Spatial prediction of soil properties in a watershed scale through maximum likelihood approach

ARTICLE *in* ENVIRONMENTAL EARTH SCIENCES · APRIL 2012

Impact Factor: 1.77 · DOI: 10.1007/s12665-011-1185-7

3 AUTHORS, INCLUDING:

Priyabrata Santra
ICAR-Central Arid Zone Research Institute (…

B S Das
IIT Kharagpur

Environ Earth Sci (2012) 65:2051–2061
DOI 10.1007/s12665-011-1185-7

ORIGINAL ARTICLE

Spatial prediction of soil properties in a watershed scale through maximum likelihood approach

Priyabrata Santra · Bhabani Sankar Das · Debashish Chakravarty

Received: 8 May 2010 / Accepted: 2 July 2011 / Published online: 26 July 2011

Abstract Surface map of soil properties plays an important role in various applications in a watershed. Ordinary kriging (OK) and regression kriging (RK) are conventionally used to prepare these surface maps but generally need large number of regularly girded soil samples. In this context, REML-EBLUP (REsidual Maximum Likelihood estimation of semivariogram parameters followed by Empirical Best Linear Unbiased Prediction) shown capable but not fully tested in a watershed scale. In this study, REML-EBLUP approach was applied to prepare surface maps of several soil properties in a hilly watershed of Eastern India and the performance was compared with conventionally used spatial interpolation methods: OK and RK. Evaluation of these three spatial interpolation methods through root-mean-squared residuals (RMSR) and mean squared deviation ratio (MSDR) showed better performance of REML-EBLUP over the other methods. Reduction in sample size through random selection of sampling points from full dataset also resulted in better performance of REML-EBLUP over OK and RK approach. The detailed investigation on effect of sample number on performance of spatial interpolation methods concluded that a minimum sampling density of 4/km^2 may successfully be adopted for spatial prediction of soil properties in a watershed scale using the REML-EBLUP approach.

Keywords Residual maximum likelihood · Best linear unbiased prediction · Kriging · Semivariogram · Soil hydraulic properties · Elevation

Introduction

Modeling runoff and sediment loss from a watershed through physically based distributed hydrological model needs information on soil properties especially the soil hydraulic properties, e.g., saturated hydraulic conductivity (K_s) and water retention parameters describing the relationship between volumetric water contents (θ) and matric potential head (h) (Du et al. 2009; Santra et al. 2011). Preparation of surface map of these properties through direct measurements at several locations within a watershed is time consuming and expensive. Spatial interpolation techniques serve as an alternative to create desired surface maps from a few measured values. Inverse distance weighting (Uygur et al. 2010) and several geostatistical methods (Goovaerts 1997) are generally used for preparing thematic soil maps. Recently, Baskan et al. (2010) and Rakhmatullaev et al. (2010) have proposed a sequential Gaussian simulation (SGS) (also, called as turning band interpolation) to interpolate soil properties. Despite continued innovations on interpolation approaches, the kriging approach continues to be one of the most common approaches to obtain the surface maps of soil properties from point-based measurements (Goovaerts 1998; Goovaerts 1999; Iqbal et al. 2005; Robinson and Metternicht 2006; Santra et al. 2008; Dong et al. 2011).

Present Address:
P. Santra (✉)
Division of Agricultural Engineering for Arid Production Systems, Central Arid Zone Research Institute, Jodhpur, Rajasthan 345001, India
e-mail: priyabrata.iitkgp@gmail.com

B. S. Das
Department of Agricultural and Food Engineering, Indian Institute of Technology, Kharagpur 721302, India

D. Chakravarty
Department of Mining Engineering, Indian Institute of Technology, Kharagpur 721302, India

The kriging approach requires a semivariogram that describes the spatial correlation structure for the variable to be interpolated. Precise estimation of semivariogram parameters is a key step for this purpose. The most common approach of estimating semivariogram parameters is the calculation of experimental semivariogram from raw data and fitting it to standard semivariogram models, such as exponential model, spherical model, Gaussian model, etc. The experimental semivariogram is commonly computed through the Matheron's (Matheron 1965) methods of moment (MoM) approach, which needs at least 100–150 sampling sites (Webster and Oliver 1992). A sample of this size is often beyond the budget of many geostatistical applications especially in a watershed-scale. As an alternative to the MoM approach, the maximum likelihood (ML) approach has been proposed for better estimation of a semivariogram structure from limited spatial data with sample sizes even below 90 (Pardo-Igúzquiza 1998; Lark 2000; Kerry and Oliver 2007). Besides the issue of sample size, the available standard semivariogram models may not adequately describe an experimental semivariogram over small lags. To fulfill this shortcoming, a more flexible model, the Matérn semivariogram model has been proposed (Matérn 1960; Stein 1999; Diggle et al. 2003; Minasny and McBratney 2005). In this approach, the spatial correlation structures of soil properties are more conveniently described using a smoothness parameter (v) in the Matérn model, specifically when soil properties are influenced by several auxiliary variables (Stein 1999; Minasny and McBratney 2007). The smoothness parameter is also important for determining the semivariogram estimator from irregularly located observations.

