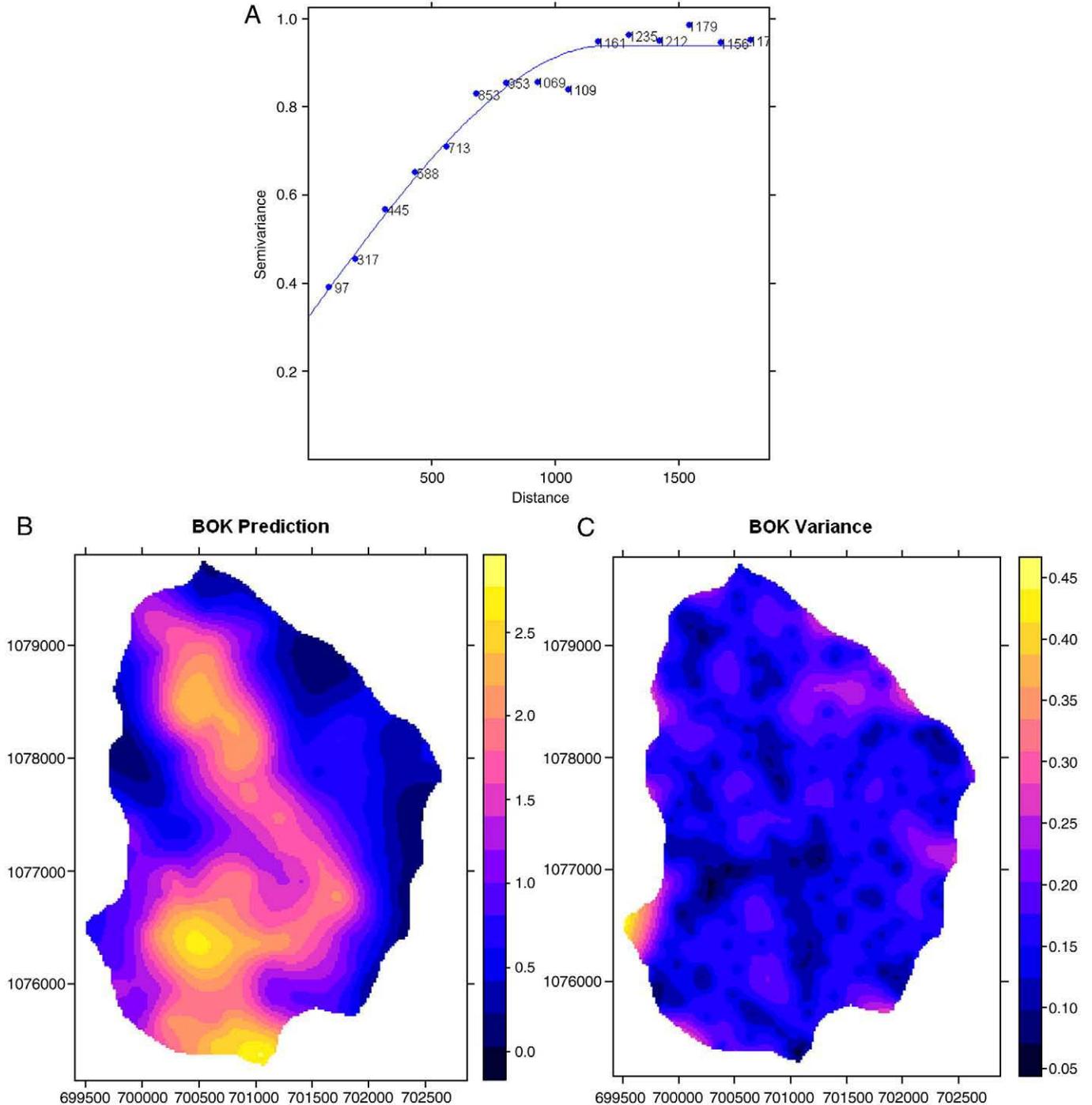


Fig. 5. Soil depth interpolated using ordinary kriging on 20 m by 20 m blocks – A) empirical variogram fitted with the spherical model, B) prediction and C) prediction variance.

Residuals from all models were used for BRK. Model results were compared with the prediction based on ordinary kriging which is based on the null regression (i.e. prediction by the mean) and with each other. Table 4 (a and b) shows the statistics and accuracy assessment of the soil depth predictions. Linear models based on other individual topographical covariables gave poor AIC values and thus were not further used.

Model LM_{lu} shows, as expected, that LU4 (exposed rock) have the highest influence in determining the soil depth of the region followed by LU3 and LU2 (Table 3a). These are contrasted to the effect of LU1 (rubber) which is associated with the highest soil depths in the region (shown in the model intercept). The AIC of this model (Table 3b) is comparable to that of LM_{all} (see next) implying that soil depth in the region can best be predicted in feature space by land use types,

without having to use additional topographical correlates. The significance (Table 3a) of the intercept and the slope values indicates that each land use class has a significant linear effect on soil depth. The adjusted R^2 of the model was 0.47; i.e. about half of the variance of soil depth in the catchment is explained by land use. This model could be used to predict soil depth; however, because the predictor is for classes and the resulting map would show one depth for each class. This is not realistic, so such a map was not produced; instead the model was used to derive residuals for BRK (below).

