Analyzing spatial data: An assessment of assumptions, new methods, and uncertainty using soil hydraulic data

B. Zimmermann,[1,2] E. Zehe,[1,3] N. K. Hartmann,[1] and H. Elsenbeer[1]

Received 21 October 2007; revised 8 April 2008; accepted 6 August 2008; published 16 October 2008.

[1] Environmental scientists today enjoy an ever-increasing array of geostatistical methods to analyze spatial data. Our objective was to evaluate several of these recent developments in terms of their applicability to real-world data sets of the soil field-saturated hydraulic conductivity (*Ks*). The intended synthesis comprises exploratory data analyses to check for Gaussian data distribution and stationarity; evaluation of robust variogram estimation requirements; estimation of the covariance parameters by least-squares procedures and (restricted) maximum likelihood; use of the Matérn correlation function. We furthermore discuss the spatial prediction uncertainty resulting from the different methods. The log-transformed data showed Gaussian uni- and bivariate distributions, and pronounced trends. Robust estimation techniques were not required, and anisotropic variation was not evident. Restricted maximum likelihood estimation versus the method-of-moments variogram of the residuals accounted for considerable differences in covariance parameters, whereas the Matérn and standard models gave very similar results. In the framework of spatial prediction, the parameter differences were mainly reflected in the spatial connectivity of the *Ks* field. Ignoring the trend component and an arbitrary use of robust estimators would have the most severe consequences in this respect. Our results highlight the superior importance of a thorough exploratory data analysis and proper variogram modeling, and prompt us to encourage restricted maximum likelihood estimation, which is accurate in estimating fixed and random effects.

Citation: Zimmermann, B., E. Zehe, N. K. Hartmann, and H. Elsenbeer (2008), Analyzing spatial data: An assessment of assumptions, new methods, and uncertainty using soil hydraulic data, *Water Resour. Res.*, *44*, W10408, doi:10.1029/2007WR006604.

1. Introduction

[2] Environmental scientists today enjoy an ever-increasing array of methods to choose from to analyze spatial data. New techniques are constantly developed that aim to combine physical laws, observations and process knowledge for modeling the variability of environmental variables, both in space and time [*Heuvelink and Webster*, 2001]. Geostatistics, which predominantly uses observations to describe the spatial variability of, for instance, soil properties, has come a long way since its introduction to soil science in the 1980s. Improvements relevant to soil hydrology include new variogram estimation techniques [*Pardo-Igúzquiza*, 1998; *Lark*, 2000b; *Minasny and McBratney*, 2005], robust estimation [*Cressie and Hawkins*, 1980; *Dowd*, 1984; *Genton*, 1998; *Lark*, 2000a], scale effects in variogram estimation [*Skøien and Bloeschl*, 2006], trend modeling [*Leuangthong and Deutsch*, 2004; *Lark et al.*, 2006], model selection [*Geiler et al.*, 1997; *Lark*, 2000a, 2000b], sample size requirements for estimating variograms [*Webster and Oliver*, 1992; *Gascuel-Odoux and Boivin*, 1994], and variogram uncertainty for the method-of-moments variogram [*Marchant and Lark*, 2004]. Although incomplete, this list hints at the considerable number of aspects the user of geostatistics, e.g., the soil hydrologist, should consider. By implication, this short list also hints at various pitfalls the unsuspecting user of geostatistics may encounter. To our knowledge, however, there has been no soil-hydrological study that evaluates some of those new developments within a geostatistical framework. This lack prompted us to attempt such an assessment, using real-world data sets of soil field-saturated hydraulic conductivity (*Ks* from here on). We measured *Ks* under different land covers to characterize its spatial and temporal variation in the tropical montane rainforest of southern Ecuador. For this study, we selected three data sets, which are composed of measurements at three soil depths under an old fallow regenerating from antecedent pasture use. Our assessment is based on answering the following research questions:

[3] 1. Do soil hydraulic data satisfy the assumption of stationarity of the mean?

[4] 2. Do we need robust estimation techniques?

[5] 3. Do we have to account for anisotropic variation?

[6] 4. In which manner do different variogram estimation techniques, e.g., restricted maximum likelihood versus method-of-moments estimation, contribute to variations in the estimated covariance parameters?

[7] 5. Should we use the flexible Matérn function, which helps to avoid misspecification of the covariance model?

[1]Institute of Geoecology, University of Potsdam, Potsdam, Germany.
[2]Now at Smithsonian Tropical Research Institute, Center for Tropical Forest Science, Balboa, Panama.
[3]Now at Institute of Water and Environment, Technische Universitaet Muenchen, Munich, Germany.

0043-1397/08/2007WR006604

Figure 1. Location of the study area.

[8] 6. How can we get a reasonable estimate of the nugget variance?

[9] 7. In which manner do the variations in the estimated covariance parameters (items 4–6) influence spatial predictions?

[10] 8. Does the existing literature regarding *Ks* spatial variability allow for any general statements about, for example, a dependence on certain environmental conditions?

[11] The answers to these questions form the basis for recommendations regarding the geostatistical analysis of soil hydraulic properties in particular, and environmental data in general.

2. Study Area

[12] Embedded in the Eastern Cordillera of the Andes of south Ecuador, our study area (Figure 1) is located in the Reserva Biósfera de San Francisco around the Estación Científica San Francisco (ECSF) (3°58′18″S, 79°4′45″W, 1860 m a.s.l.). Steep slopes (30–50°) covered with ''Lower Montane Rain Forest'', which gradually changes to ''Lower Montane Cloud Forest'' on higher ground [*Bruijnzeel and Hamilton*, 2000], characterize the north-facing slopes, where there are no signs of recent human intervention. In contrast, the south-facing slopes have been subject to human influence for decades.

[13] Mean annual precipitation (1998–2005) amounts to 2273 mm [*Rollenbeck et al.*, 2007], and mean annual air temperature (1999–2002) is 15.5 °C [*Motzer et al.*, 2005]. A drier period is between November and January, and May and June are the wettest months [*Motzer et al.*, 2005].

[14] The bedrock consists mainly of weakly metamorphosed Palaeozoic schists and sandstones with some quartz veins [*Wilcke et al.*, 2003], which belong to the Chiguinda unit (Mapa geológico del Ecuador; Instituto Geográfico de Militar y Ministerio de Energía y Minas). Soils are classified as Inceptisols and Histosols [*Soil Survey Staff*, 1999; *Schrumpf et al.*, 2001] with a large percentage of silt.

[15] For this study, we selected a plot of 50 by 70 meters in an old fallow situated at the south-facing valley side such that the x-coordinate approximately follows the contour line. Main elevation differences occur in main slope direction (i.e., in direction of increasing y-coordinate), but there is also a topographic gradient following the contour line, as the left side of the plot is located near a gully and the right side on a ridge. After a recovery of at least 10 years from antecedent pasture use, the fallow is covered by bracken (Pteridium aquilinum), which is very abundant in the pasture areas of this region, and succession vegetation. Among the most abundant succession plant families are *Orchidaceae*, *Asteraceae*, *Ericaceae*, *Melastomataceae*, *Poaceae*, *Rosaceae*, *Gleicheniaceae*, *Lycopodiaceae*, *Bromeliaceae*, and *Myrsinaceae*.

3. Methods

3.1. Field Measurements and Sampling Strategy

[16] We measured the field-saturated hydraulic conductivity in situ in the mineral soil at the depths of 12.5, 20 cm and 50 cm with an Amoozemeter, which is a Compact Constant-Head Permeameter [Ksat Inc., Raleigh; *Amoozegar*, 1989a]. The procedure required that we auger a cylindrical hole with radius r to the desired depth, establishing a constant head H such that $H/r \geq 5$, and monitoring the outflow from the device until a steady-state flow rate is attained; at which point *Ks* can be calculated via the Glover solution [*Amoozegar*, 1989b, 1993]. In a comparison of *Ks* computed in this manner with *Ks* based on different Alpha parameter values [*Elrick and Reynolds*, 1992], the Glover equation gave results that were comparable to the results obtained by the fixed Alpha value approach (A. Amoozegar, personal communication, 1992).

[17] We used a combined design- and model-based sampling design to satisfy the requirements both for the intra-site spatial analysis presented here, and for inter-site comparisons to be published elsewhere. First we chose the location of the measurement plot such that its x-axis corresponded to a contour line of the investigated hillslope (Figure 2). We then superimposed a grid consisting of grid cells of a size of 2 m^2; a total of 30 grid cells was then included in our sample using a random selection algorithm. Within each selected cell we placed five fixed measuring points (Figure 2). The spacing among those points was based upon experience from a former study in the Brazilian Amazon, where spatial patterns of *Ks* emerged only at a high sampling resolution (i.e., short lag distances) [*Sobieraj et al.*, 2004]. Hence we emphasized small separation distances, starting with the smallest possible point distance of 0.25 m (Figure 2). This distance emerged from field trials which showed that yet smaller distances would cause prior measurements to interfere with subsequent ones.

[18] The described procedure resulted in an irregular sampling scheme with a sample size of 150 per soil depth, and somewhat less at 50 cm depth where augering was not possible in all places because of a large percentage of stones. We chose this sample size in reference to *Webster and Oliver* [1992] and *Gascuel-Odoux and Boivin* [1994] who recommended to sample at some 150 to 200 points in an area to estimate the variogram of that area.

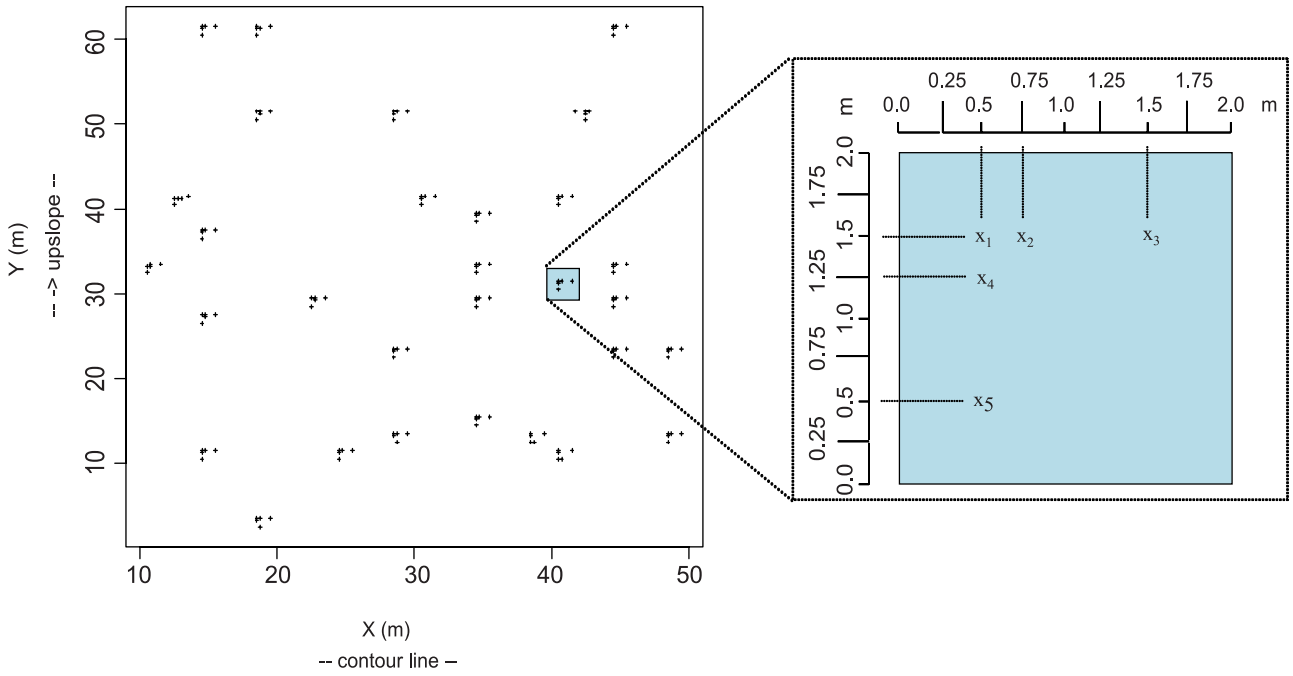

Figure 2. Sampling grid.