Among different kriging methods, ordinary kriging (OK) is considered as the best method (Issaks and Srivastava 1989) and is most suitable for preparing soil maps (Goovaerts 1999). When two soil properties are correlated, co-kriging is observed to be a better approach than OK (Basaran et al. 2010). In case of interrelations between spatial attributes, multivariate geostatistical methods such as factorial kriging may be a better option (Goovaerts 1992, 1994; Lin 2002). If the goal of geostatistical analysis is to detect the spatial patterns of extreme values, such as hot spots of pollutants in soil, indicator kriging (IK) or sequential indicator simulation (SIS) are the suitable options (Juang et al. 2004). Recently, the kriging approach was integrated with triangular network interpolation for severely skewed data with several peak values (Wu et al. 2010). In the presence of trend in dataset, regression kriging (RK) is popular among pedometricians because it is easy to use and performs better than OK and co-kriging (Odeh et al. 1995; Baxter and Oliver 2005; Herbst et al. 2006; Simbahan et al. 2006). However, Cressie (1993) and Lark et al. (2006) pointed out the theoretical biasness of RK in estimating the semivariogram from the residuals and therefore suggested the pedometricians to use a statistically sound method called REML-EBLUP (REsidual Maximum Likelihood estimation of semivariogram parameters followed by Empirical Best Linear Unbiased Prediction of a variable). The REML approach is capable of handling spatial data with presence of trend and estimates the semivariogram parameters and trend parameters directly from the data (Patterson and Thompson 1971; Kitanidis 1983; Stein 1999). The trend in spatial data could be because of geographical coordinates (internal trend) or ancillary variables (external trend). Moreover, the REML-EBLUP will be advantageous when the external trend variable is relatively cheap to measure, such as topographical indices derived from a digital elevation model, or a remotely sensed measurement of the land surface (Lark et al. 2006). Recently, Minasny and McBratney (2007) reported the superiority of the REML-EBLUP approach to OK and RK for spatial data with presence of trend. Therefore, in a hilly watershed, where sampling in regular grid is a difficult task, REML-EBLUP approach may be a good option for spatial prediction of soil properties. In hill slope pedogenic processes, there is also chance of soil variation due to topography, which may be taken as ancillary information in REML-EBLUP approach. Therefore, in this study, the REML-EBLUP approach was evaluated for preparation of surface map of several soil properties in a hilly watershed of Eastern India. The performance of REML-EBLUP was compared with conventionally used spatial interpolation methods such as OK and RK. Another problem for geostatistical applications is identification of the optimum number of samples required to apply them in a watershed scale. It was reported in earlier literature that in case of small number of soil samples, REML-EBLUP performs better than OK and RK (Lark et al. 2006; Minasny and McBratney 2007). Therefore, we have also tested the comparative performance of OK, RK, and REML-EBLUP with different sample sizes in a hilly watershed.

Materials and methods

Study area

The present study was carried out at the Dengei Pahad Watershed (DPW), which is a part of the western catchment of Chilika Lake system in Orissa, India, and falls within the Northeastern Ghat agro-climatic zone under hot and sub-humid climate (Fig. 1). It is located between 19°49′48″–19°52′8.4″N and 85°13′55.2″–85°14′34.8″E. The average annual rainfall of the study area is 1,130 cm, of which the major portion occurs during the monsoon

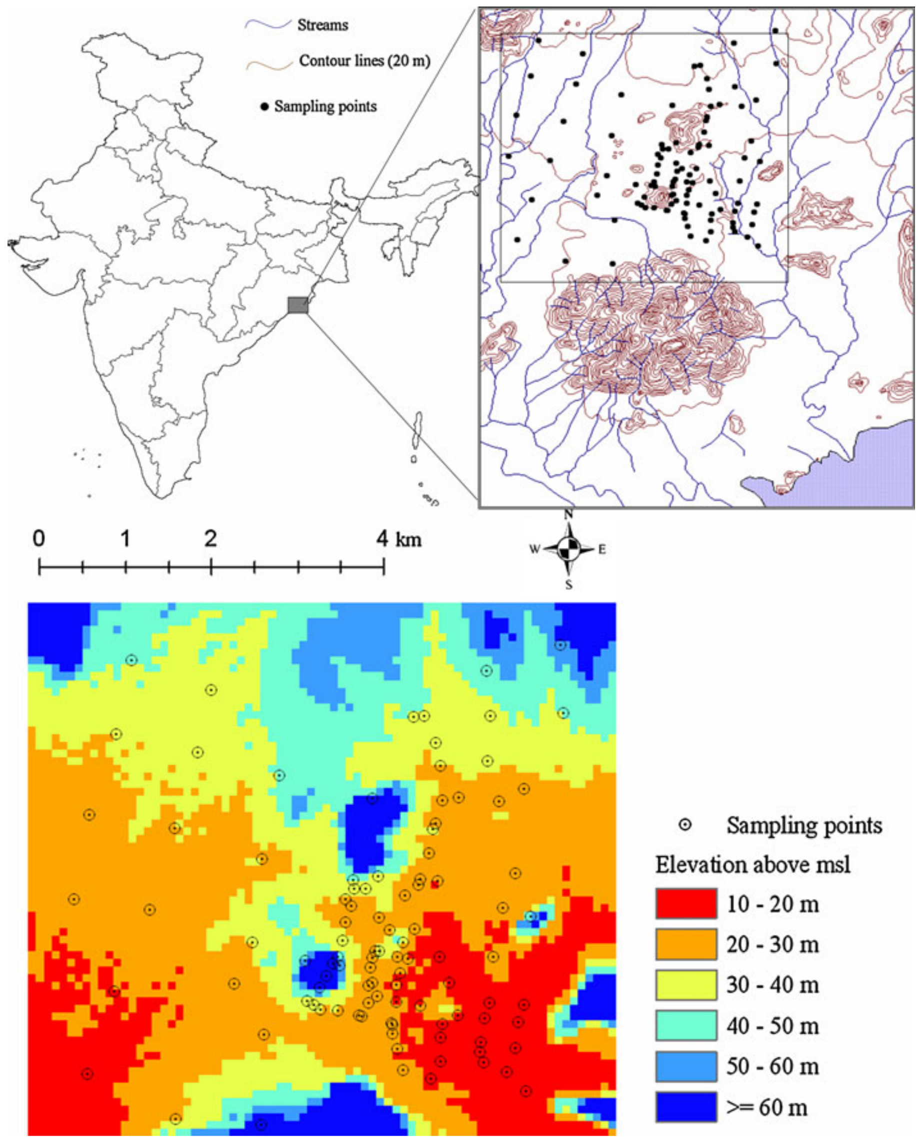

Fig. 1 Location of the Dengei Pahad Watershed (DPW) at the western catchment of Chilika Lake and the sampling points within the watershed

season from June to September. The area is a hilly terrain with the mean sea level varying from 5 to more than 451 m. The hills and isolated rocky knobs break the watershed into small but well-cultivated fields.

Soil sampling

A total of 100 surface soil samples was collected from the DPW from an area of 6 × 7 km. Out of all soil samples, 23 samples were collected on a 200 m × 600 m grid during first sampling campaign and rest of the samples were collected on an average grid size of 1.5 km × 1.5 km. Geographical coordinates and elevation of each sampling location was recorded using a handheld global positioning system (GPS). A few locations were also cross-checked with a differential GPS (DGPS). During each sampling, undisturbed soil cores (5 cm internal diameter and 6 cm long) and approximately 500 g of surface soil (~10 cm deep top soil) were collected from each site. After sampling, soil cores were carried to the laboratory and stored in a refrigerator and the loose soil was air dried, ground, and passed through a 2 mm mesh sieve. These sieved soils were analyzed for different soil properties. The digital elevation model of the DPW was extracted from the shuttle radar topography mission (SRTM) data of the area, which was downloaded from http://srtm.csi.cgiar.org/.