Model LM_{all} achieved an adjusted R^2 of 0.49, only a bit higher than the single-predictor model from land use. Fig. 4 shows the soil depth predicted using this model. The land use dependence of soil depth in the catchment is represented in the map. The effect of the topographic variables, even though not significant in the model, is evident. Model

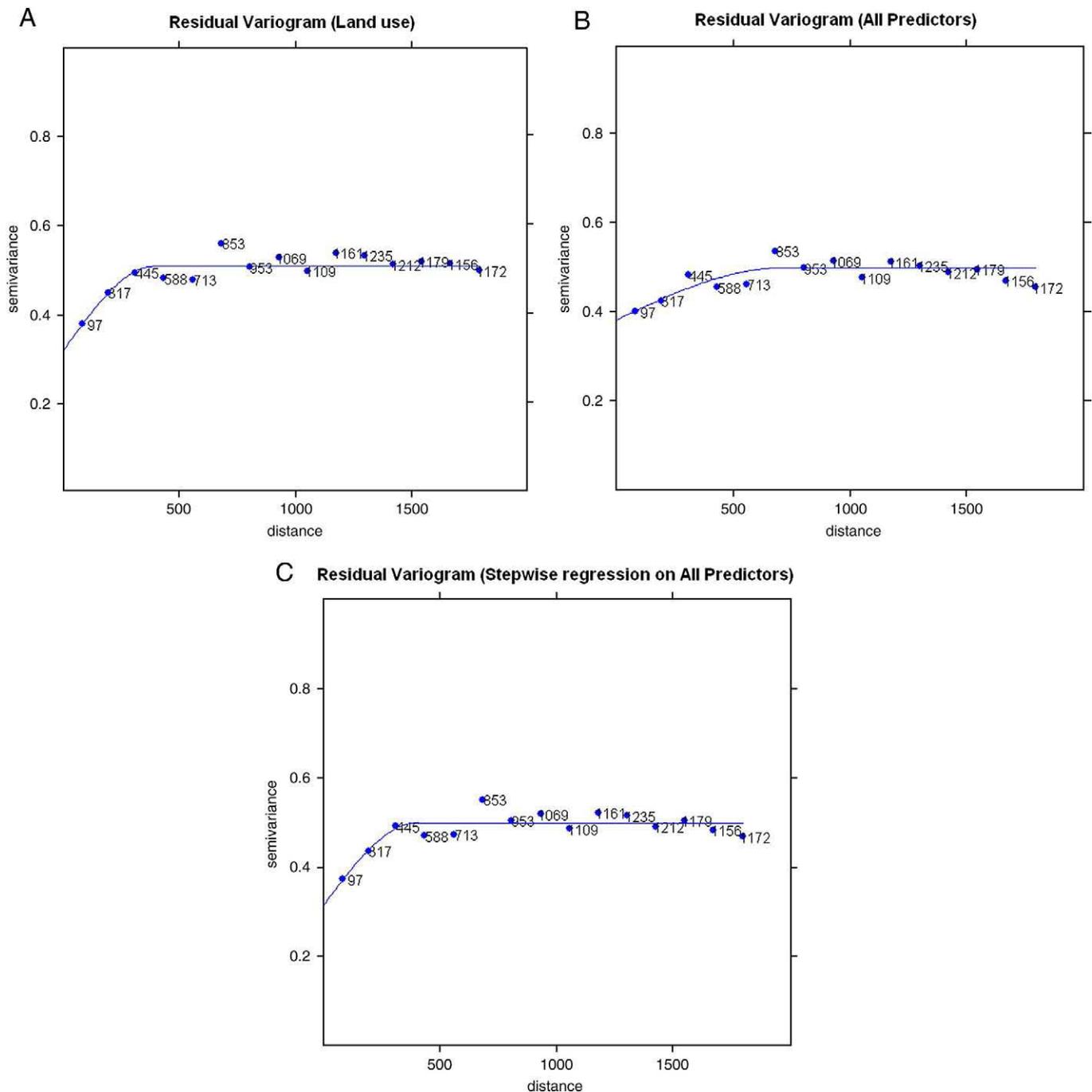


Fig. 6. Residual variogram fitted with the spherical model – A) land use, B) all predictors and C) backward stepwise regression using all variables.

LM_{stall} reduced the number of predictors from ten to five with a lower AIC value but with the same R^2 as that of LM_{all} 0.49. Accuracy assessment indicates that LM_{all} is a slightly better predictor of soil depth than the more parsimonious LM_{stall} (Table 4b).

3.2. Ordinary kriging on blocks

Ordinary kriging (OK) uses the spatial correlation structure of the data to compute weights for linear prediction from known points. An experimental variogram (Fig. 5A) was derived from the training data set. An isotropic spherical model with a nugget value of 0.32, range of 1213 m and a structural sill of 0.615 was visually and subsequently automatically fitted using *gstat* to this, and the model was used for OK.

The resultant map (Fig. 5B) matches the general trend of soil depth in the catchment as it clearly shows that valleys have a higher soil depth than steep slopes. However OK predictions are by construction smooth and in this case does not represent reality, as the land use dependence of the soil depth is not reflected in the prediction. Minimum soil depth in the catchment was best predicted by this method (10 cm) compared to others; however it underestimated the mean and maximum soil depth; standard deviation of the prediction is 64 cm which is lower than that of the sampling data (Table 4a). This method yielded the lowest R^2 and highest RMSE and NRMSE (Table 4b) thereby proving to be the poorest predictor.

Kriging also produces a map of the prediction variance (Fig. 5C), which is determined by the arrangement of sample points and the variogram model. Prediction variance of this model ranges from about 0.05 (in the centre) to 0.45 (in the extreme W of the map, far from sample points).