3.2. Data Analysis

[19] For all data analysis except robust variogram estimation we used the language and environment of *R*, version 2.2.1. [*R Development Core Team*, 2004]; many of the geostatistical methods were implemented in the libraries geoR [*Ribeiro and Diggle*, 2001] and gstat [*Pebesma*, 2004]. We used the Fortran-code of *Rousseeuw and Croux* [1993] to calculate experimental variograms with the estimator proposed by *Genton* [1998].

3.2.1. Exploratory Data Analysis

[20] We first checked the univariate distribution of the *Ks* data by means of diagnostic plots (boxplots, quantile-quantile plots, histograms), and used the Box-Cox transformation [*Box and Cox*, 1964] to find the most appropriate transformation for achieving Gaussian behavior. We also produced **h**-scattergrams [*Webster and Oliver*, 2001], which are scatterplots of point pairs separated by a fixed distance to scrutinize the data for outlying values.

[21] We used diagnostic plots (plots of the spatial data, which were divided into quintiles; plots of the data versus the coordinates) to explore the data for non-stationarity of the mean that may be caused by local trends according to

$$z(\mathbf{x}) = \mu(\mathbf{x}) + \varepsilon(\mathbf{x}), \tag{1}$$

where $z(\mathbf{x})$ is the observed variable at location $\mathbf{x}$, $\mu(\mathbf{x})$ is the local mean, i.e., it represents a deterministic drift of the variable at location $\mathbf{x}$, and $\varepsilon(\mathbf{x})$ is the random component at location $\mathbf{x}$ that should be normally distributed with zero mean and that satisfies the second-order stationarity required for the geostatistical analysis. Hence we calculated the residual values $\varepsilon(\mathbf{x})$ at every location x and used them for the subsequent geostatistical analyses of the method-of-moments variogram if the F-test [*Fisher*, 1972] showed that the trend coefficients were significantly different from 0 ($\alpha = 0.05$).

3.2.2. Geostatistical Analysis

3.2.2.1. Experimental Variogram Estimation

[22] As a first step in the variogram analysis, we calculated isotropic experimental variograms using the "classical" variogram estimator of *Matheron* [1962]

$$2\hat{\gamma}_M(\mathbf{h}) = \frac{1}{N(\mathbf{h})}\sum_{i=1}^{N(\mathbf{h})}\{z(\mathbf{x}_i) - z(\mathbf{x}_i+\mathbf{h})\}^2, \tag{2}$$

where $z(\mathbf{x}_i)$ is the observed value at location $\mathbf{x}_i$, $N(\mathbf{h})$ are the pairs of observations that are separated by lag $\mathbf{h}$.

[23] The inspection of diagnostic plots cannot totally rule out the presence of extreme values, because data can qualify as "spatial outliers" when compared to their close neighbors even though they do not appear unusual in a histogram of the whole data set [*Lark*, 2002]. Hence we also produced robust experimental variograms with estimators proposed by *Cressie and Hawkins* [1980], *Dowd* [1984], and *Genton* [1998]. The Cressie-Hawkins' estimator is given by:

$$2\hat{\gamma}_{CH}(\mathbf{h}) = \frac{\left\{\frac{1}{N(\mathbf{h})}\sum_{i=1}^{N(\mathbf{h})}\left|z(\mathbf{x}_i - z(\mathbf{x}_i+\mathbf{h})\right|^{\frac{1}{2}}\right\}^4}{0.457 + \frac{0.494}{N(\mathbf{h})} + \frac{0.045}{N^2(\mathbf{h})}}. \tag{3}$$

[24] The Dowd estimator is

$$2\hat{\gamma}_D(\mathbf{h}) = 2.198\{median(|y_i(\mathbf{h})|)\}^2, \tag{4}$$

where $y_i(\mathbf{h}) = z(\mathbf{x}_i) - z(\mathbf{x}_i + \mathbf{h})$, i = 1,2,..., $N(\mathbf{h})$; and Genton's estimator is given by

$$2\hat{\gamma}_G(\mathbf{h}) = \left(2.219\{|y_i(\mathbf{h}) - y_j(\mathbf{h})|; i<j\}_{\binom{H}{2}}\right)^2, \tag{5}$$

with $y_i(\mathbf{h})$ defined as for equation (4) and H = integer part $(n/2) + 1$, $n = N(\mathbf{h})$.

[25] For an in-depth discussion of the efficiency and robustness of robust estimators we refer to *Lark* [2000a].

[26] Since our sampling was highly irregular we had to define lag classes, which are groups of different individual lag distances between point pairs, such that every comparison constitutes one estimate only; i.e., every possible point pair is used in one particular lag class only. This procedure is somewhat arbitrary, so we used different lag classes to compute the experimental variograms, and we then compared their influence on the theoretical variogram models. On the one hand, we calculated experimental variograms using lags in accordance with the unsystematic sampling design. That is to say, the lag classes mirror the sampling emphasis on small separation distances and hence, are narrower for small compared to large lags (''unevenly spaced experimental variograms''). On the other hand, we chose 25-cm lags if the minimum number of point pairs was at least 30 (''evenly spaced experimental variograms'').

[27] We also checked our data for anisotropy by calculating the experimental variograms in four different directions, one representing the contour line (the x-coordinate, see Figure 2), one the slope (the y-coordinate, see Figure 2) and the other two are in between those two extremes, i.e., in 45° and 135°, respectively.

3.2.2.2. Estimation of Covariance Parameters

[28] To determine the need for robust estimation techniques, we fitted three standard theoretical variogram models (exponential, Gaussian, spherical) to the unevenly spaced experimental variograms derived from the ''classical'' Matheron and the three robust estimators by ordinary least squares. The ''best'' model was chosen by the minimum sum of squares from the fit, which is a conventional measure for the goodness-of-fit in a least-squares fitting procedure. We then adopted the approach of *Lark* [2000a] who compared the variograms derived from all above-mentioned estimators by cross-validation or with a validation subset using a statistics $\theta(\mathbf{x})$ defined as

$$\theta(\mathbf{x}) = \frac{\left\{z(\mathbf{x}) - \hat{Z}(\mathbf{x})\right\}^2}{\sigma^2_{\mathbf{K},\mathbf{x}}}, \tag{6}$$

where $z(\mathbf{x})$ is the observed value at location $\mathbf{x}$, $\hat{Z}(\mathbf{x})$ is the kriged estimate and $\sigma^2_{\mathbf{K},\mathbf{x}}$ the kriging variance. If kriging errors follow a Gaussian distribution, $\theta(\mathbf{x})$ will be distributed as χ^2 with one degree of freedom. Since the median of the standard χ^2 distribution with one degree of freedom is 0.455, the median of $\theta(\mathbf{x})$ is also 0.455 when a correct variogram is used to interpolate intrinsic data. A sample median significantly less than 0.455 suggests that kriging overestimates the variance whereas one which is greater than 0.455 underestimates the variance. In the former case, this may be due to the effects of outliers on the variogram, and in the latter case due to the effect of non-normality on a robust estimator. In order to compute confidence limits for the median of $\theta(\mathbf{x})$, *Lark* [2000a] quotes standard texts for the distribution of the sample median of a large sample of $2n + 1$ data, which are random variables of median $\tilde{y}$ and probability density function $f(\tilde{y})$; hence, $\tilde{y}$ is a normally distributed variable drawn from a population with variance $\sigma^2_{\tilde{y}}$ where

$$\sigma^2_{\tilde{y}} = \frac{1}{8n * f(\tilde{y})^2}. \tag{7}$$

[29] In the framework of the statistic $\theta(\mathbf{x})$, $f(\tilde{y})$ is the pdf of the χ^2 distribution with 1 d.f., the median of which is 0.455.

[30] At this point the confidence limits can be computed by

$$0.455 \pm 1.96 * \sqrt{\sigma^2_{\tilde{y}}}. \tag{8}$$

[31] We used cross-validation to compare the classical and robust variogram estimators. For the soil depths of 12.5 and 20 cm (n = 150), the upper and lower 95% confidence limits for the median of $\theta(\mathbf{x})$ were calculated as 0.286 and 0.624, respectively; for the 50 cm depth (n = 105) they were 0.251 and 0.659, respectively. *Lark* [2000a] proposed to select the estimator with the median value of $\theta(\mathbf{x})$ closest to 0.455, if the median of $\theta(\mathbf{x})$ indicates that Matheron's estimator (which otherwise is the estimator of choice because of its efficiency) is significantly influenced by outliers. If all robust estimates are similar, then he recommended Genton's, because of its efficiency.

[32] If the median of $\theta(\mathbf{x})$ ruled out the presence of influential extreme values and hence supported using the Matheron estimator, we compared two lines of geostatistical analysis, one of which is based on fitting a theoretical model to an experimental variogram whereas the other uses maximum likelihood, which estimates the covariance parameters directly from the data.

[33] In the former case, we fitted the model by least squares (LS) to the unevenly spaced experimental variogram both over the whole (abbr. LS1) and over half the separation distance (abbr. LS2), and to the evenly spaced experimental variogram (abbr. LS3); in case of significant trend coefficients we used the residuals of the regression model for the calculation of the experimental variograms. We again used the above-mentioned standard variogram models (exponential, Gaussian, spherical) and goodness-of-fit criterion, and two different weights: equal weights (ordinary least squares, abbr. OLS), and weights which correspond to the number of point pairs in each lag class (weighted least squares, abbr. WLS).