Measurement of soil properties

Soil particle size distribution, pH, water retention behavior (ψ–θ), and saturated hydraulic conductivity (K_s) of collected soil samples were determined in the laboratory using standard procedures (Tables 1, 2). Pressure head-water content (ψ–θ) data of each soil sample was fitted to the van

Table 1 Laboratory determination of soil properties of the collected soil samples from Dengei Pahad watershed (DPW)

Soil properties	Methodology	Reference
Soil particle size distribution in three size fractions: sand, silt and clay	International pipette method	Gee and Bauder (1986)
Soil pH	Measuring pH of soil water slurry (1:2.5)	–
Soil water retention at potential from 10 to 80 kPa	Tempe cell apparatus connected to a pressure manifold	Klute (1986)
Saturated hydraulic conductivity	Constant head permeameter (Eijelkamp Agrisearch Equipment, Netherlands)	Klute and Dirrksen (1986)

Table 2 Descriptive statistics of measured soil properties from Dengei Pahad watershed (DPW)

Soil properties[a]	Min	Max	Mean	σ[b]	CV[c]	Skewness	Kurtosis
Sand	7.85	84.60	48.71	17.80	36.53	−0.11	−0.72
Clay	4.60	63.97	34.99	14.93	42.71	0.00	−0.99
pH	3.44	8.72	6.83	1.20	19.14	−0.63	−0.42
$\ln(K_s)$	−0.82	6.26	3.09	1.77	53.21	−0.28	−0.58
$\ln(\alpha)$	−5.81	1.51	−2.24	1.48	63.72	−0.04	−0.13
n	1.03	1.54	1.23	0.09	34.53	0.76	0.98

Clay, clay content (%); pH, soil reaction; $\ln(K_s)$, natural log of saturated hydraulic conductivity (K_s, cm day^{-1}); $\ln(\alpha)$, natural log of van Genuchten parameter, (α, cm^{-1}); n, van Genuchten parameter

[a] Sand, sand content (%)

[b] σ, Standard deviation

[c] CV, Coefficient of variation calculated as (σ/mean) × 100

Genuchten (VG) water retention model (van Genuchten 1980):

$$S_e = \left(\frac{\theta - \theta_r}{\theta_s - \theta_r}\right) = \left[\frac{1}{1 + (\alpha|\psi|)^n}\right]^{1-\frac{1}{n}} \quad (1)$$

where S_e is the relative saturation, θ_r is the residual soil water content (cm^3 cm^{-3}), θ_s is the saturated soil water content (cm^3 cm^{-3}), ψ is the matric potential head (cm) and α (cm^{-1}) and n are shape parameters of the water retention curve. Measured water retention data were fitted in a VG model through optimization of θ_r, α, and n for each soil sample using the Solver function of MS-EXCEL spreadsheet program. Parameter θ_s was equated with soil porosity and was provided as a constant to the VG model.

Identification of trend in spatial data

The presence of trend in spatial data was identified through modeling the trend in the following linear functions:

$$z(x, y) = \beta_0 + \beta_1 \times x + \beta_2 \times y \quad (2)$$

$$z(x, y) = \beta_0 + \beta_1 \times t \quad (3)$$

where $z(x, y)$ is the soil property at a spatial coordinate (x, y), and β_0, β_1 are coefficients of trend function, and t is the external variable. The coefficients of this trend model were computed through least square fitting and were further used to formulate the trend design matrix in RK and REML-EBLUP.

Spatial interpolation of soil properties

Three spatial interpolation methods were evaluated for preparation of surface map of soil properties within the sampling area of the DPW: OK, RK, and REML-EBLUP. The OK, RK, and REML-EBLUP methods were executed using Matlab codes available from the personal website of Dr. B. Minasny (http://www.usyd.edu.au/su/agric/acpa/software). Performance of these three interpolation methods was evaluated in the spatial dataset from DPW to explore whether the REML-EBLUP approach performs better than OK and RK in a watershed scale.

In the OK and RK approaches, experimental semivariogram values were calculated from directly measured (hereinafter referred as to original data) and de-trended data (hereinafter referred as to residuals), respectively, using MoM approach (Matheron 1965):

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \{z(x_i) - z(x_i + h)\}^2 \quad (4)$$

where, $N(h)$ is the number of sample pairs at a given lag h, $z(x_i)$ and $z(x_i + h)$ are the measured values of the random variable z at place x_i and place separated by h from x_i, respectively. During calculation of experimental semivariogram values, maximum separation distance was fixed as half of the extent of the sampling area. The calculated experimental semivariogram values were then fitted in a standard semivariogram model. In the OK approach, the fitted semivariogram parameters from the original data were used for prediction of soil properties at unsampled locations. In the RK approach, residuals were kriged first followed by summing the kriging results with output from trend function, which is generally known as the RK of type C (Odeh et al. 1995). In case of REML-EBLUP approach, semivariogram parameters were estimated directly from the data through REML approach by minimizing the negative log-likelihood function and the detailed equations are given in Minasny and McBratney (2007). The fundamental assumption of the REML approach is that the spatial data follows a multivariate Gaussian distribution with the joint probability density function (pdf) of the measurements:

$$p(z/\beta,\theta) = (2\pi)^{-\frac{n}{2}}|K|^{\frac{1}{2}}\exp\left\{-\frac{1}{2}(z - M\beta)^T K^{-1}(z - M\beta)\right\} \quad (5)$$

where, z is a vector that contains n data, θ contains the parameters of the covariance matrix, K $(n \times n)$ is the variance–covariance matrix, and Mβ represents the trend (M is the design matrix for trend and β is the coefficients). In the REML approach, semivariogram parameters were estimated according to Matern model:

$$\gamma(h) = c_0 + c_1\left(1 - \frac{1}{2^{v-1}\Gamma(v)}\left(\frac{h}{r}\right)^v K_v\left(\frac{h}{r}\right)\right) \quad (6)$$

where, K_v is a modified Bessel function of the second kind of order v (Abramowitz and Stegun 1972), Γ is the gamma function, and v is smoothness parameter ($v > 0$). The REML estimated semivariogram parameters were then used to predict soil properties at unsampled locations using BLUP method.