Geostatistical interpolation on blocks rather than points supports results in much lower kriging prediction variances (Goovaerts, 1999) which are more realistic for predictive modeling studies. In this case the median kriging prediction variance lowered from 0.52 (punctual) to 0.18 (block). The predictions themselves hardly change, the maximum being -0.01 m between the punctual and the block methods.

3.3. Regression kriging on blocks

Residuals from LM_{lu} , LM_{all} and LM_{stall} were used to create empirical variograms, which were fitted with spherical models (Table 3c and Fig. 6). As expected by theory, the nuggets are similar to the original variogram and the structural sill and range are both much reduced (compare Fig. 5A): half the variability and much of the longer-range dependence is explained by the feature-space linear model and hence not reflected in the variograms.

The residual variograms from LM_{lu} and LM_{stall} showed a steep climb from nugget to sill, while that of LM_{all} had a more gradual rise to the sill. The gradual rise of the variogram curve of LM_{all} can be attributed to the presence of variables such as distance from rivers and slope which has a gradual rise in values, while variables such as land use and aspect vary abruptly in the region inducing a much steeper rise to the sill with the curves of LM_{lu} and LM_{stall} . In all three cases the nugget is a large fraction of the total sill, meaning that unexplained variation dominates after the feature-space effects are removed. These respective variograms were used in block kriging to predict soil depth with a) land use as predictor (BRK_{lu}), b) with all predictors (BRK_{all}) and c) with backwards stepwise method using all predictors (BRK_{stall}).

3.3.1. BRK_{lu}

Fig. 7A shows the result of BRK_{lu} ; the boundaries between land uses are evident. However, within each land use category soil depth varies smoothly; thus the advantages of both feature-space modelling and local spatial prediction are combined. The observed maximum soil depth was underestimated, while the mean over estimated it

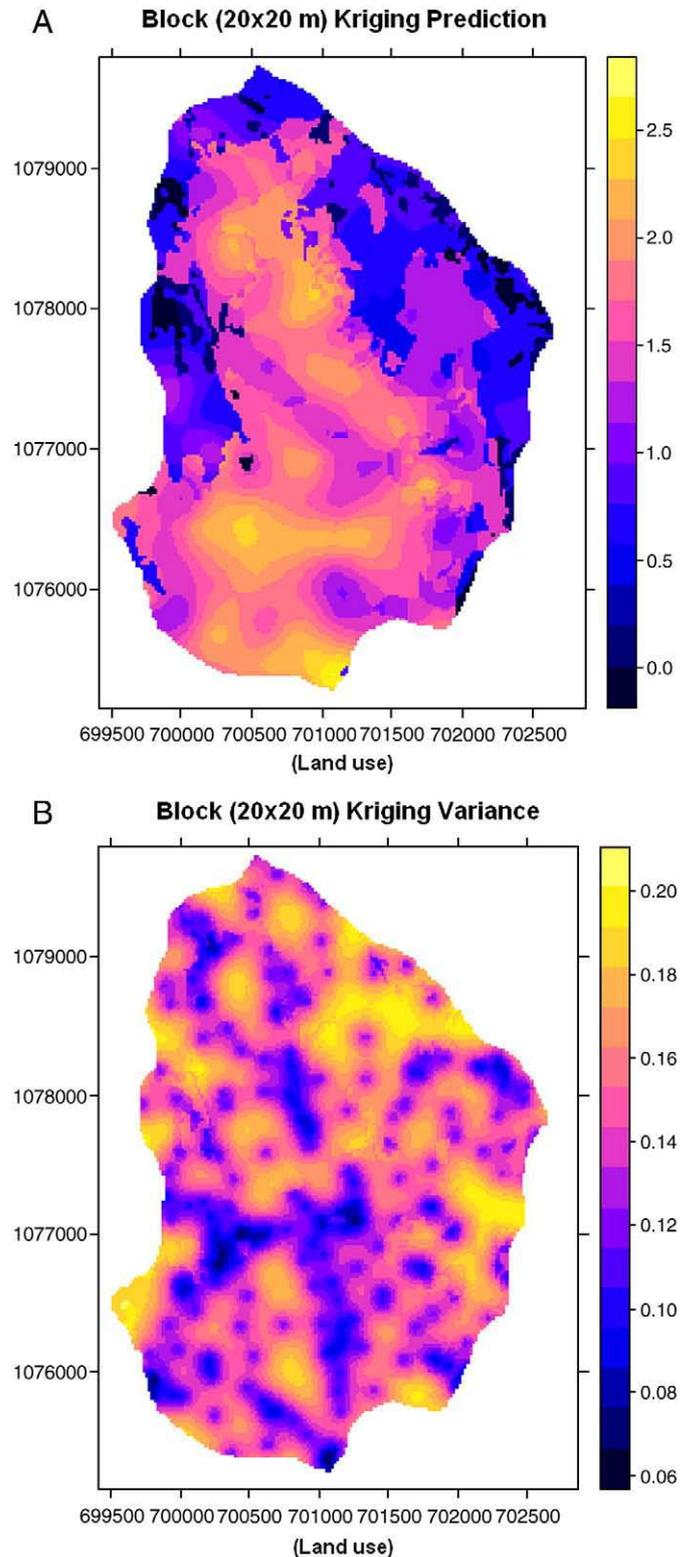


Fig. 7. Soil depth interpolated using regression kriging on 20 m by 20 m blocks with land use – A) prediction and B) prediction variance.