[34] The principle of maximum likelihood estimation of variogram parameters is as follows: n observed data are assumed to be from a multivariate Gaussian distribution with a mean vector $\mathbf{m}$ of length n and a covariance matrix $\mathbf{\Sigma}$; the joint probability density of the data is

$$g(\mathbf{z}) = (2\pi)^{-n/2}|\mathbf{\Sigma}|^{-1/2}\exp\left\{-\frac{1}{2}(\mathbf{z}-\mathbf{m})^T\mathbf{\Sigma}^{-1}(\mathbf{z}-\mathbf{m})\right\}. \tag{9}$$

[35] The vector $\mathbf{z}$ contains the n data, and the vector $\mathbf{p}$ contains the parameters of the covariance matrix $\mathbf{\Sigma}$. If the mean $\mathbf{m}$ and the covariance $\mathbf{\Sigma}$ are unknown and depend on parameter vectors $\mathbf{m}$ and $\mathbf{p}$, respectively, one can regard $\mathbf{z}$ as fixed and $g(\mathbf{z})$ as a function of $\mathbf{m}$ and $\mathbf{p}$, which is called the

likelihood function $L(\mathbf{m}, \mathbf{p})$. Maximizing the likelihood, or minimizing the negative log-likelihood $-\log L(\mathbf{m}, \mathbf{p})$ yields the parameter estimates. If a trend is apparent, the joint determination of the trend and the covariance parameters are prone to bias. The solution to this problem is to use restricted maximum likelihood (REML) to estimate the variance parameters, because it removes the dependence of the estimates on the nuisance parameter $\mathbf{m}$. Next, a mixed modeling procedure is used to estimate the fixed effects (e.g., the coefficients of a trend model), the random effects (the spatially dependent random variation), and the random error (nugget variation). For a detailed mathematical description of the method we refer to *Lark et al.* [2006].

[36] Within the framework of likelihood estimation, we also used the Matérn model in addition to the three standard correlation functions. The semivariance of this model is given by

$$\gamma(\mathbf{h}) = c_0 + c_1\left(1 - \frac{1}{2^{\nu-1}\Gamma(\nu)}\left(\frac{\mathbf{h}}{r}\right)^{\nu} K_{\nu}\left(\frac{\mathbf{h}}{r}\right)\right), \qquad (10)$$

where $\mathbf{h}$ is the separation distance, c_0 is the nugget variance, $c_0 + c_1$ is the sill variance, ν is the smoothness parameter ($\nu > 0$), Γ is the gamma function, r is the distance parameter, and $K\nu$ is a modified Bessel function of the second kind of order ν. The greater flexibility of the Matérn model compared with standard models is based on its smoothness parameter ν, which represents several theoretical models; examples are given by *Minasny and McBratney* [2005]. They stated that ν implies the "roughness" or "smoothness" of the spatial process ($\nu \rightarrow 0$: rough spatial process; $\nu \rightarrow \infty$: smooth spatial process). They concluded that ν should be estimated by restricted maximum likelihood because a weighted nonlinear least-squares estimation could give misleading results. *Lark et al.* [2006] stated that REML estimates are consistent as they converge in probability to the parameters to be estimated with less bias than both maximum likelihood estimates and method-of-moment estimates obtained from residuals of a fitted trend.

[37] REML estimation explicitly assumes a multivariate normal distribution, which cannot be verified because we only have one realization of the full joint distribution of a variable [*Pardo-Igúzquiza*, 1998]. Among others, *Lark et al.* [2006] discussed this problem in the context of spatial prediction of soil properties. *Chilès and Delfiner* [1999] suggested, in the framework of Gaussian simulation, that at least the bivariate distributions of the (transformed) data could be checked, e.g., via those $\mathbf{h}$-scattergrams which we already used for outlier detection. These plots ought to be elliptical when the point pairs are auto-correlated, or circular in the uncorrelated case; deviations from the normal distribution can be easily detected (Figure 3).

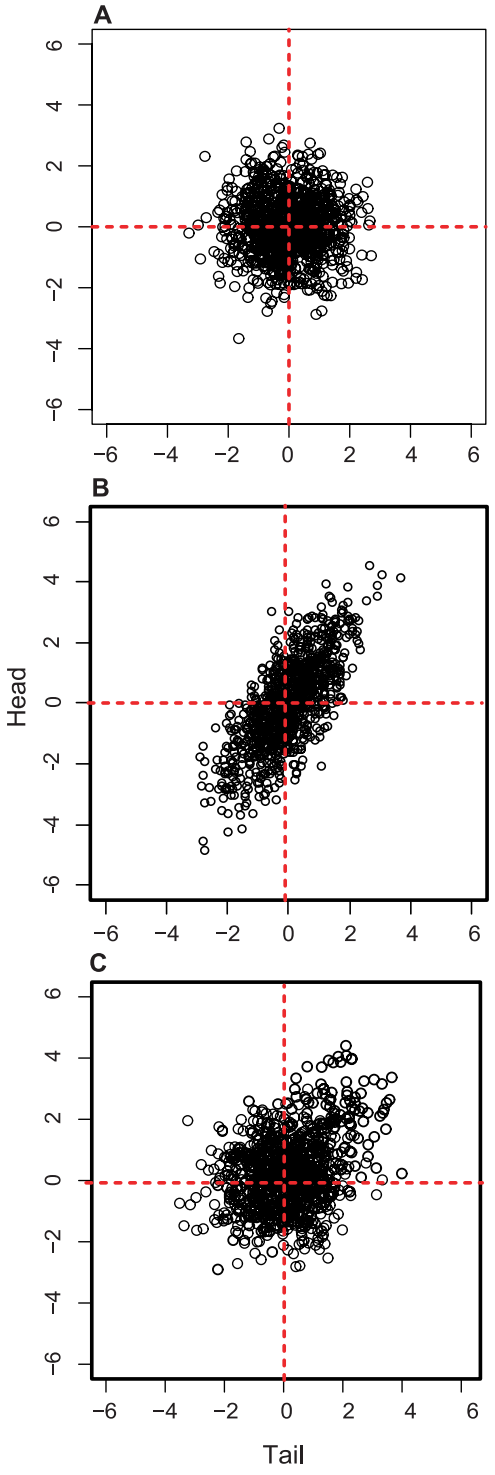

Figure 3. Examples of $\mathbf{h}$-scattergrams. These are plots of all pairs of the measured variable at locations $\mathbf{x}$ separated by a certain distance; the value at the start of the distance vector $\mathbf{h}$, $z(\mathbf{x})$, is called the tail value, and the value at the end of the distance vector, $z(\mathbf{x} + \mathbf{h})$, is the head value. The vertical and horizontal red lines correspond to the population mean. (A) Uncorrelated point pairs of a population that consists of two sets of randomly generated data of distribution N(0, 1), where N(μ, σ^2) denotes a normal distribution of mean μ and variance σ^2. (B) Correlated point pairs of a population that consists of two sets of randomly generated data, one of which is set 1 from Figure 3A and the other is the sum of the two sets of Figure 3A. (C) Uncorrelated point pairs of a population that consists of two sets of randomly generated data of distribution N(0, 1), each of which comprises a 10% contamination of a population with distribution N(2, 1).

Table 1. Comparison of Selected Studies of *Ks* Spatial Variability

Reference	Land use	Texture	Method[b]	Soil depth (m)	n	Scale triplet[a]		Trend Y/N	Variogram modeling				
						Spacing[c] (m)	Extend[d] (m)		Estimator[e]	Model[f]	Range (m)	Effective range (m)[g]	Nugget/sill (%)
Mohanty et al. [1991]	No-tillage agriculture	loam	CCHP (field)	0.15 (field)	66[h]	4.6	~90	Y	CH	Nug	0	0	100
			CHP (lab)	0.15 (lab)	66[h]				CH	Nug	0	0	100
				0.30 (field)	66[h]				CH	Nested[i],	23+60[i], 0[j]	23+60[i], 0[j]	28[i], –[j]
				0.30 (lab)	66[h]				CH	Nug[j]	18+46[i], 0[j]	18+46[i], 0[j]	17[i], –[j]
Mallants et al. [1996]	Agriculture	sandy loam	CHP (lab)	0–25	60	0.1, 0.9	31	N	M	Sph	3.5	3.5	65
				25–55	60				M	Nug	0	0	100
				55–100	60				M	Nug	0	0	100
Reynolds and Zebchuk [1996]	No-tillage pasture	silty clay	CCHP (field)	0.5	82	20	~200	?	M	Exp	18	54	0
			AHM (field)	0.5	68				M	Exp	21	63	0
Buttle and House [1997]	Forest	?	SRI (field)	surface	35	~15	~100	?	M	Gau, anis	0[k], 23[l]	0[k], 40[l]	?
			CHP (lab)	till bedrock	35					Exp, anis	0[k], 35[l]	0[k], 105[l]	?
Bosch and West [1998]	Tillage agriculture	loamy sand, sand	CCHP (field)	0.25	18/38[m]	0.5, 5	140	Y[m]	M	Lin/Lin[m]	5/38	5/38	46/62
				0.5	18/38[m]				M	Lin/Gau[m]	2/57	2/99	96/6
				1.4	18/38[m]				M	Lin/Gau[m]	17/67	17/115	18/15
				2	18/38[m]				M	Gau/Lin[m]	15/16	26/16	7/67
Mohanty and Mousli [2000]	No-tillage agriculture	loam	CHP (lab)	0.15	66[h]	4.6	~90	Y	M	Gau, anis	60, 143	104, 248	41
				0.3	66[h]				M	Sph, anis	46, 116	46, 116	30
Sobieraj et al. [2002]	Primary rainforest	sandy (clay) loam	CCHP (field)	0.2	18	25	250	N	M	Nug	0	0	100
				0.3	18				M	Nug	0	0	100
		sandy clay		0.5	18				M	Nug	0	0	100
Shukla et al. [2004]	Tillage agriculture	(sandy) loam, (loamy) sand	CCHP (field)	~0.15	60	25	?	?	M	Sph	154	154	54
Iqbal et al. [2005]	Agriculture	fine-silt	FHM (lab)	surface	209	79.4	~1000	?	M	Exp	94	282	31
		coarse silt		subsurface	209				M	Exp	110	330	50
		loam		deep	209				M	Exp	111	333	27
Duffera et al. [2007]	Farmland	loamy sand, sandy loam	CHP (lab)	0.4–1.2	60	~60	~350	?	M	Nug	0	0	100
				0.19–0.27	60				M	Nug	0	0	100
				0.34–0.42	60				M	Nug	0	0	100
				0.49–0.57	60				M	Nug	0	0	100
				0.64–0.72	60				M	Nug	0	0	100

[a]Support for lab methods in general 0.076 × 0.076 m; unknown for field methods.
[b]CCHP, compact constant-head permeameter; CHP, constant head permeameter; AHM, auger hole measurements; SRI, single-ring infiltrometer; FHM, falling-head method.
[c]Smallest separation distance.
[d]Longest separation distance.
[e]For experimental variograms, M: Matheron estimator, CH: Cressie-Hawkins estimator.
[f]Nug: pure nugget, Exp: exponential, Sph: spherical, Gau: Gaussian, Lin: linear, anis: anisotropic.
[g]For exponential model: range*3; for Gaussian model: range*$3^{0.5}$.
[h]With three replicates.
[i]Log-transformed data with trend.
[j]After trend removal.
[k]Normal to stream valley.
[l]Parallel to stream valley.
[m]Pine flat (clay increase with depth) and Troup.