Comparison of spatial interpolation methods

Each spatial interpolation method was evaluated through a leave-one-out cross-validation approach (Davis 1987). The performance of each spatial interpolation method was assessed using the root-mean-squared residual (RMSR) and mean squared deviation ratio (MSDR):

$$\mathrm{RMSR} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}[z(x_i) - \hat{z}(x_i)]^2} \quad (7)$$

$$\mathrm{MSDR} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\{z(x_i) - \hat{z}(x_i)\}^2}{\sigma_i^2}\right) \quad (8)$$

where $z(x_i)$ is the observed values of the variable at the location x_i, $\hat{z}(x_i)$ is the predicted values with variance σ_i^2 at the location x_i, and n is the number of sampling location. The RMSR estimates the accuracy of prediction (e.g., larger RMSR values indicate less accuracy of prediction). The MSDR measures the goodness of fit of the theoretical estimate of error (Bishop and Lark 2008). If the correct semivariogram model is used, the MSDR values should be close to 1 (Lark 2000; Minasny and McBratney 2005; Kerry and Oliver 2007).

Effect of sample number

A few recent reports stated that REML-EBLUP method is best suited for limited spatial data (Lark et al. 2006; Minasny and McBratney 2007). In the present study the spatial data consisting of 100 sampling locations in a watershed may be considered as limited data set. Since, sampling in a watershed is tedious even to collect 100 samples; we wanted to test the performance of OK, RK, and REML-EBLUP with spatial data smaller than 100 samples. For this purpose, two subsets from the full dataset of 100 (hereinafter referred as S_1) were created in two subsequent steps through random selection of sampling points. The criteria followed during random selection was to maintain a specific minimum distance between sampling pairs, which is greater than that originally adopted in this study and thus resulted into spatial data sets with different sampling density. During creation of the first subset, minimum separation distance between samples was considered as 250 m, which resulted in 78 soil samples (hereinafter referred as S_2). During creation of the second subset, minimum separation distance between samples was considered as 500 m, which resulted in 54 soil samples (hereinafter referred as S_3). The sampling locations for the dataset S_1, S_2, and S_3 are depicted in Fig. 1. The performance of OK, RK, and REML-EBLUP were tested with the spatial data S_1, S_2, and S_3 for spatial prediction of different soil properties.

Results and discussions

Descriptive statistics of spatial data

The average sampling density for this study was 2.4 km^{-2}. The average sampling distance between sampling pairs was 2,213 m with a highest and lowest separation distance of 26 and 7,423 m, respectively. Thus, the sampling strategy adopted in this study was considered as irregularly distributed samples.

Table 3 Semivariogram parameters of soil properties derived through method of moment (MoM) approach on original variables and de-trended residuals and through residual maximum likelihood (REML) approach

Soil properties[a]	MoM estimated semivariogram parameters						REML estimated semivariogram parameters (Matérn model)			
	Original variables (exponential model)			Residuals (exponential model)						
	C_0	C_1	Range (m)	C_0	C_1	Range (m)	C_0	C_1	Range (m)	v
Sand	86.10	227.4	288	153.2	101.4	459	173.57	105.04	191	6
Clay	190.1	30.25	229	129.2	46.08	107	139.29	48.54	196	6
pH	0.813	8.894	22384	0.75	19.92	50000	0.98	0.99	640	2
Ln(K_s)	1.933	1.141	248	–	–	–	2.004	0.354	187	3
Ln(α)	0.583	1.665	232	0	2.048	132	0	2.106	131	0.5
n	0.002	0.009	1892	0.002	0.009	2184	0.001	0.008	527	0.5

–, Semivariogram model can not be fitted for this property; C_0, nugget; C_1, Sill; v, smoothness parameter of the Matérn model; Clay, clay content (%); pH, soil pH; ln(K_s), natural log of saturated hydraulic conductivity (K_s) (cm day^{-1}); ln(α), natural log of van Genuchten parameter, α (cm^{-1}); n, van Genuchten parameter

[a] Sand = sand content (%)

Different physical and chemical properties of soils collected from the DPW watershed are summarized in Table 2. Soils at this site had wide variation in physical and chemical properties as is expected in a hill slope catena. Sand and clay contents varied widely within the watershed. Sand contents were observed to increase with elevation and the reverse was true for clay contents. A total of seven different textural classes was observed in this watershed although the majority of soil samples had either light clay texture having clay content >25% ($N = 36$) or heavy clay texture having clay content >45% ($N = 25$) indicating the presence of fine-textured soils in the study area. Soils were generally acidic with pH values being as low as 3.44 because of the presence of Fe- and Al-oxides. Saturated hydraulic conductivity also showed wide variation ($K_s > 1{,}000$ cm day^{-1} for coarse textured soils and $K_s < 0.1$ cm day^{-1} for heavily textured soils). The van Genuchten parameter, α was log-normally distributed with a mean value of -2.26. The mean value for the other van Genuchten parameter, n was 1.22. The Kolmogorov–Smirnov test revealed that except for EC, K_s, and α, all other soil properties were normally distributed at 5% level of significance.

Particle density (ρ_p), bulk density (ρ_b) and organic carbon content of DPW were also measured as auxiliary variables. Soil ρ_p ranged from 2.40 to 3.00 g cm^{-3} with the mean ρ_p (2.55 g cm^{-3}) less than the particle density (2.65 g cm^{-3}) generally assumed for mineral soils. Similarly, the mean ρ_b was 1.51 g cm^{-3} with the range between 1.80 and 2.00 g cm^{-3} for the red-colored soils of high-elevation areas and pastures. Very loose soils of sandy clay loam texture with ρ_b as low as 1.00–1.20 g cm^{-3} were also observed for few cultivated areas. Mean OC content of the soil samples was high (>0.75%) as per the rating for Indian soils (ICAR 2006).

Trend in spatial data

The directional trend due to easting and northing was found negligible in the present dataset according to Eq. (2). Log transformed elevation (m) had significant trend on most soil properties. Sand content, clay content, pH, and ln(K_s) specifically showed significant trend. Such trend was mainly due to the undulating topography which finally affected the soil forming processes. The sampling locations with high elevations were associated with high sand content whereas locations with low elevations had high clay content, and pH. Eroded finer soil particles from surface soil were carried by runoff water and deposited in low-lying areas, which resulted in high clay content at low elevation. Soils at high-elevation areas of the watershed were rich in sesquioxide (Fe- and Al-oxides), which resulted into low pH for those areas.