(Table 4b). Fig. 7B shows the soil depth prediction variances, ranging from 0.07 to 0.20.

This model is a better predictor than LM_{stall} (Table 4b). Thus land use combined with local observations can be used to predict soil depth in this region with reasonable confidence.

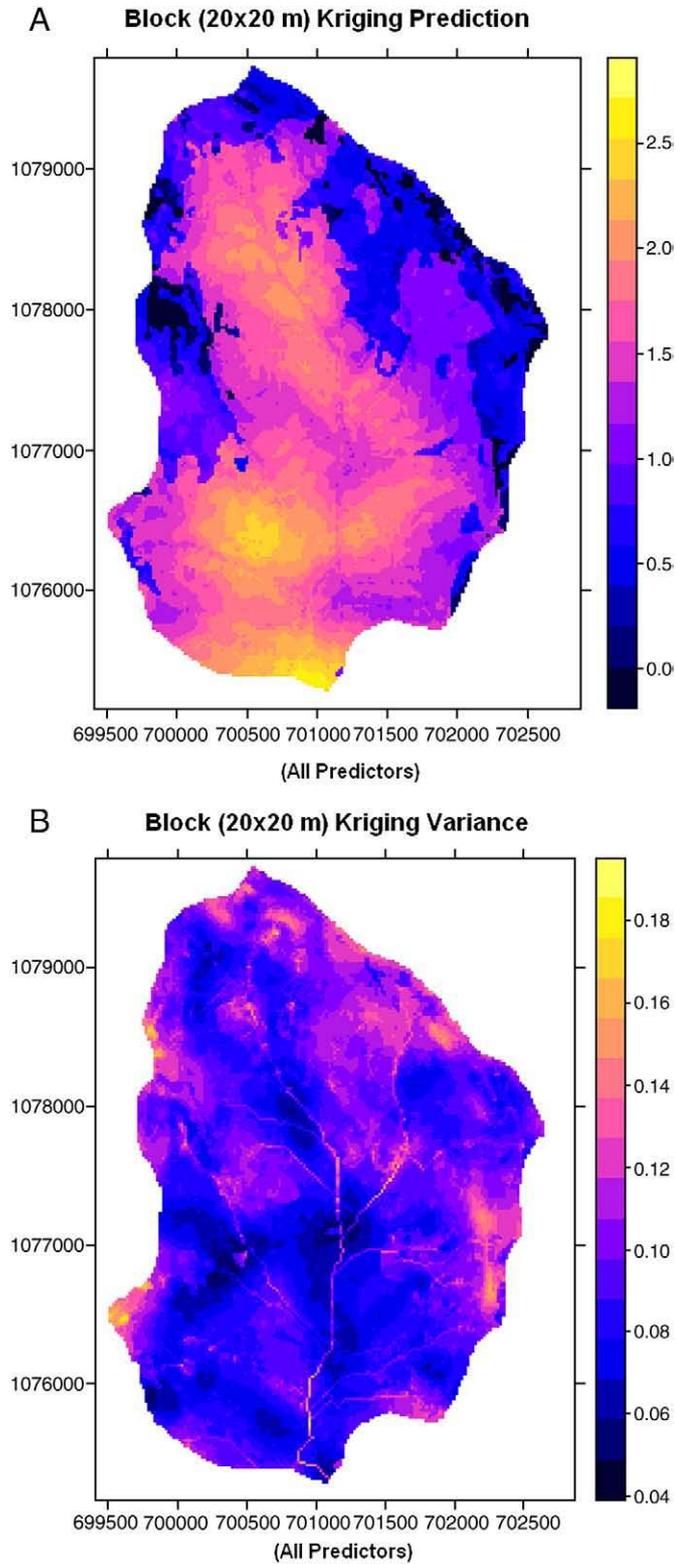


Fig. 8. Soil depth interpolated using regression kriging on 20 m by 20 m blocks with all variables – A) prediction and B) prediction variance.

3.3.2. BRK_{all}

Fig. 8A shows the result of BRK_{all}. This map appears the most realistic, based on field knowledge. The dependence of soil depth on land use is well represented and so is the variation of soil depth within each land use. Influence of altitude, aspect and slope are clearly evident. Regions of thin and no soil cover also seem to be well represented (cf. Fig. 2). The observed maximum and mean soil depth

was marginally under-estimated, by 28 cm and 5 cm respectively. Fig. 8B shows the soil depth prediction variance based on this model, ranging from 0.05 to 0.19.

This model is the best predictor: R^2 is the highest and RMSE and NRMSE are the lowest of all models (Table 4b).