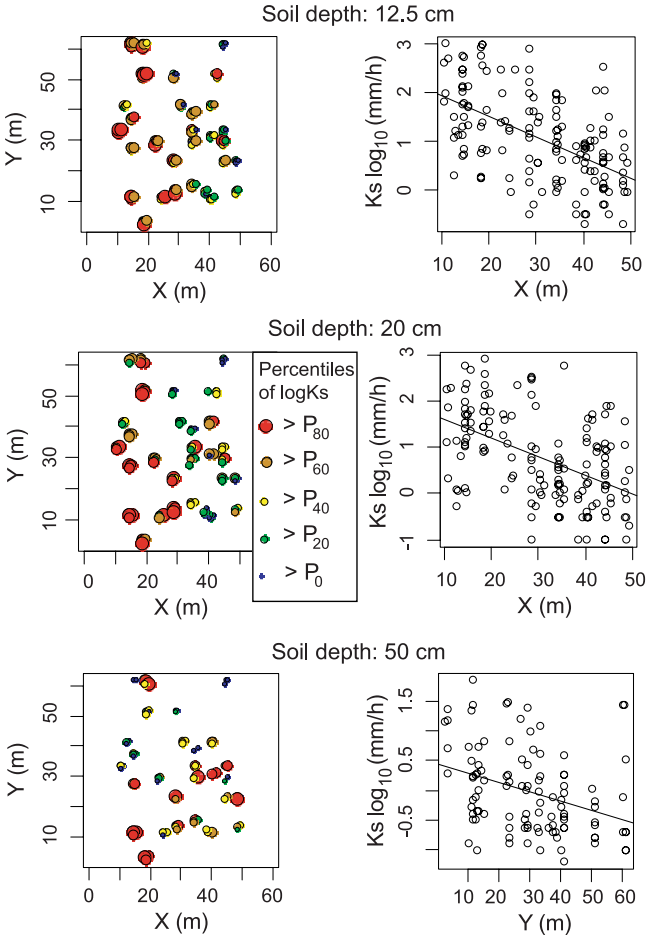

Figure 4. (left) Plots of the log-transformed data divided into quintiles (greater point sizes indicate higher *Ks*) and (right) linear regression of the transformed data with the respective coordinate as the independent variable.

[38] Since we wanted to evaluate the parameters of the variogram models that were derived by the described methods, we calculated the effective range in each case to allow for comparability of the correlation lengths independent of the model type. The effective range for the spherical model coincides with the distance parameter; for the exponential model, the effective range is approximately the distance parameter times 3, and for the Gaussian model it is estimated by the distance parameter times the square root of 3 [*Webster and Oliver*, 2001]. For the Matérn model we estimated an effective range at 95 % of the sill.

[39] As to how to deal with the nugget, *Minasny and McBratney* [2005] proposed to fix the nugget to the variance at the shortest possible separation distance (the smallest lag) due to difficulties to simultaneously estimate the nugget variance of the Matérn model with REML if the spatial process is rough. We adopted this proposal and tried both approaches within the REML estimation framework: estimating the nugget simultaneously, and fixing it to the semivariance at our shortest lag of 0.25 cm, which was calculated according to equation (2) (Matheron estimator).

3.2.2.3. Spatial Prediction

[40] In order to evaluate the influence on spatial prediction of the differences in variogram parameters involved in the described estimation techniques, we first produced maps by kriging. To this end, we interpolated the residual values via simple kriging to account for any trends, and then added the trend coefficients back to the interpolated residuals [*Goovaerts*, 1997]. We furthermore calculated the differences between kriged *Ks*-values that arose because of two different variogram models. For this purpose, we chose one covariance model as the reference against which the other models are compared.

[41] We then performed 100 conditional sequential Gaussian simulations, again using all different variogram models. For every model, we post-processed the 100 simulations to summarize the spatial uncertainty information. We displayed the outcome in probability maps: at each simulated grid node, the probability of exceeding a given threshold is evaluated as the proportion of the 100 simulated values that exceed that threshold [*Goovaerts*, 1997]. To assess the hydrological consequences of the spatial structure of *Ks*, e.g., the spatial distribution of possible impeding layers, we calculated for every grid node the probability that *Ks* is exceeded by some location estimate of local rainfall intensities. We extracted these intensities from ten-minute rainfall records for the period October 2005 to September 2006 (unpublished data T. Peters, Erlangen) of a climate station located some 100 meters away from our plot. From this data set, we calculated maximum 30-minute rainfall intensities of rainfall events that were separated from another by a dry period of at least 2 h. Location estimates of those intensities (e.g., median, upper quartile, maximum) can serve as 'thresholds': where they exceed simulated *Ks* values, impeding layers are likely to exist.

4. Results

4.1. Exploratory Data Analysis

[42] The Box-Cox transformations revealed that taking logarithms is the most appropriate transformation for the right-skewed *Ks* data, which corresponds to a multitude of other studies of *Ks* (see Table 1 for examples from the literature).

[43] Significant trends exist for the log-transformed data at the three soil depths, which appear to be a first-order polynomial on the x- (12.5, 20 cm depth) or the y-coordinate (50 cm depth, respectively (Figure 4). Thus they can easily be removed by linear regression. At the shallow soil depths, *Ks* decreases from the left to the right side of the plot (i.e., from gully to ridge), which we attribute to topography-dependent differences in soil depth and vegetation cover. At the 50 cm depth, larger *Ks* characterize the lower parts of the plot; however, this trend is not as consistent as it is for the topsoil. In summary, a trend denotes a local feature of our data and may be an effect of either selecting too small an extent [*Skøien and Bloeschl*, 2006] or of a structured variability that in turn is a fingerprint of the landscape organization, in our case slope affecting soil formation.

[44] Consequently we performed all subsequent analyses with the residuals of the linear regression or estimated the trend coefficients as fixed effects within the REML estimation of the covariance function, respectively. The univariate distribution of the residuals appears approximately Gaussian with a somewhat larger coefficient of skewness at 50 cm

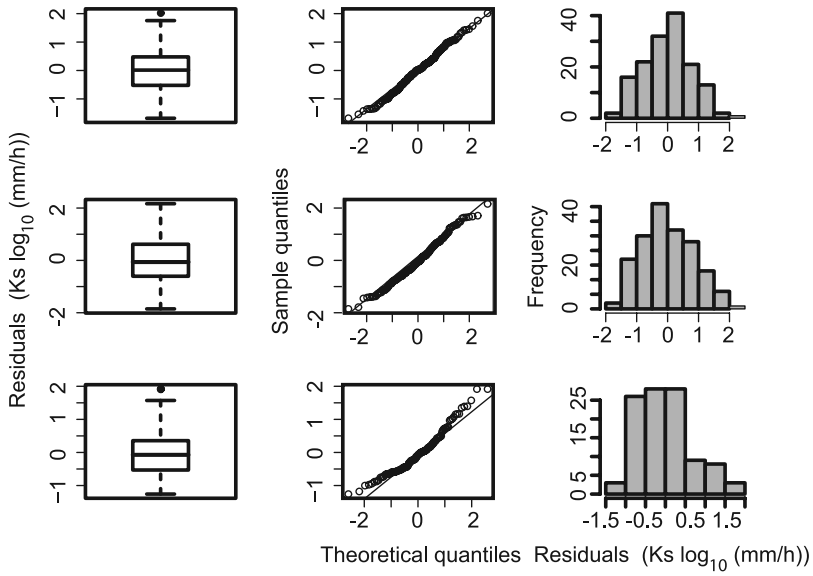

Figure 5. Univariate distribution of the residuals from regression for the (top) 12.5 cm, (middle) 20 cm, and (bottom) 50 cm soil depth.

soil depth (Figure 5). **H**-scattergrams display bivariate normal distributions and do not reveal any outlying values (Figure 6). They also indicate weak and very short autocorrelations at 12.5 cm depth; the linear correlation coefficient is only 0.35 for lags between 0.25 and 0.9 meters, and measurements only 1.1 meter apart are uncorrelated. Autocorrelations are more pronounced at the 20-cm and 50-cm depths, but vanish after a few meters.

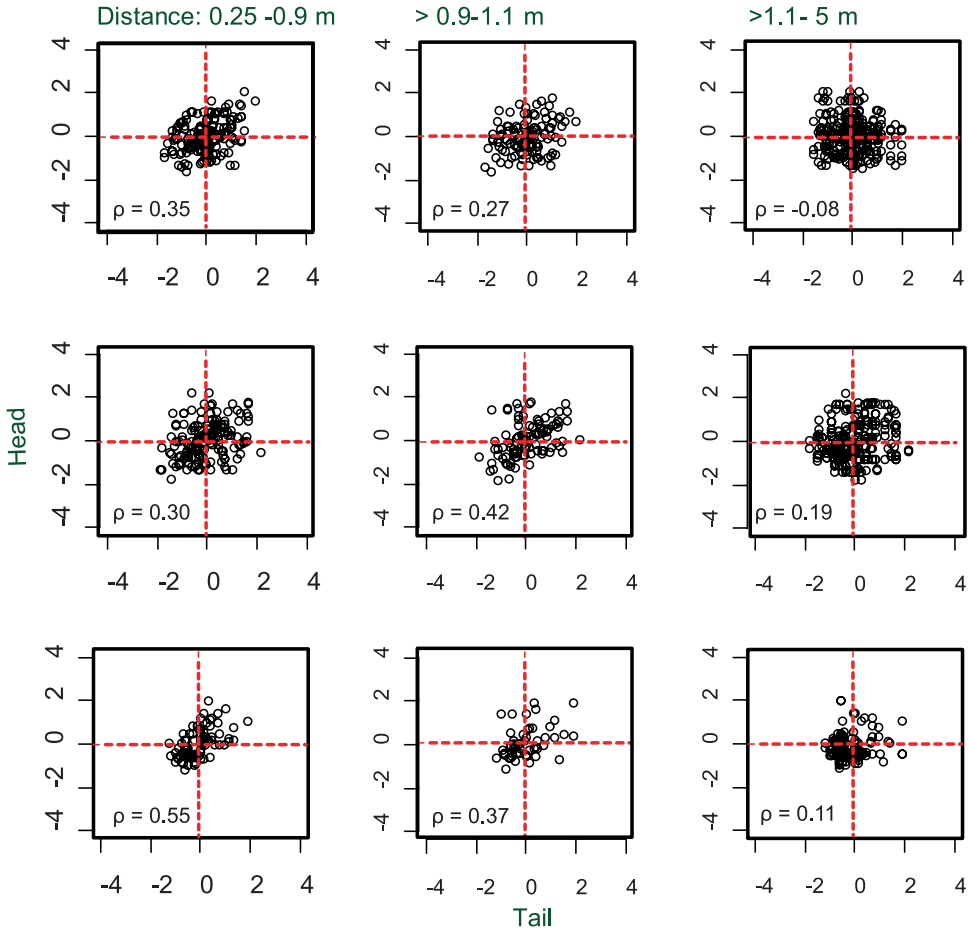

Figure 6. Bivariate distributions of the residuals for the (top) 12.5 cm, (middle) 20 cm, and (bottom) 50 cm soil depth; ρ is the Pearson correlation coefficient.