Semivariogram parameters of soil properties

The semivariogram parameters of soil properties computed through MoM approach and REML approach are presented in Table 3. The range parameter of most soil properties was found in between 200 and 300 m except for pH and n. Similar to original variables, the range parameter for de-trended data on pH and n was quite high. Here it is noted that fitting semivariogram models for de-trended data on ln(K_s) was not possible. The REML estimated Matérn semivariogram parameters were quite similar with the semivariogram parameters of residuals for most soil properties. The v parameter of Matérn model was 6 for sand and clay content, which indicated the smooth spatial variation of these two soil properties.

Table 4 Root-mean-squared residual (RMSR) and mean squared deviation ratio (MSDR) of predicted soil properties through ordinary kriging (OK), regression kriging (RK), and residual maximum likelihood estimation of covariance parameter followed by empirical best linear unbiased prediction (REML-EBLUP)

Soil properties[a]	Root-mean-squared residual (RMSR)			Mean squared deviation ratio (MSDR)		
	OK	RK	REML-EBLUP	OK	RK	REML-EBLUP
Sand	15.62	15.07	14.43	1.045	0.991	0.964
Clay	14.18	13.43	12.45	0.915	1.009	0.960
pH	1.12	1.11	1.11	1.159	1.173	0.997
Ln(K_s)	1.57	1.54	1.48	0.888	0.982	0.966
Ln(α)	1.37	1.40	1.37	1.068	1.005	0.988
n	0.08	0.08	0.08	1.606	1.526	1.051

Clay, clay content (%); pH, soil pH; ln(K_s), natural log of saturated hydraulic conductivity (K_s) (cm day^{-1}); ln(α), natural log of van Genuchten parameter, α (cm^{-1}); *n*, van Genuchten parameter

[a] Sand, sand content (%)

Comparisons of spatial interpolation methods

The RMSR of predicted soil properties through OK, RK, and REML-EBLUP is presented in Table 4. The REML-EBLUP approach showed smallest RMSR values for all soil properties. This was expected because REML is statistically sound and we had further combined this approach with the Matérn model, which was considered as a more flexible semivariogram model than the other standard models. The OK and RK methods were next in performance to the REML-EBLUP method. The RMSR values from Table 4 revealed that REML-EBLUP was 4.41% better than OK and 2.93% better than RK. Almost similar magnitude of improvement (about 4–6%) by REML-EBLUP approach over OK was reported by Chai et al. (2008).

The goodness of fit of the prediction error in terms of MSDR values for OK, RK, and REML-EBLUP approach are presented in Table 4. The REML-EBLUP approach showed the MSDR values very near to 1 for all soil properties. The deviation of MSDR values from 1 was more for OK and RK approaches than for REML-EBLUP approach. Therefore, the REML-EBLUP approach was revealed as the more reliable method than OK and RK. The MSDR values for REML-EBLUP approach, which are presented in Table 4 also matches with the previously reported values (Minasny and McBratney 2007; Kerry and Oliver 2007; Chai et al. 2008). The cross-validation results revealed that the REML-EBLUP approach may be preferred over RK and OK for spatial prediction of soil properties in a watershed scale.

Sample number versus prediction performance of REML-EBLUP

In the above discussions, the performance of REML-EBLUP was shown better than OK and RK in predicting soil properties within the DPW. This improved performance of REML-EBLUP was obtained from a spatial data with $N = 100$ from a watershed of 42 km^2 area, which was tedious to generate through sample collection followed by laboratory analysis. Moreover, the improvement of REML-EBLUP over OK and RK was around only 4%. Therefore, we wanted to explore the capability of REML-EBLUP even with less number of soil samples (<100), where OK or RK may not be a suitable option (Webster and Oliver 1992; Kerry and Oliver 2007). The comparative performance of REML-EBLUP, OK, and RK in two smaller datasets, S_1 ($N = 78$) and S_2 ($N = 54$) was therefore further tested. With reduction in sample number, the predictive performance of REML-EBLUP was better than OK and RK across all soil properties (Fig. 2a). With the spatial dataset S_2 ($N = 78$), the REML-EBLUP approach was 6.56% and 3.72% better than OK and RK, respectively. With further decrease in sample number to $N = 54$ (S_3), the improvement of REML-EBLUP over OK and RK approach was 15.48% and 5.55%, respectively. The improvement of REML-EBLUP over OK was found quite significant but not so much over RK. This is because of the presence of trend in the present spatial dataset and RK approach generally performs better in case of spatial data with trend. It is also hereby noted that decrease in sample number from 100 to 54 had resulted in average increase of RMSR by 19.27 and 7.28% for OK and RK, respectively, whereas for REML-EBLUP it was an average decrease by 4.25%. Besides the improvement in RMSR values, the deviation of MSDR from 1 was also less for REML-EBLUP than OK and RK (Fig. 2b). With the spatial dataset S_1 ($N = 100$), the average deviation of MSDR from 1 was 0.03 for REML-EBLUP, whereas for OK and RK, it was 0.18 and 0.12, respectively. When the sample number was decreased to $N = 54$, the average deviation of MSDR from 1 was 0.04 for REML-EBLUP, and was quite high for OK and RK (0.23 and 0.28, respectively). These results suggested that REML-EBLUP approach may effectively be