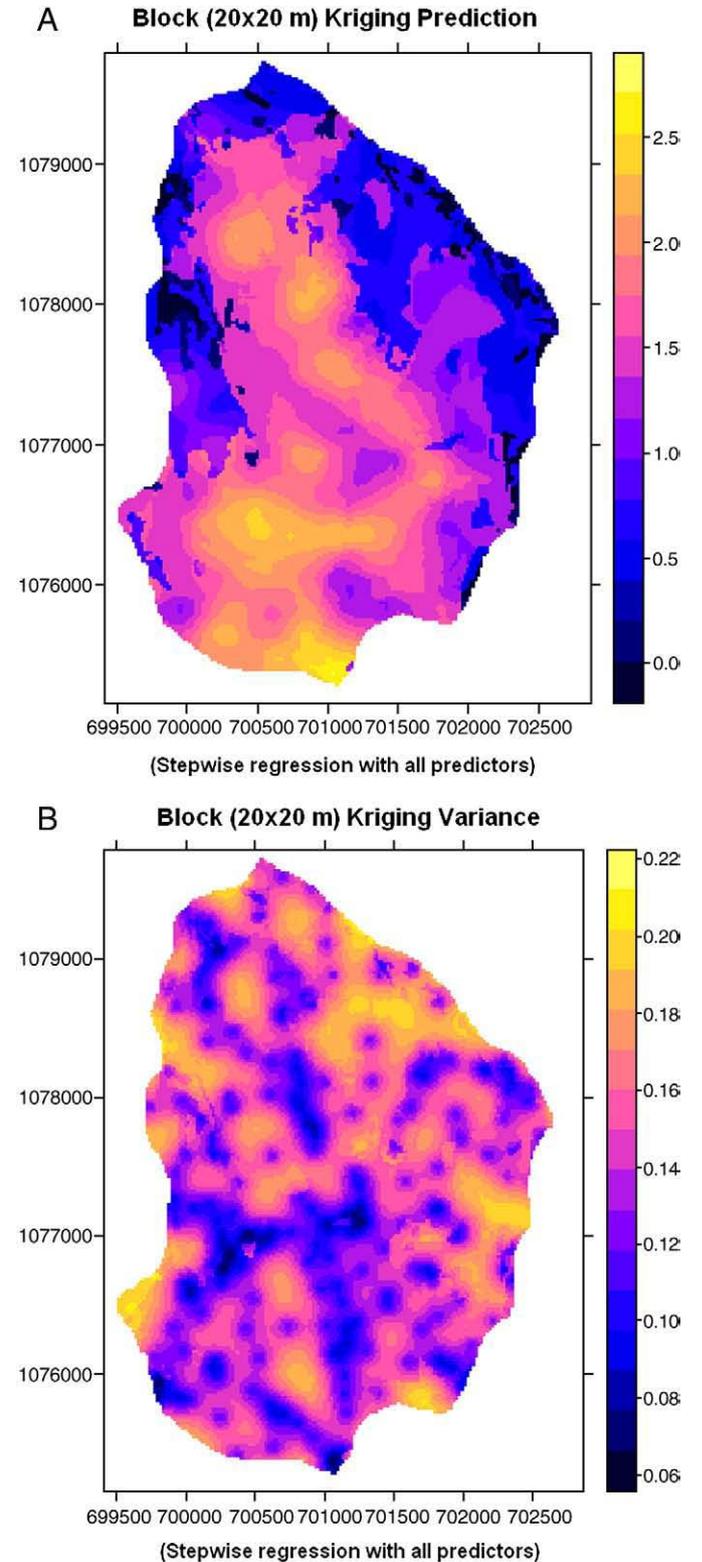


Fig. 9. Soil depth interpolated using regression kriging on 20 m by 20 m blocks with backward stepwise regression method using all variables – A) prediction and B) prediction variance.

3.3.3. BRK_{stall}

Considering the AIC value of the feature-space model LM_{stall} (Table 3b), predictions based on this model should provide the best results. On the contrary, BRK_{stall} yielded the poorest result which can be attributed to the removal of several topographic covariates; although they did not statistically contribute to a parsimonious linear model, the residuals based on the more complete model better represented local spatial variability. Fig. 9A shows the predictions and Fig. 9B shows the prediction variance. The observed maximum and mean soil depths were over-estimated (Table 4a) compared to the measured maximum and mean. The range of variance was 0.07–0.21.

Agreement between the observed and the predicted (R^2) is the poorest, and RMSE and NRMSE are the highest for this model (Table 4b). However, these measures are all quite close for the three BRK models. It is thus evident that even when a less complex LM model yields the lowest AIC value compared to that of a more complex one it cannot be taken for granted that the simpler model will be a more accurate model for the use in regression kriging.

3.4. Stochastic simulation

Geostatistical interpolation is optimal at each block (if the model of spatial dependence is correct), but when viewed over the entire field, the results are unrealistically smooth if there is any nugget effect (Goovaerts, 1997) as is the case here. The nugget effect represents short-range variability and measurement error that can not be resolved by spatial interpolation, and thus adjacent locations are not expected to be identical. Kriging by construction produces a smooth surface, because a slight change in location produces only a slight

change in weights for the weighted average. When smooth maps are used in distributed models that simulate transport across the landscape they will often result in unrealistically smooth flows. This is overcome by conditional stochastic (Gaussian) simulation: reproducing the statistics of the random field modeled by the fitted variogram, respecting the observation points. Each simulation is a different possible representation of reality. The *gstat* package of R (Pebesma, 2004) implements stochastic simulation as an option of its *krige* method.

Fig. 10 shows four simulations (SRK_{all} 1 to 4) of the random field implemented with BRK_{all} . The overall pattern is similar because the sample points and land use classes are respected, but the noise introduced by the nugget effect is different in each case. Summary statistics and accuracy (Table 4b) are comparable with those from BRK_{all} , as expected.

4. Conclusion

Feature-space linear models gave reasonably accurate predictions of soil depth in this catchment mainly because of the strong influence of land use. The all-predictors model was slightly better than the one reduced by backwards stepwise regression, despite the apparent parsimony of the latter. Ordinary kriging was the least accurate and also poorly-represented field knowledge of the complex pattern of soil depth. Regression kriging corrected this deficiency; resulting maps represented both the covariables and local variability. Performance of regression kriging improved when all variables (LM_{all}) were used for residual variogram computation as against using only land use (LM_{lu}) information or a parsimonious model selected through step method (LM_{stall}). Block kriging substantially reduced prediction variance, compared to punctual kriging since decisions are made on areas rather than points the resulting prediction variance is useful for land use management.