Table 2. Results of Robust Estimation

Soil depth (cm)	Estimator[a]	Model[b]	Nugget	Partial sill[c]	Sill	Range	Nugget/sill *100 (%)	Median of θ(x)	Mean of θ(x)
12.5	M	Exp	0.43	0.16	0.58	3.24	72.83	0.457	1.00
	D	Exp	0.36	0.29	0.64	3.24	55.53	0.535	1.08
	CH	Exp	0.45	0.16	0.61	3.24	73.68	0.433	0.97
	G	Exp	0.53	0.07	0.60	17.82	88.32	0.505	1.03
20	M	Gau	0.40	0.34	0.75	8.10	54.05	0.483	1.15
	D	Exp	0.33	0.50	0.83	6.48	39.85	0.534	1.19
	CH	Exp	0.40	0.40	0.81	4.86	50.02	0.477	1.00
	G	Exp	0.46	0.22	0.68	4.86	68.29	0.423	0.97
50	M	Exp	0.23	0.27	0.49	9.72	45.56	0.426	1.04
	D	Gau	0.17	0.28	0.45	11.34	37.11	0.755	1.75
	CH	Exp	0.26	0.30	0.56	21.05	46.70	0.445	1.06
	G	Exp	0.36	0.16	0.52	6.48	68.60	0.350	0.77

[a]Experimental variogram estimators proposed by Matheron (M), Dowd (D), Cressie-Hawkins (CH), and Genton (G).
[b]Selected theoretical variogram model, Exp: exponential; Sph: spherical; Gau: Gaussian.
[c]Sill variance less the nugget variance.

[45] We did not encounter problems in trend decomposition, such as heteroscedastic variance of residuals or linear constraints on the residuals [*Leuangthong and Deutsch*, 2004].

4.2. Geostatistical Analysis

4.2.1. Experimental Variogram Estimation

[46] Robust and classical (Matheron) variogram estimation differed considerably with respect to both the selected model and the theoretical variogram parameters nugget, sill and range at all depths (Table 2). Since the variograms depend strongly on the selected lag classes, we were not able to delineate any systematic features of the several estimators. The median of $\theta(\mathbf{x})$ always remains within the confidence limits when using the Matheron estimator, and likewise for the robust estimators in most instances. Thus the Matheron estimator can be used for all data sets, and robust estimation is not required at all.

[47] Anisotropic behavior is not evident (Figure 7). Obviously, if calculating directional variograms for the log-transformed data containing the trend component one could assume zonal anisotropy (Figure 8); this is particularly conspicuous at the shallow soil depths, where the strongest spatial correlation (without reaching a sill due to the unbounded variance) coincides with the direction of the trend component.

4.2.2. Estimation of Covariance Parameters

[48] Greatest variations in the estimated covariance parameters characterize the 50 cm depth, where the spatial autocorrelation is strongest, in contrast to the other soil depths where all models suggest a very short correlation (Table 3 and Figure 9). At the 20 cm depth, the models of the evenly spaced experimental variograms have comparatively short ranges, but only the WLS model of the unevenly spaced experimental variogram, which was fitted over half the maximum lag, appears out of range compared to most of the models. At this depth, it was impossible to fit the Matérn model because of a flat likelihood function that does not reach a distinctive maximum. The most pronounced features at the 50 cm depth include (1) the somewhat larger sill of the REML models, which is certainly due to the different estimate of the trend coefficients compared to those of the linear regression, (2) the similarity between the Matérn and the respective standard model, and (3) the similarity of the ML models of the residual data with the LS models of the evenly spaced experimental variograms. The largest deviation from the bulk of models is

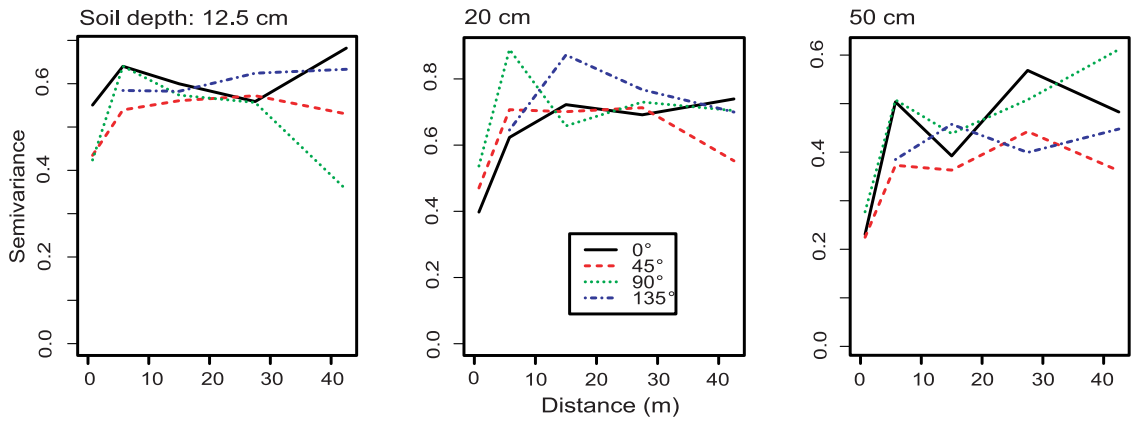

Figure 7. Directional variograms of the residuals: 90° follows the direction of the contour line (i.e., the x-coordinate) and 0° is the direction of the slope. Angle tolerance is 22.5°.

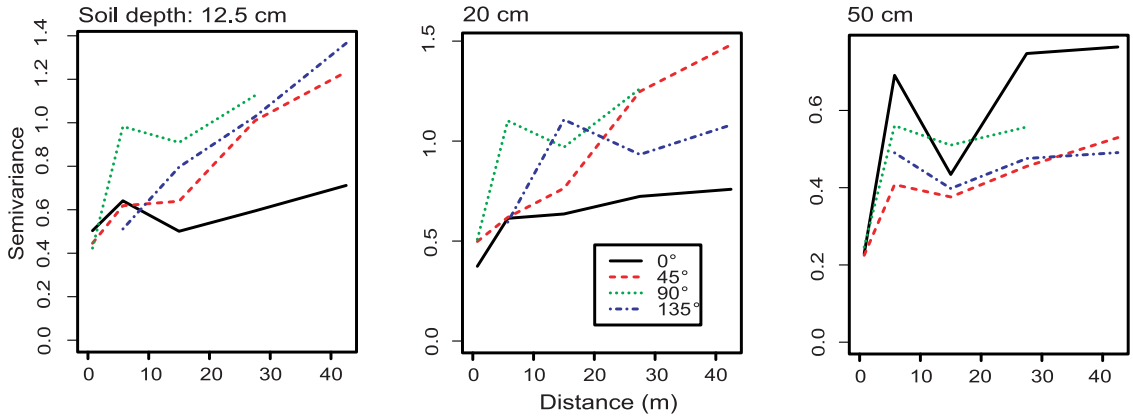

Figure 8. Directional variograms of the log-transformed data before trend decomposition: 90° follows the direction of the contour line (i.e., the x-coordinate) and 0° is the direction of the slope. Angle tolerance is 22.5°.

associated with the OLS1 estimate, which composes a comparatively large range and sill. This is due to the fact that this model fits with equal weights to all points of the unevenly spaced experimental variogram (Figure 9); i.e., the large semivariance at the largest lag (possibly an edge effect) has to be incorporated in the estimate.

[49] The median of $\theta(\mathbf{x})$ does not appear to be a sensitive indicator of the above differences since it shows only small and insignificant variation; moreover, its optimum in the sense of being closest to the expectation does not correspond to the model that gives the minimum sum of squares or the largest likelihood, respectively (Table 3). When considering the mean of $\theta(\mathbf{x})$, however, a somewhat clearer picture emerges since it is always very close to the expectation for the likelihood models (Table 3). A systematic behavior emerges at least at 50 cm depth, where the mean of $\theta(\mathbf{x})$ exceeds the expectation for the models of the unevenly spaced experimental variograms and falls below it for those of the evenly spaced experimental variograms (not shown).

4.2.3. Spatial Prediction

[50] Because of the rather weak spatial structure at the shallow soil depths, we focus in the following on the prediction for the 50 cm soil depth.

Table 3. Summary of the REML Models

Parameters and evaluation of REML models	Soil depth (cm)				
	12.5		20.0[a]	50.0	
Model[b]	Mat	Sph (Exp)	Gau (Gau)	Mat	Exp/Sph[c](Gau)
Nugget[d]	0.17 (0.39)	0.11 (0.39)	0.44 (0.50)	0.14 (0.16)	0.11/0.12 (0.16)
Partial sill[e]	0.41 (0.22)	0.47 (0.22)	0.30 (0.28)	0.37 (0.35)	0.40/0.38 (0.34)
Sill	0.58 (0.61)	0.58 (0.61)	0.74 (0.78)	0.51 (0.51)	0.51/0.50 (0.50)
Range (m)	0.03 (0.70)	0.76 (1.15)	4.85 (4.88)	1.30 (1.04)	1.92/3.84 (1.60)
Effective range[f] (m)	∼0.6 (∼2)	0.76 (3.45)	8.40 (8.45)	∼4.5 (4)	5.76/3.84 (2.77)
Nugget/Sill	0.29 (0.64)	0.19 (0.64)	0.59 (0.64)	0.24 (0.31)	0.22/0.24 (0.32)
ν[g]	44.16 (0.83)	-	-	0.86 (1.26)	-
LogL[h]	−162.6 (−163.5)	−162.5 (−163.5)	−167.6 (−168.1)	−89.83 (−89.87)	−89.87/−89.87 (−89.95)
β_0[i]	2.4807 (2.4414)	2.4836 (2.4396)	2.3717 (2.3733)	1.0984 (1.0907)	1.1237/1.0601 (1.0515)
β_1[i]	−0.0425 (−0.0419)	−0.0425 (−0.0420)	−0.0410 (−0.0411)	−0.0034 (−0.0033)	−0.0040/−0.0025 (−0.0024)
β_2[i]	−0.0026 (−0.0021)	−0.0026 (−0.0021)	−0.0105 (−0.0104)	−0.0169 (−0.0168)	−0.0171/−0.0166 (−0.0165)
θmedi[j]	0.480 (0.457)	0.474 (0.437)	0.505 (0.443)	0.432 (0.402)	0.431/0.446 (0.382)
Deviance E{θmedi}[k] (%)	2.5 (0.2)	1.9 (−1.8)	5 (−1.2)	−2.3 (−5.3)	−2.4/−0.9 (−7.2)
θmean[l]	1.026 (0.931)	0.974 (0.929)	0.991 (0.882)	0.989 (0.972)	0.990/1.008 (0.985)
Deviance E{θmean}[m] (%)	2.6 (−6.9)	−2.6 (−7.1)	−0.9 (−11.8)	−1.1 (−2.8)	−1.0/0.8 (−1.5)

[a]Impossible to fit the Matérn model.
[b]Correlation function: Ma, Matérn; Sph, spherical; Exp, exponential; Gau, Gaussian.
[c]Indifferent log-likelihood.
[d]Values in parentheses refer to the estimated parameters when the nugget was fixed to the semivariance of the first separation distance.
[e]Sill variance less the nugget variance.
[f]For exponential model: range*3; for Gaussian model: range*$3^{0.5}$; for Matérn model: implied visually at 95 % of the sill.
[g]Smoothness parameter of the Matérn function.
[h]Maximized log-likelihood.
[i]Fixed effects.
[j]Median of the θ-statistics.
[k]Deviance from the expectation of the median of $\theta(\mathbf{x})$.
[l]Mean of the θ-statistics.
[m]Deviance from the expectation of the mean of $\theta(\mathbf{x})$.