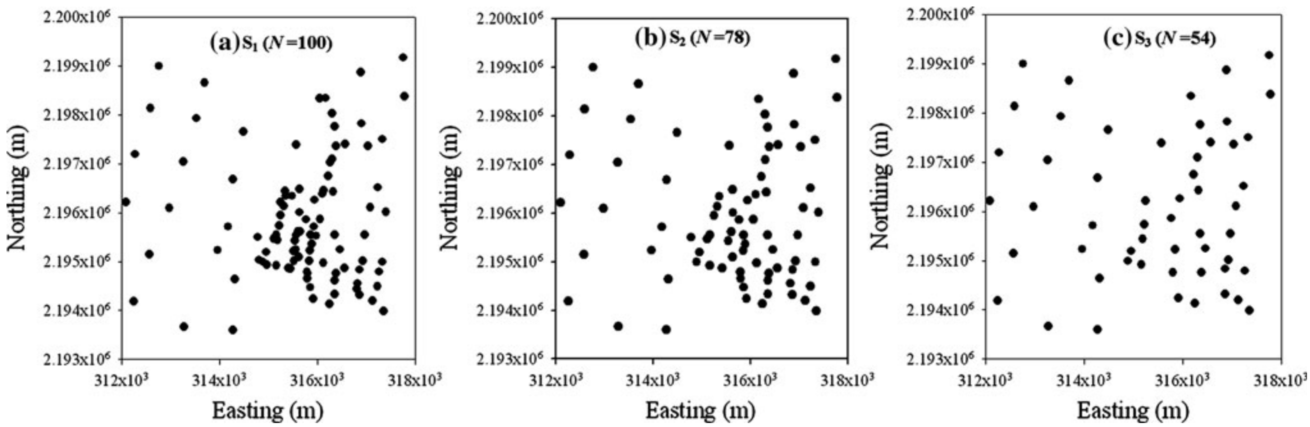

Fig. 2 Sampling points with different simulated grid sizes **a** S_1 = sampling points with all 100 soil samples, **b** S_2 = sampling points with assumption of sampling in a minimum of 250 m sampling grid ($N = 78$); **c** S_3 = sampling points with assumption of sampling in a minimum of 500 m sampling grid ($N = 54$)

implemented in watershed scale with minimum sampling separation distance of 500 m, which is equivalent to maximum sampling density of 4 km^{-2}.

Surface map of soil properties

Evaluation of OK, RK, and REML-EBLUP in the present spatial data from a hilly watershed showed REML-EBLUP as the best method. Therefore, surface map of each soil property within the sampling area of the DPW was prepared through REML-EBLUP approach and is presented in Figs. 3 and 4. Spatial pattern of sand and clay content within the sampling area was closely related with elevation grid. The top soil from high-elevation areas, especially in the northern part, was eroded by water and thus exposed the coarse-textured subsurface layer, which resulted in higher sand content for those areas. Reversely, the deposition of eroded finer particles in low-lying areas resulted in higher clay content. Soil pH varied greatly within the sampling area and ranged from 4.5 to 8.1. Saturated hydraulic conductivity (K_s) within the sampling area ranged from as low as ~1.5 cm day^{-1} to as high as ~640 cm day^{-1}. The low-lying areas at the southeast corner were high in clay content and low in sand content and thus resulted in very low K_s (<10 cm day^{-1}). On the contrary, rocky outcrops and coarse-textured soil in hilly areas, which mostly occur in the northern parts of the sampling area, were high in K_s (>50 cm day^{-1}). The VG parameter, α, which is the inverse of air entry potential of soil, showed erratic spatial distribution within the sampling area as it was influenced by several basic soil properties. The VG parameter n, which indicates the steepness of water release curve upon drying from its saturation, ranged from 1.103 to 1.449 within the sampling area. The higher the value of n, the steeper the curve, indicating quicker release of water from soil pore spaces after applying a little amount of suction.

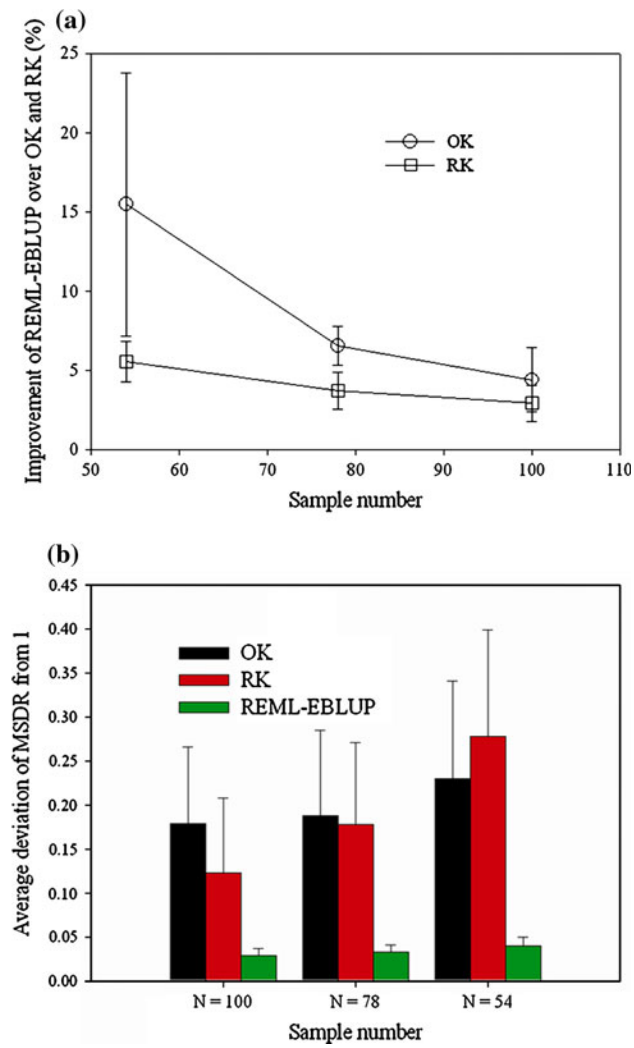

Fig. 3 Effect of sample number on prediction performance of ordinary kriging, regression kriging, and REML-EBLUP; **a** root mean-squared residual (RMSR) vs sample number, **b** mean squared deviation ratio (MSDR) versus sample number