All the geostatistical models gave similar prediction accuracy because, in this landscape, land uses, especially the rock outcrops, are major controlling factors for soil depth. Thus when no other data is available, BRK_{lu} is a good predictor of soil depth in strongly influenced anthropogenic landscapes such as the Western Ghats scarps. Accuracy improves marginally with the addition of topographic predictors, but unlike in less-intervened areas, these are not particularly important. With the highest R^2 , RMSE close to other models and the lowest kriging variance, BRK_{all} is the 'best' predictor of soil depth amongst the interpolation techniques evaluated. In other landscapes covariates such as stream power index, landform index, viewshed and soil types should be examined for their predictive quality. In light of the previous slope stability modeling experience in the region (Kuriakose et al., 2009b) it is evident that the overall uncertainty will be reduced significantly by using the continuous soil depth map derived using BRK_{all} . The stochastic simulation results coupled with Monte Carlo procedures is an appropriate tool for accurate error propagation and the prediction variance can improve the sensitivity analysis substantially.

Prediction variance cannot be better than measurement precision; the three possible errors of the knocking pole method could also contribute to the nugget effect of the variograms. Predictions can be improved to any desired precision up to the magnitude of the nugget of the residual variogram, by more efficient and intensive sampling, using the regression kriging prediction variances as an optimization criterion as proposed by Brus and Heuvelink (2007). A further advantage of kriging over non-geostatistical interpolation methods is that the spatial field can be simulated, providing a realistic representation of spatial variation of soil depth for distributed physically based modeling. Thus, the methods and model presented here are generally applicable in the anthropogenically modified terrain of the Western Ghats scarp lands, depending on the availability of relevant feature-space covariates in each particular landscape.

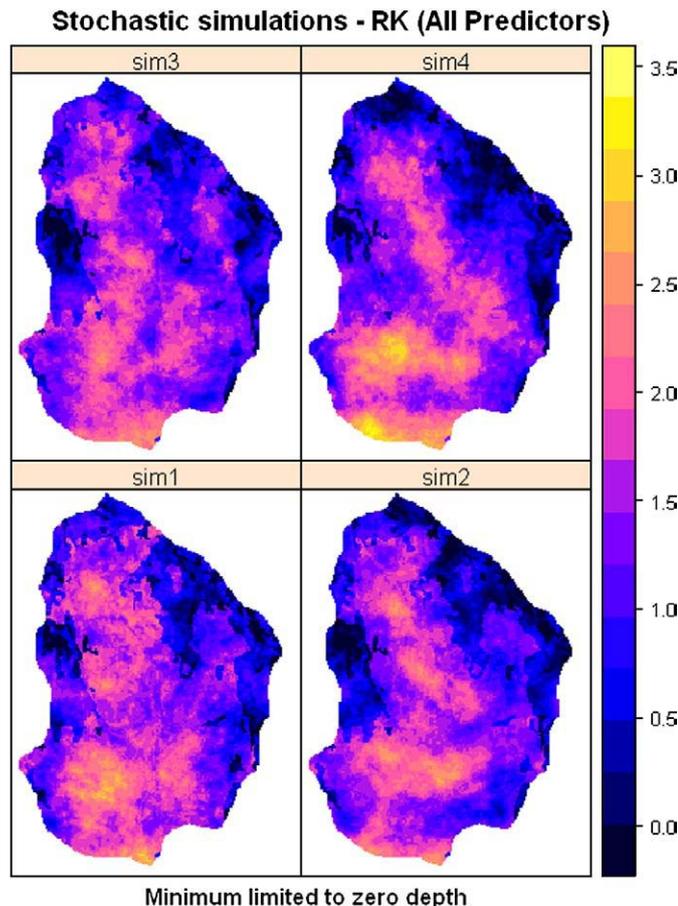


Fig. 10. Stochastic simulation of soil depth interpolated using regression kriging on 20 m by 20 m blocks with all variables.

Acknowledgement

We are thankful to Mr. P.C George, Member of the Legislative Assembly, Poonjar Constituency, for his constant support and help in carrying out the research. We thank Dr. M. Baba, Director, and Mr. G. Sankar, Scientist, CESS for their support in carrying out this research. We thank Mr. Jaibi N. Jacob, Vagamom, Mr. Krishnankutty, Meladukkam and Mr. Siji for the field assistance that they provided. Thanks are due to the Members and the Secretary of the Tikoy Panchayath and Fr. Roy P. Thomas, Vicar, CSI Church, Meladukkam. We also thank the efforts of the reviewers for helping us improve the article to its present form.