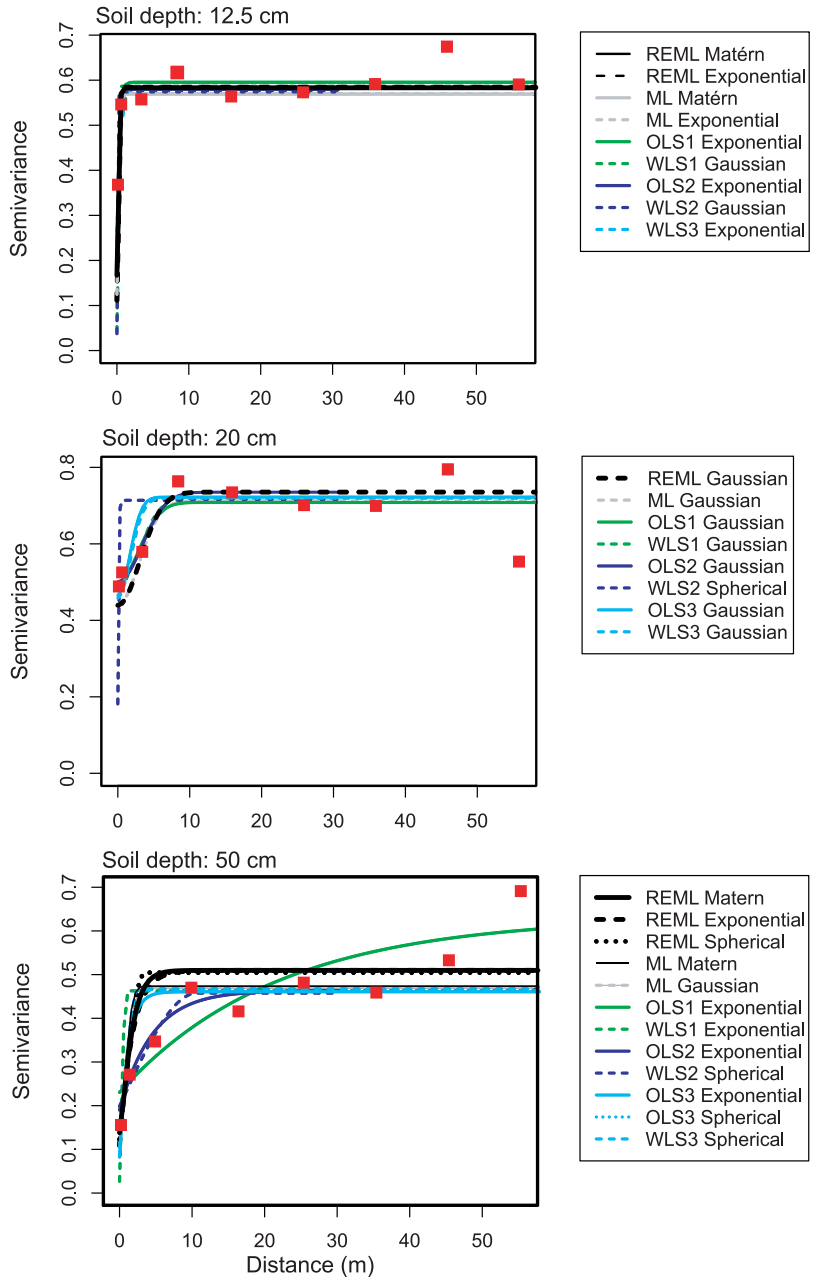

Figure 9. Fitted variogram models. Estimation methods are restricted maximum likelihood (REML), maximum likelihood (ML), ordinary least squares (OLS), and weighted least squares. OLS1 and WLS1 were fitted to an unevenly spaced (i.e., according to the sampling design) experimental variogram over the whole separation distance, OLS2 and WLS2 are based on the same experimental variogram fitted over half the separation distance, and OLS3/WLS3 are fitted to the evenly spaced experimental variograms (one lag every 25 cm). Red squares show the unevenly spaced experimental variogram.

[51] In general, the differences among the kriged maps are restricted to certain locations within the plot, that is to say, to the left-lower corner and a cluster near the right upper corner (Figure 10). Because of the similarity of the Matérn and the standard REML models, the kriging predictions do not differ much, with some locally larger deviations for the spherical model. The latter is also true for the ML and the LS estimates of the evenly spaced experimental variogram but here some clusters exist with more severe over- or underestimation compared to the REML-Matérn reference model. Largest deviations are associated with the LS models of the unevenly spaced experimental variograms both for those that were fitted over the whole as well as over half the separation distance. The kriging map based on a model with a linear trend shows a completely different spatial pattern, in addition to the absolute deviations.

[52] The assessment of the spatial uncertainty revealed that the average number of *Ks* values in every of the 5 probability classes is quite similar regardless of the respective model (Table 4). For instance, between 25 and 32 % of the simulated values are subject to a very small probability

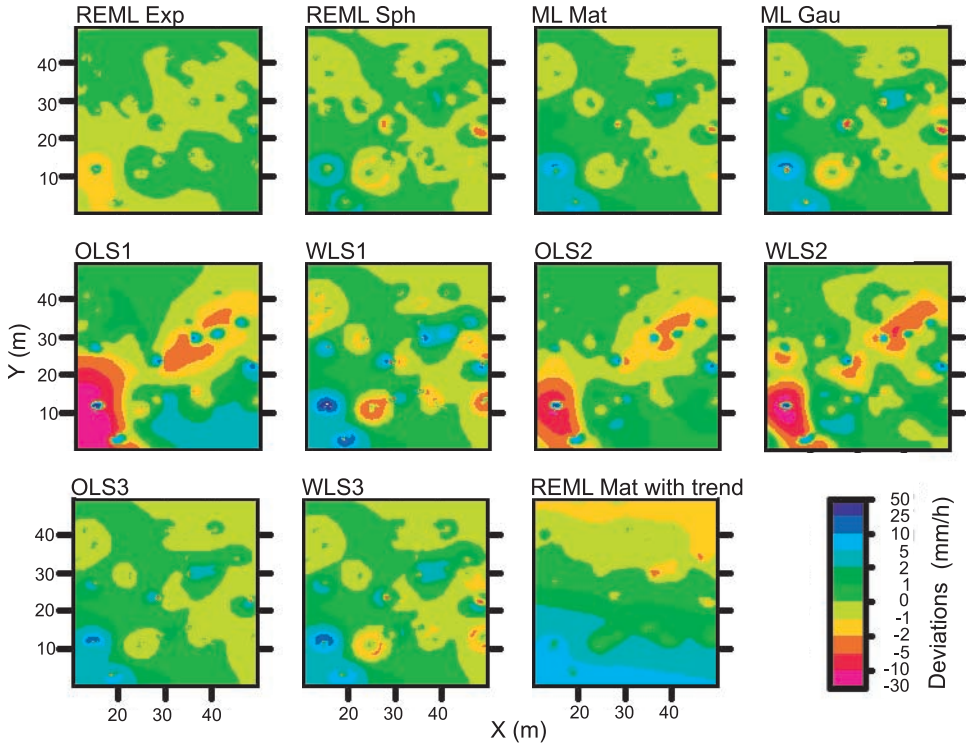

Figure 10. Differences between kriged *Ks*-values due to differences between the estimated covariance parameters at 50 cm soil depth. We chose the REML-Matérn model as the reference with which the other models are compared. These are (from left to right and from top to bottom) the REML-exponential, the REML-spherical, the ML-Matérn, the ML-Gaussian, the OLS model fitted to a unevenly spaced experimental variogram over the whole separation distance (OLS1), the WLS model fitted to a unevenly spaced experimental variogram over the whole separation distance (WLS1), the OLS model fitted to a unevenly spaced experimental variogram over half the separation distance (OLS2), the WLS model fitted to a unevenly spaced experimental variogram over half the separation distance (WLS2), the OLS model fitted to the evenly spaced experimental variograms (OLS3), the WLS model fitted to the evenly spaced experimental variograms (WLS3), and a REML-Matérn model without fixed effects (i.e., no trend elimination).

(i.e., below 20 %) of being exceeded by the median of the maximum 30-minute rainfall intensities. Hence, in our example, model differences are insignificant for the mean probability. However, the spatial arrangement, e.g., the existence or nonexistence, respectively, of connected patches of similar permeability, clearly depends on the covariance parameters used (Figure 11). Because of the comparatively long range of the OLS1 model, it simulated pronounced clusters of high or low probabilities, suggesting an impeding layer at the upper left corner and an area of unrestricted percolation in the lower left parts of the plot. Smaller but still pronounced clusters were produced by the

Table 4. Probability That the Median of the Maximum 30−Minute Rain Intensities Exceeds *Ks*

Method[a]	Number of predicted Ks values/class (%): mean±standard deviation				
	Class[b] 1 "very unlikely"	Class 2 "unlikely"	Class 3 "indifferent"	Class 4 "likely"	Class 5 "very likely"
REMLMat/REMLSph	25.8 ± 0.7	43.1 ± 1.2	28.2 ± 0.8	2.8 ± 0.7	0.1 ± 0.1
MLMat/MLGau	24.1 ± 0.3	46.7 ± 1.7	27.5 ± 1.1	1.6 ± 0.3	0.1 ± 0.0
OLS1/WLS1	28.7 ± 11.2	43.7 ± 11.1	22.8 ± 6.8	4.9 ± 6.6	0.0 ± 0.0
OLS2/WLS2	32.3 ± 1.3	38.8 ± 1.7	23.3 ± 0.1	5.5 ± 0.4	0.1 ± 0.1
OLS3/WLS3	23.3 ± 0.4	48.0 ± 0.7	27.0 ± 0.2	1.7 ± 0.1	0.1 ± 0.0

[a]Model are summarized regarding the estimation method: REML/ML: Matérn and a standard model estimated by REML/ML; OLS1/WLS1 & OLS2/WLS2: ordinary and weighted least squares models, whole and half distance fit, resp., of visually selected experimental variograms; LS3: LS models of regular exp. variograms.
[b]Probability classes: 1: 0 − < 2, 2: > = 2 − < 4, 3: > = 4 − <6, 4: > = 6 −< 8, 5: > = 8 − < 10.