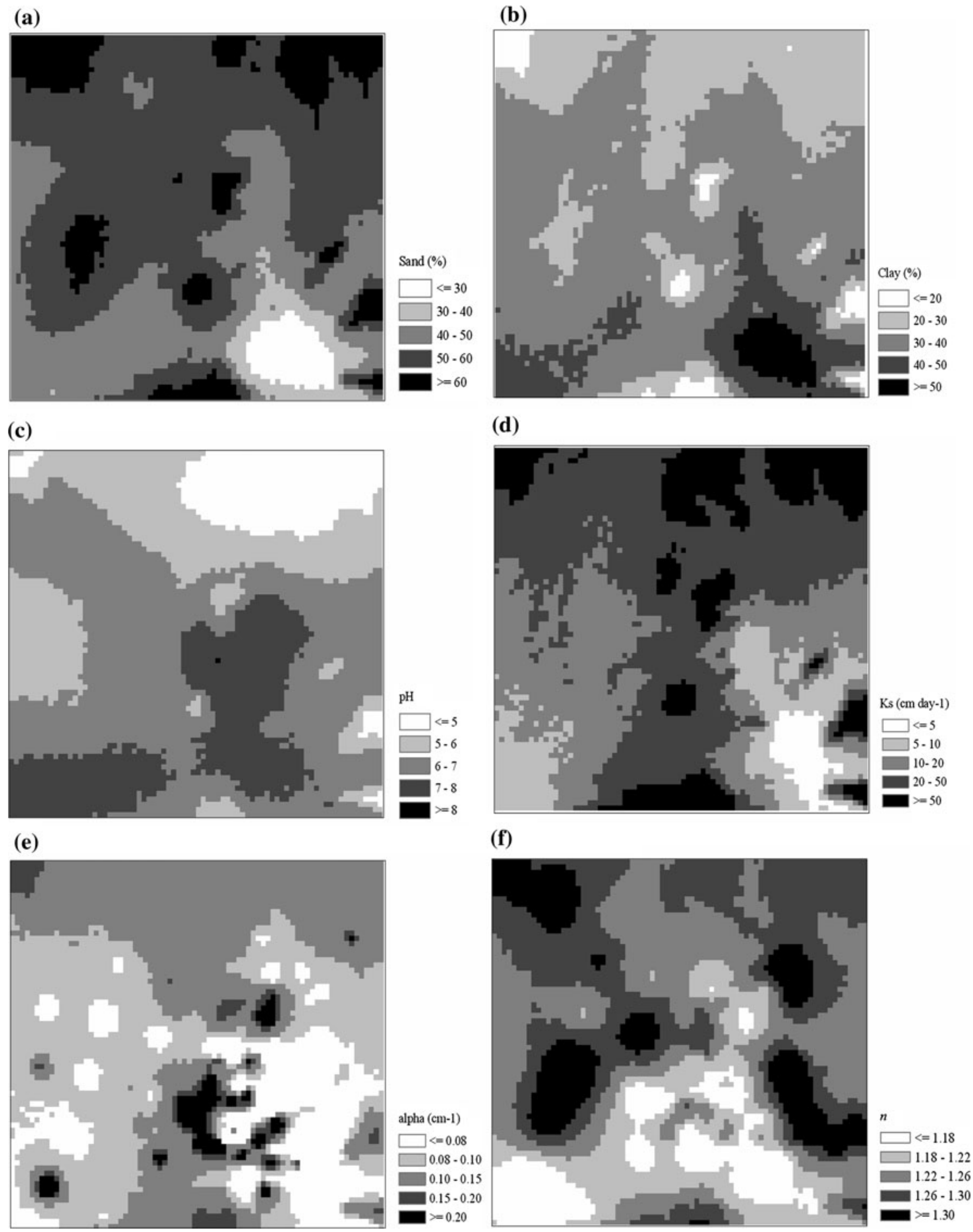

Fig. 4 Surface map of soil properties within the sampling area of Dengei Pahad Watershed (DPW) as predicted through REML-EBLUP approach on spatial data S_1 with sample number 100

Summary and conclusion

The REML-EBLUP approach was applied to a sparse dataset ($N = 100$) collected over an irregular grid from the DPW (area = 42 km^2) in the Catchment area of Chilika Lake, India, for generating surface maps of different soil physical and physicochemical properties. Wide variation in soil properties was observed in the dataset due to undulating topography within the sampling area. Removal of surface soils through runoff and deposition of fine clays at depression areas in the watershed is a regular process during monsoon season (June–September). Hence, the trend of elevation on most soil properties was observed in the dataset. Previous literatures had shown the REML-EBLUP approach as a potential method to interpolate soil properties in a limited dataset with trend but mostly were reported from a small scale study. In the present study, the REML-EBLUP approach was tested in a microwatershed having three most frequently observed features of watershed-scale spatial data: irregular distribution of sampling points, limited number of data and the existence of strong spatial trends in the datasets. The performance of REML-EBLUP in the dataset was also compared with commonly used geostatistical methods such as RK and OK. The REML-EBLUP was found better than OK and RK by an average improvement of 4%. Even in the reduced dataset ($N = 54$), which was created through random removal of sampling points, the REML-EBLUP approach was found 15.48% better than OK and 5.55% better than RK. The uncertainty of prediction was also lower for REML-EBLUP approach than OK and RK. Therefore, it may be concluded that REML-EBLUP approach may successfully be used to generate surface map of soil properties in a watershed scale specifically when the availability of larger datasets is a difficult task. Results also show that the minimum separation distance in watershed-scale studies should be at least 500 m (equivalent to the sampling density of 4 km^{-2}) for spatial interpolation of soil properties through REML-EBLUP approach.

Acknowledgments The senior author acknowledges the Director, Central Arid Zone Research Institute, Jodhpur, India for granting permission for study leave to IIT, Kharagpur. The authors acknowledge the financial support provided by the Chilika Development Authority, Bhubaneswar, India for conducting a part of the study.

References

Abramowitz M, Stegun IA (1972) Handbook of mathematical functions with formulas, graphs, and mathematical tables. New York, Dover

Basaran M, Erpul G, Ozcan AU, Saygin DS, Kibar M, Bayramin I, Yilman FE (2010) Spatial information of soil hydraulic conductivity and performance of cokriging over kriging in a semiarid basin scale. Environmental Earth Science doi:10.1007/s12665-010-0753-6

Baskan O, Cebel H, Akgul S, Erpul G (2010) Conditional simulation of USLE/RUSLE soil erodibility factor by geostatistics in a Mediterannean catchment, Turkey. Environ Earth Sci 60:1179–1187

Baxter SJ, Oliver MA (2005) The spatial prediction of soil mineral N and potentially available N using elevation. Geoderma 128: 325–339

Bishop TFA, Lark RM (2008) Reply to “Standardized vs. customary ordinary cokriging…” by A. Papritz. Geoderma 146(1–2):397–399

Chai X, Shen C, Yuan X, Huang Y (2008) Spatial prediction of soil organic matter in the presence of different external trends with REML-EBLUP. Geoderma 148:159–166