References

- Bakker, M.M., Govers, G., Kosmas, C., Vanacker, V., van Oost, K., Rounsevell, M., 2005. Soil erosion as a driver of land-use change. *Agriculture, Ecosystems & Environment* 105 (3), 467–481.
- Barbiéro, L., Parate, R.H., Desclotres, M., Bost, A., Furian, S., Kumar, M.M.S., Kumar, C., Braun, J.-J., 2007. Using a structural approach to identify relationships between soil and erosion in a semi-humid forested area, South India. *Catena* 70, 313–329.
- Bathurst, J.C., Moretti, G., El-Hames, A., Begueria, S., Garcia-Ruiz, J.M., 2007. Modelling the impact of forest loss on shallow landslide sediment yield, Ijuez river catchment, Spanish Pyrenees. *Hydrology and Earth System Sciences* 11 (1), 569–583.
- Bertoldi, G., Dietrich, W.E., Miller, N.L., Rigon, R., 2004. Role of bedrock heterogeneities and soil thickness on the saturation overland flow dynamics in headwater catchments: field observation and simulation using a distributed hydrological model. In: Teuling, A.J., Leijnse, H., Troch, P.A., Sheffield, J., Wood, E.F. (Eds.), 2nd International CAHMDA Workshop on: The Terrestrial Water Cycle: Modelling and Data Assimilation Across Catchment Scales, Princeton, USA, pp. 25–27.
- Beven, K.J., Kirkby, M.J., 1979. A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin* 24 (1), 43–69.
- Boer, M., Del Barrio, G., Puigdefabres, J., 1996. Mapping soil depth classes in dry Mediterranean areas using terrain attributes derived from a digital elevation model. *Geoderma* 72 (1–2), 99–118.
- Bourgeon, G., 2001. A survey of soils and weathering patterns through land system mapping in the Western Ghats Region. In: Gunnell, Y., Radhakrishnan, B.P. (Eds.), *Sahyadri, The Great Escarpment of the Indian Subcontinent*. Geological Society of India, Bangalore, India, pp. 855–904.
- Brus, D.J., Heuvelink, G.B.M., 2007. Optimization of sample patterns for universal kriging of environmental variables. *Geoderma* 138 (1–2), 86–95.
- Catani, F., Segoni, S., Falorni, G., 2007. Accurate basin scale soil depth modelling and its impact on shallow landslides prediction. European Geosciences Union General Assembly 2007: Geophysical Research Abstracts, vol. 9(10828). Vienna, Austria. <http://www.cosis.net/abstracts/EGU2007/10828/EGU2007-J-10828.pdf>.
- D'Odorico, P., 2000. A possible bistable evolution of soil thickness. *Journal of Geophysical Research* 105 (B11), 25927–25935.
- Deepthi, R., Balakrishnan, S., 2005. Climatic control on clay mineral formation: evidence from weathering profiles developed on either side of the Western Ghats. *Journal of Earth System Sciences* 115 (5), 545–556.
- DeRose, R.C., Trustrum, N.A., Blaschke, P.M., 1991. Geomorphic change implied by regolith–slope relationships on steepland hillslopes, Taranaki, New Zealand. *Catena* 18 (5), 489–514.
- Dietrich, W.E., Reiss, R., Hsu, M.-L., Montgomery, D.R., 1995. A process-based model for colluvial soil depth and shallow landsliding using digital elevation data. *Hydrological Processes* 9, 383–400.
- Fuhlendorf, S.D., Smeins, F.E., 1998. The influence of soil depth on plant species response to grazing within a semi-arid savanna. *Plant Ecology* 138, 89–96.
- Gessler, P.E., Moore, I.D., McKenzie, N.J., Ryan, P.J., 1995. Soil-landscape modelling and spatial prediction of soil attributes. *International Journal of Geographical Information Systems* 9 (4), 421–432.
- Goovaerts, P., 1997. *Geostatistics for Natural Resources Evaluation*. Applied Geostatistics. Oxford University Press, Oxford, 483 pp.
- Goovaerts, P., 1999. Geostatistics in soil science: state-of-the-art and perspectives. *Geoderma* 89, 1–45.
- Gunnell, Y., 1998. Present, past and potential denudation rates: is there a link? Tentative evidence from fission-track data, river sediment loads and terrain analysis in the South Indian shield. *Geomorphology* 25, 135–153.
- Gunnell, Y., 2000. The characterization of steady state in earth surface systems: findings from the gradient modelling of an Indian climosequence. *Geomorphology* 35 (1–2), 11–20.
- Heimsath, A.M., Dietrich, W.E., Nishizumi, K., Finkel, R.C., 1999. Cosmogenic nuclides, topography, and the spatial variation of soil depth. *Geomorphology* 27 (1–2), 151–172.
- Hengl, T., 2007. A practical guide to geostatistical mapping of environmental variables. European Commission, Joint Research Centre and Institute for Environment and Sustainability, Luxembourg. URL: http://eussoils.jrc.it/ESDB_Archive/eussoils_docs/other/EUR22904en.pdf.
- Hengl, T., Heuvelink, G.B.M., Stein, A., 2004. A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma* 120 (1–2), 75–93.
- Ihaka, R., Gentleman, R., 1996. R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics* 5 (3), 299–314.
- Jasrotia, A.S., Singh, R., 2006. Modelling runoff and soil erosion in a catchment area, using the GIS, in the Himalayan region, India. *Environmental Geology* 51, 29–37.
- Kuriakose, S.L., Jetten, V.G., van Westen, C.J., Sankar, G., van Beek, L.P.H., 2008. Pore water pressure as a trigger of shallow landslides in the Western Ghats of Kerala, India: some preliminary observations from an experimental catchment. *Physical Geography* 29 (4), 374–386.