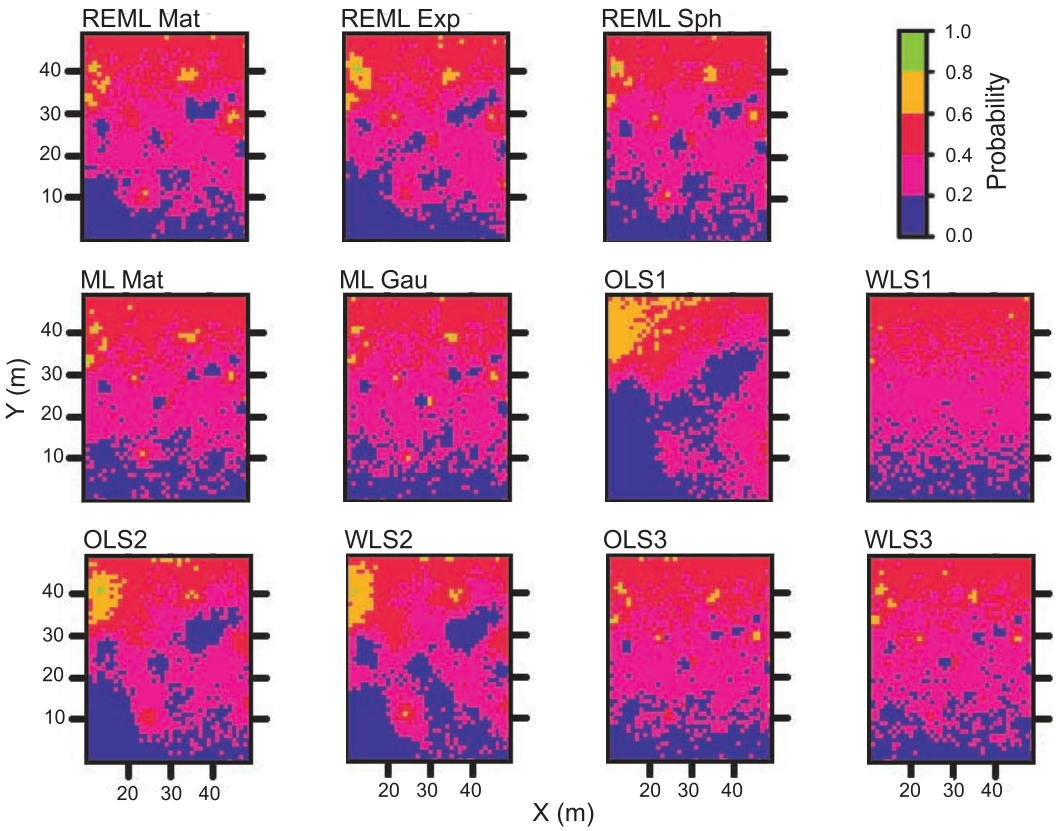

Figure 11. Probability maps displaying the probability that the median of the maximum 30-minute rain intensities exceeds the soil permeability at 50 cm soil depth. All fitted covariance models were used to run the conditional simulations. Please refer to Figure 10 for abbreviations.

LS2 models, whereas the WLS1 model resulted in a complete random pattern, which is, by and large, also true for the ML and LS3 estimates. The REML models take a middle position with some connected patches; in addition, the probability map again mirrors the small differences between the Matérn and the standard covariance models.

[53] In summary, the predictions reproduced the differences in the covariance parameters. The unevenly spaced experimental variogram based on lag distance groups and the resulting smoothing, yielded a fundamentally different pattern, in our case the existence of a spatially explicit impeding layer. The kriging maps (Figure 11) confirm the simulation results, and local deviations between them can be as high as 50 mm/h.

[54] Since we also wished to assess the importance of various parts of the variogram analysis for application purposes, we calculated, by way of one example, differences among kriging maps, for which we considered three cases: (1) different model estimation, (2) ignoring the trend component, and (3) using the robust Cressie-Hawkins estimator (Figure 12). Not surprisingly, the violation of the intrinsic hypothesis is most crucial for the shallow soil, where the linear trend was particular strong. At both soil depths, also the use of the robust instead of the classical estimator has severe consequences for kriging. This comparison highlights the outstanding importance of a thorough exploratory data analysis, and the risk of arbitrarily using robust estimation techniques.

5. Discussion

5.1. Exploratory Data Analysis

[55] There are several approaches to check the assumption of stationarity. We used diagnostic plots and evaluated apparent trends by testing the significance of the regression slope estimate. Within the framework of REML estimation *Lark et al.* [2006] used the Wald statistics to evaluate the null hypothesis that the fixed effects are zero. If the user chooses REML estimation, this approach can be adopted in addition to the visual diagnostics. Although it is established that the sill variance should not be equated to the sample variance [e.g., *Barnes*, 1991], a comparison between these measures can be used as a control. The comparison depends on boundary conditions that *Barnes* [1991] summarized in a rule-of-thumb: the data must be somewhat evenly distributed over an area with an extend greater than three times the range of the variogram. Since the rule-of-thumb is valid in our case, sill and sample variance of our data should approximately agree. Table 5 shows the exact match of those measures when the trend is taken into account, and the discrepancy between them when the trend is ignored. In the latter case, the continuously increasing semivariance with

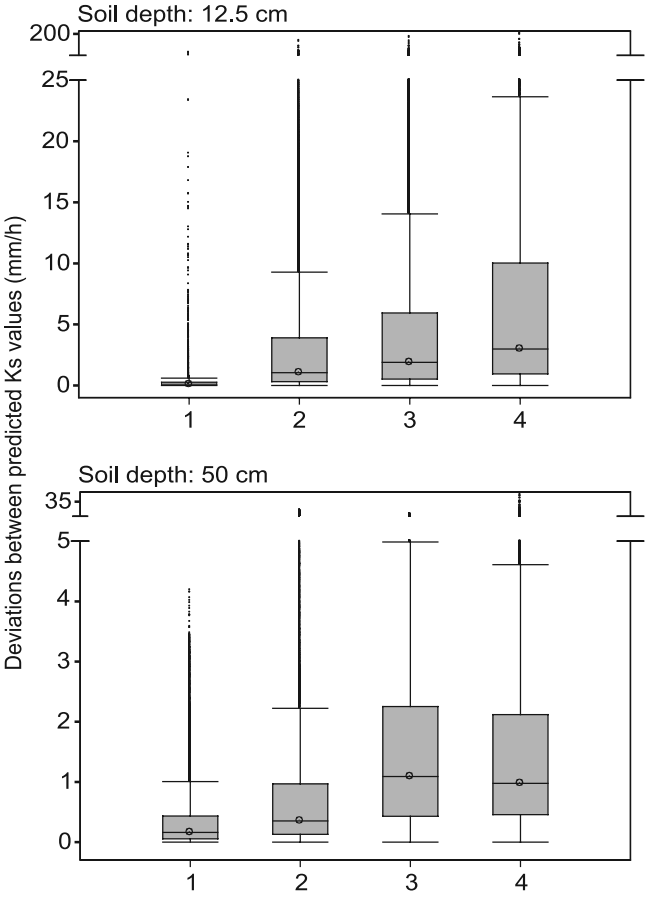

Figure 12. Boxplots of the absolute differences between predicted *Ks*-values that are due to different aspects of the variogram analysis denoted by the numbers 1–4 below the x-axis. The REML-Matérn model serves as the reference, which is compared to (1) the best standard REML model (selection criterion: largest maximized likelihood), (2) the best least-squares model (selection criterion: median of $\theta(\mathbf{x})$), (3) the best standard REML model without fixed effects (i.e., no trend elimination), and (4) the ordinary least squares model for the experimental variogram of the Cressie-Hawkins estimator (see Table 2 for parameters).

increasing lag in direction of the trend component (Figure 8) yields estimates of the sill variance that exceed the sample variance (Table 5).

5.2. Geostatistical Analysis

5.2.1. Experimental Variogram Estimation: Robust Estimation Requirements

[56] Robust estimation coincides with considerable differences in the covariance parameters nugget, sill, and range. In our case, the Matheron estimator shows both a median and a mean of the θ-statistics which are very close to the expectation. This is also true for the Cressie-Hawkins estimator, whereas the other two robust estimators performed worse regarding this criterion, particularly the Genton estimator. The absences of extreme values that may negatively affect the classical estimator were confirmed by the univariate and bivariate distributions (see Figures 5 and 6). Hence our study corroborates the proposal by *Lark* [2000a] that the θ-statistics may qualify as a criterion to choose between classical and robust estimation. The arbitrary use of a robust estimator instead of the classical one might be disadvantageous, first because robust estimators all assume an underlying distribution that is multivariate normal and second because they are less efficient than Matheron's estimator [*Lark*, 2000a] in the absence of outliers. Another drawback of robust estimation is the limited comparisons with other studies where the classical estimator has been used.

[57] In summary, our results confirm the usefulness of the statistic $\theta(\mathbf{x})$. In this study, we used a rather rough approach because the confidence limits for the median of $\theta(\mathbf{x})$ ignore spatial dependence between the locations. For this reason *Lark* [2002] developed a method to compute Monte Carlo confidence intervals in the framework of modeling soil properties as contaminated regionalized variables.

5.2.2. Estimation of Covariance Parameters

[58] Fitting parametric models to experimental variograms remains the method of choice in dealing with the spatial variability of environmental variables (see Table 1 for examples concerning saturated hydraulic conductivity), although some authors encourage estimation of the covariance model by restricted maximum likelihood [e.g., *Pardo-Igúzquiza*, 1998; *Stein*, 1999; *Lophaven et al.*, 2002], especially for unsystematic sampling [*Lark*, 2000b]. Regarding hydrological modeling, *Unlu et al.* [1990] and *Russo et al.* [1997] used REML to estimate the covariance functions of parameters of some hydraulic properties including *Ks*. In the former study, REML estimates of the sill and range turned out more accurate and consistent than the corresponding OLS- and ML-based estimates. Because there is no objective "test" criterion that "proves" the advantage of the REML procedure (the θ-statistic is a rather rough measure because it ignores correlation between the separate validation points), we discuss the basis for our preference in the spatial modeling of our *Ks* data in the following.