Cressie N (1993) Statistics for spatial data. Wiley, New York

Davis BM (1987) Uses and abuses of cross-validation in geostatistics. Math Geol 19:241–248

Diggle PJ, Riberio PJ Jr, Christensen OF (2003) An introduction to model-based geostatistics. In: Moller J (ed) Spatial statistics and computational methods. Springer, New York, pp 43–86

Dong J, Yu M, Bian Z, Wang Y, Di C (2011) Geostatistical analysis of heavy metal distribution in reclaimed mine land in Xuzhou, China. Environ Earth Sci 62:127–137

Du Y, Delfs JO, Kalbus E, Park CH, Wang W, Kolditz O (2009) A regional hydrological soil model for large-scale applications: Computational concept and application. J Environ Hydrol 17: Paper No. 7, pp 16

Gee GW, Bauder JW (1986) Particle size analysis. In: Klute A (ed) Methods of soil analysis, Part 1, Agronomy Monograph No. 9. ASA and SSSA, Madison, pp 383–412

Goovaerts P (1992) Factorial kriging analysis: a useful tool for exploring the structure of multivariate spatial soil information. J Soil Sci 43:597–619

Goovaerts P (1994) Study of spatial relationships between two sets of variables using multivariate geostatistics. Geoderma 62:93–107

Goovaerts P (1997) Geostatistics for natural resources evaluation. Applied geostatistics series. Oxford University Press, New York

Goovaerts P (1998) Geostatistical tools for characterizing the spatial variability of microbiological and physico-chemical soil properties. Biol Fertil Soils 27:315–334

Goovaerts P (1999) Geostatistical in soil science: state-of-the-art and perspectives. Geoderma 89:1–45

Herbst M, Diekkrüger B, Vereecken H (2006) Geostatistical coregionalisation of soil hydraulic properties in a micro-scale catchment using terrain attributes. Geoderma 132:206–221

Iqbal J, Thomasson JA, Jenkins JN, Owens OR, Whisler FD (2005) Spatial variability analysis of soil physical properties of alluvial soils. Soil Sci Soc Am J 69:1338–1350

Issaks EH, Srivastava RM (1989) Applied geostatistics. Oxford university press, New York

Juang KW, Chen YS, Lee DY (2004) Using sequential indicator simulation to asses the uncertainty of delineating heavy-metal contaminated soil. Environ Pollut 127:229–238

Kerry R, Oliver MA (2007) Comparing sampling needs for semivariograms of soil properties computed by the method of moments and residual maximum likelihood. Geoderma 140:383–396

Kitanidis PK (1983) Statistical estimation of polynomial generalized covariance functions and hydrologic applications. Water Resour Res 19:909–921

Klute A (1986) Water retention: laboratory methods. In: Klute A (ed) Methods of soil analysis, Part 1, Agronomy Monograph No. 9, 2nd edn. ASA and SSSA, Madison, pp 635–662

Klute A, Dirrksen C (1986) Hydrualic conductivity and diffusivity. In: Klute A (ed) Methods of soil analysis—part 1, physical and mineralogical methods. ASA and SSSA, Madison, pp 687–734

Lark RM (2000) Estimating semivariograms of soil properties by the method of moments and maximum likelihood. Eur J Soil Sci 51:717–728

Lark RM, Cullis BR, Welham SJ (2006) On spatial prediction of soil properties in the presence of a spatial trend: the empirical best linear unbiased predictor (E-BLUP) with REML. Eur J Soil Sci 57:787–799

Lin YP (2002) Multivariate geostatistical methods to identify and map spatial variations of soil heavy metals. Env Geol 42:1–10

Matérn B (1960) Spatial variation. In: Meddelanden fran Statens Skogsforskningsinstitut, 49, No. 5, 2nd edn. 1986, Lecture Notes in Statistics, No. 36. Springer, New York

Matheron G (1965) Les variables regionalisées et leur estimation. Masson, Paris

Minasny B, McBratney AB (2005) The Matérn function as a general model for soil semivariograms. Geoderma 128:192–207

Minasny B, McBratney AB (2007) Spatial prediction of soil properties using EBLUP with the Matérn covariance function. Geoderma 140:324–336

Odeh IOA, McBratney AB, Chittleborough DJ (1995) Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. Geoderma 67:215–226

Pardo-Igúzquiza E (1998) Inference of spatial indicator covariance parameters by maximum likelihood using MLREML. Comput Geosci 24:453–464

Patterson HD, Thompson R (1971) Recovery of inter-block information when block sizes are unequal. Biometrika 58:545–554

Rakhmatullaev S, Marache A, Huneau F, Coustumer PL, Bakiev M, Motelica-Heino M (2010) Geostatistical approach for the assessment of the water reservoir capacity in arid regions: a case study of the Akdarya reservoir, Uzbekistan. Environ Earth Sci. doi:10.1007/s12665-010-0711-3

Robinson TP, Metternicht G (2006) Testing the performance of spatial interpolation techniques for mapping soil properties. Comput Electron Agric 50:97–108

Santra P, Chopra UK, Chakraborty D (2008) Spatial variability of soil properties in agricultural farm and its application in predicting surface map of hydraulic property. Current Sci 95(7):937–945

Santra P, Das BS, Chakravarty D (2011) Delineation of hydrologically similar units in a watershed based on fuzzy classification of soil hydraulic properties. Hydrol Process 25:64–79

Simbahan GC, Dobermann A, Goovaerts P, Ping J, Haddix ML (2006) Fine-resolution mapping of soil organic carbon based on multivariate secondary data. Geoderma 132:471–489

Stein ML (1999) Interpolation of spatial data: some theory for kriging. Springer, New York

Uygur V, Irvem A, Karanlik S, Akis R (2010) Mapping of total nitrogen, available phosphorus and potassium in Amik plain, Turkey. Environ Earth Sci 59:1129–1138

van Genuchten MTh (1980) A close-form equation for predicting the hydraulic conductivity of unsaturated soils. Soil Sci Soc Am J 44:892–898

Webster R, Oliver MA (1992) Sample adequately to estimate semivariograms of soil properties. J Soil Sci 43:177–192

Wu C, Wu J, Luo Y, Zhang H, Teng Y, Degloria SD (2010) Spatial interpolation of severely skewed data with several peak values by the approach integrating kriging and triangular network interpolation. Environ Earth Sci. doi:10.1007/s122665-010-0784-z