- Kuriakose, S.L., Sankar, G., Muraliedharan, C., 2009a. History of landslide susceptibility and a chorology of landslide prone areas in the Western Ghats of Kerala, India. *Environmental Geology* 57 (7), 1153–1568.
- Kuriakose, S.L., van Beek, L.P.H., van Westen, C.J., 2009b. Parameterizing a physically based shallow landslide model in a data poor region. *Earth Surface Processes and Landforms* 34 (6), 867–881.
- Mendonça Santos, M.L., Guenat, C., Bouzelboudjen, M., Golay, F., 2000. Three-dimensional GIS cartography applied to the study of the spatial variation of soil horizons in a Swiss floodplain. *Geoderma* 97 (3–4), 351–366.
- Meyer, M.D., North, M.P., Gray, A.N., Zald, H.S.J., 2007. Influence of soil thickness on stand characteristics in a Sierra Nevada mixed-conifer forest. *Plant and Soil* 294 (1–2), 113–123.
- Minasny, B., McBratney, A.B., 1999. A rudimentary mechanistic model for soil production and landscape development. *Geoderma* 90, 3–21.
- Murray, M.R., 1999. Kriging soil depth: the benefits of prior statistical analysis. In: Arkinson, P.M., Riding, A.E., Tate, N.J. (Eds.), *Geostats-UK '99 Conference*, 22nd November, University of Leicester, UK, pp. 101–116.
- Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I – a discussion of principles. *Journal of Hydrology* 10 (3), 282–290.
- Odeh, I.O.A., Chittleborough, D.J., McBratney, A.B., 1991. Elucidation of soil–landform interrelationships by canonical ordination analysis. *Geoderma* 49 (1–2), 1–32.
- 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 (3–4), 215–226.
- Okimura, T., Kawatani, T., 1987. Prediction of slope failure using the estimated depth of the potential failure layer. *Journal of Natural Disaster Science* 11 (1), 67–89.
- Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. *Computers & Geosciences* 30 (7), 683–691.
- Penížek, V., Borůvka, L., 2006. Soil depth prediction supported by primary terrain attributes: a comparison of methods. *Plant, Soil and Environment* 52 (9), 424–430.
- Putthacharoen, S., Howeler, R.H., Jantawat, S., Vichukit, V., 1998. Nutrient uptake and soil erosion losses in cassava and six other crops in a Psamment in eastern Thailand. *Field Crops Research* 57 (1), 113–126.
- Saulnier, G.-M., Beven, K., Oblod, C., 1997. Including spatially variable effective soil depths in TOPMODEL. *Journal of Hydrology* 202 (1–4), 158–172.
- Styczen, M.E., Morgan, R.P.C., 1995. Engineering properties of vegetation. In: Morgan, R.P.C., Rickson, R.J. (Eds.), *Slope Stabilization and Erosion Control: A Bioengineering Approach*. E & FN Spon, London, UK, pp. 5–58.
- Talebi, A., Troch, P.A., Uijlenhoet, R., 2008. A steady-state analytical slope stability model for complex hillslopes. *Hydrological Processes* 22 (4), 546–553.
- Thampi, P.K., Mathai, J., Sankar, G., Sidharthan, S., 1998. Evaluation study in terms of landslide mitigation in parts of Western Ghats, Kerala. Research report submitted to the Ministry of Agriculture, Government of India, Centre for Earth Science Studies, Government of Kerala, Thiruvananthapuram, India, 100 pp.
- Treiber, M., Krusinger, A.E., 1975. Inferential techniques for soil depth determinations. Part I. *Coleogyne ramosissima* Torr (Black-Brush). Accession Number: ADA024355, Army Engineer Topographic Labs, Fort Belvoir, USA. <http://stinet.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA024355>, 29 pp.
- Treiber, M., Krusinger, A.E., 1979. Inferential techniques for soil depth determinations. Part II. *Artemisia filifolia* Torr (Sand Sagebrush). A449660, Army Engineer Topographic Labs, Fort Belvoir, USA. <http://www.stormingmedia.us/44/4496/A449660.html>, 30 pp.
- Tsai, C.C., Chen, Z.S., Duh, C.T., Horng, F.W., 2001. Prediction of soil depth using a soil-landscape regression model: a case study on forest soils in southern Taiwan. *National Science Council of the Republic of China, Part B: Life Sciences* 25 (1), 34–39.
- Uchida, T., Tamura, K., Mori, N., 2008. A simple method for producing probabilistic shallow landslide hazard maps using soil thickness dataset. European Geosciences Union General Assembly 2008: Geophysical Research Abstracts, vol. 10 (EGU2008-A-03941), Vienna, Austria. <http://www.cosis.net/abstracts/EGU2008/03941/EGU2008-A-03941.pdf>.
- van Beek, L.P.H., 2002. Assessment of the Influence of Changes in Landuse and Climate on Landslide Activity in a Mediterranean Environment. PhD Thesis, University of Utrecht, Utrecht, The Netherlands, 363 pp.
- Venables, W.N., Ripley, B.D., 2002. *Modern Applied Statistics with S*. Springer-Verlag, New York, 518 pp.
- Victor, O.C.D., 1962. *Kurisumala: A Socio-Economic Survey*. St. Joseph's Apostolic Seminary, Alwaye, India, 87 pp.
- Wang, J., Endreny, T.A., Hassett, J.M., 2006. Power function decay of hydraulic conductivity for a TOPMODEL-based infiltration routine. *Hydrological Processes* 20, 3825–3834.
- Zhu, A.X., Hudson, B., Burt, J., Lubich, K., Simonson, D., 2001. Soil mapping using GIS, expert knowledge, and fuzzy logic. *Soil Science Society of America Journal* 65 (5), 1463–1472.
- Ziadat, F.M., 2005. Analyzing digital terrain attributes to predict soil attributes for a relatively large area. *Soil Science Society of America Journal* 69 (5), 1590–1598.