[59] Maximum-likelihood techniques allow the estimation of covariance parameters directly from the data, whereas least-squares techniques rely on an experimental variogram. A related criticism of maximum likelihood estimation is that the sampling scatter directly influences the parameter estimation [*Chilès and Delfiner*, 1999]. Least-squares methods frequently use lag classes to calculate the experimental variogram; hence, a semivariance estimate for a particular lag bin is an average of all contributing point pairs within it. This "averaging effect" is of major importance when broader or unequal lag distance classes are used. In our case, this is true for the unevenly spaced experimental variograms, which reflect the unsystematic sampling and which hence use lag classes that are narrower for short

Table 5. Comparison of Sample Variance and Sill Variance

Soil depth (cm)	Variance[a] of lgKs	Variance[b] of residuals	Sill[c]	Sill[d]
12.5	0.8	0.6	1.8	0.6
20	0.9	0.7	1.2	0.7
50	0.5	0.5	0.6	0.5

[a]Sample variance of the log−transformed *Ks* data.
[b]Sample variance of the residuals of the linear trend.
[c]Sill variance of REML model when trend is ignored.
[d]Sill variance of the REML-Matérn model.

separation distances than for longer ones. Although the dependence of the experimental variogram on the choice of the lag classes is a well-known fact, there is no general rule for this choice (except that they should not overlap). Hence this step in the variogram analysis is always arbitrary, and especially so when the sampling is unsystematic. At any rate, it is questionable if point measurements should be summarized in classes which are then used for modeling even though the subsequent prediction or simulation is again based on very small units, e.g., a grid cell in a prediction grid. Another aspect is the generally greater number of point pairs in some, in our case the middle, lag classes, which leads to a decrease of scatter in the respective region of an experimental semivariogram.

[60] The simultaneous estimation of fixed effects is a further benefit of REML since the variogram of the residuals of the linear regression is a biased estimate of the variogram of the random component of spatial variation [*Cressie*, 1993]. Nevertheless, the residuals should be computed for exploratory purposes, e.g., to check their bivariate distribution and to assess the plausibility of the effective range. The **h**-scattergrams (Figure 6) indicated that the correlation lengths are within the range of a few meters. Thus the large and small effective ranges, respectively, of the LS-fitted models for the whole-distance variogram at 50 cm depth appear rather implausible according to these scattergrams in contrast to the effective ranges of the REML models. That is to say, **h**-scattergrams help to at least roughly estimate the correlation length.

[61] In summary, our results provide convincing evidence for the gains associated with the use of restricted maximum likelihood to estimate fixed and random effects in geostatistical modeling (referred to as the "REML-E-BLUP" approach by *Lark et al.* [2006]), so long as the bivariate distributions are approximately normal.

5.2.3. Role of the Nugget Value

[62] *Cambardella et al.* [1994] proposed that the strength of spatial dependence can be characterized by the nugget-to-sill ratio. For instance, a ratio smaller than 25 % suggests a strong spatial dependence. Our criticism regarding this approach derives from a theoretical point of view: because the nugget is usually estimated by extrapolation from a covariance model, it cannot be verified by the data. Our results for the REML models (Table 3) illustrate what this means practically: at the 12.5-cm depth, the nugget-to-sill ratio indicates strong spatial dependence, whereas for the 20-cm depth one would claim a moderate spatial dependence at best. Since the **h**-scattergrams (Figure 6) did not reveal a stronger correlation of point pairs separated by the smallest separation distance for the 12.5-cm depth compared to the 20-cm depth, it is questionable if the strength of the autocorrelation differs much between these depths. Fixing the nugget to the semivariance at the shortest lag yields more plausible nugget-to-sill ratios. The notion that it may be advantageous to fix the nugget instead of estimating it simultaneously was also put forward by *Minasny and McBratney* [2005] who reported difficulties in nugget estimation when using the Matérn model with a small smoothness parameter. In addition, our results (Table 3) show that ν can heavily depend on the nugget value, as well as the other parameters of the covariance function. Because of the very short-ranging autocorrelation at the shallowest soil depth, where the difference between fixed and estimated nugget is most pronounced, the influence of those differences on spatial prediction cannot be investigated. In cases where the range allows for meaningful predictions and where the estimated nugget considerably differs from the fixed one, the decision as to which one to use may be difficult and will depend on the user's experience. Uncertainty in the semivariance estimation at the shortest lag due to a small number of contributing point pairs may encourage the user to estimate the nugget by model extrapolation, which is unavoidable if his or her sampling scheme does not allow for a very short lag relative to the extent. At any rate, a sampling design which allows for calculating the semivariance at short lags is highly advantageous. For instance, estimating the parameters of the Matérn correlation function requires short lags [*Marchant and Lark*, 2007].

5.2.4. Spatial Prediction

[63] *Goovaerts* [1997] argued that a-priori decisions, such as the assumption of stationarity and the random function model for modeling uncertainty, are far more consequential for interpolation than the choice of a variogram model. This is particularly true in situations where, for instance, the intrinsic hypothesis does not hold (see, for example, Figure 12 for the 12.5-cm depth), or where the choice of simulation algorithm is not appropriate for a specific problem (e.g., Gaussian versus indicator simulation). However, our results also show that the estimation method of the covariance parameters can become important; both the differences in spatial connectivity and in the absolute deviations are likely to affect applications, e.g., the use of the interpolated permeability values as an input for hydrological models.

6. Comparison With Other Studies of *Ks* Spatial Variability

[64] It is currently impossible to relate the spatial structure of soil permeability to some easily predictable or measurable feature such as land use, soil texture, or soil depth (Table 1). To complicate things even more, the choice of a sample scale triplet (support, spacing, and extend, see *Bloeschl* [1999]) relative to the scale of the underlying soil variability affect the estimates of the variogram parameters; e.g., apparent correlation lengths always increase when increasing spacing, extent or support [*Western and Bloeschl*, 1999].

[65] Table 1 reveals that agriculture, for example, seems to promote spatial patterns that are completely random, but also strong and extensive autocorrelations; data from forested land are still too rudimentary to allow any conclusions. Deviations from the intrinsic hypothesis are found in many cases (Table 1, column "Trend"); hence, special attention should be given to this part of the variogram analysis. For the majority of the studies, the sample sizes fall considerably below the recommended minimum of 150–200 for the method-of-moments variogram [*Webster and Oliver*, 1992]. This shortcoming certainly influences the results and hence this comparison. The required sample sizes for REML estimation may differ from this recommendation, or may even be smaller, but this has not yet been addressed adequately.

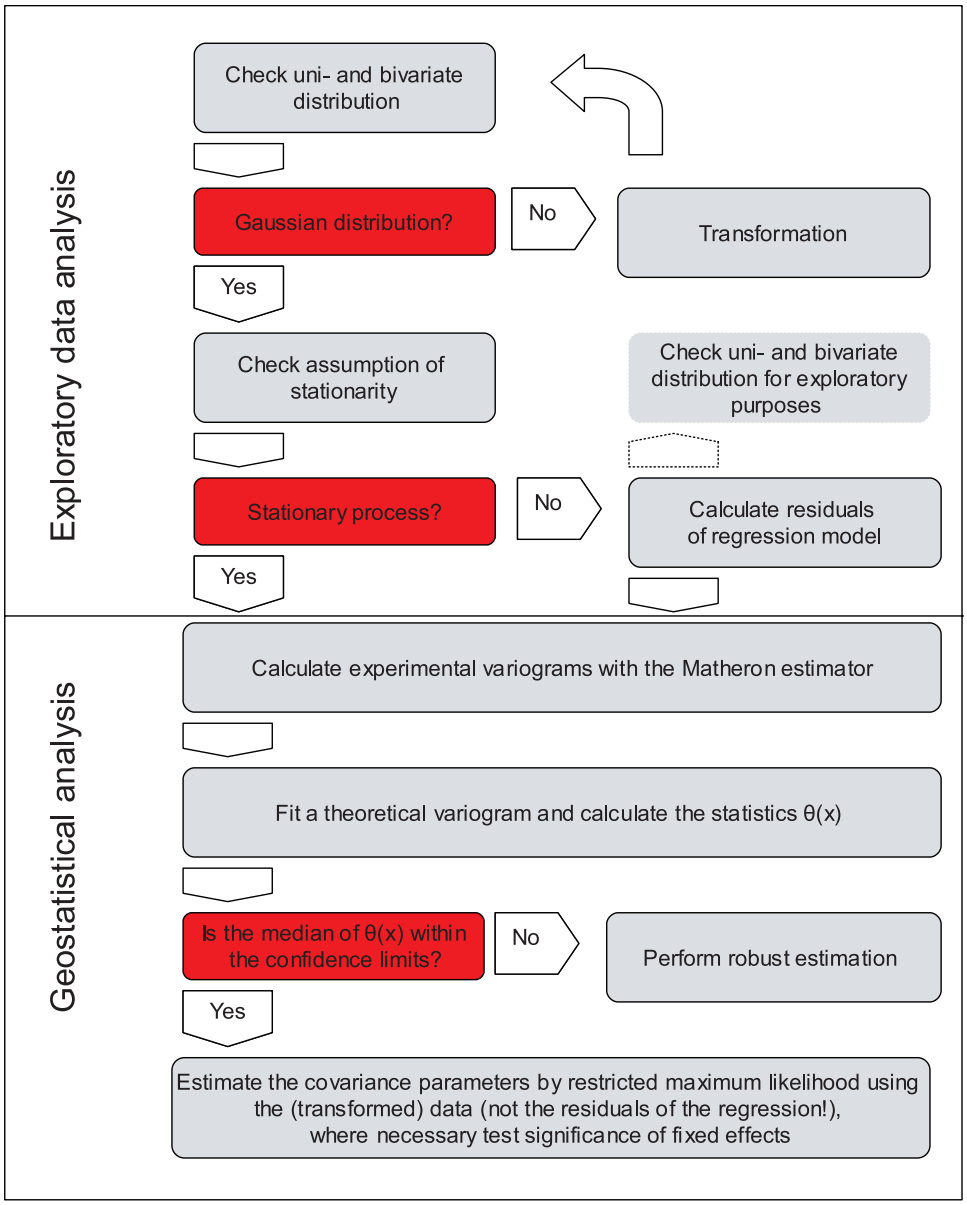

Figure 13. Proposed flowchart for spatial data analysis. The red boxes contain the most critical decisions.

[66] Where spatial structure exists, it is attributed to a well-developed and stable soil structure [*Reynolds and Zebchuk*, 1996], to uniform texture and pore size distribution developed by long-term, no-tillage agricultural practices [*Mohanty and Mousli*, 2000], or to intrinsic variations in soil texture and mineralogy [*Shukla et al.*, 2004]. The absence of spatial autocorrelation is presumably caused by biogenic macroporosity, which overrides possible texture and topography dependence [*Sobieraj et al.*, 2002], by uneven breaking of soil structure due to freezing and thawing [*Mohanty et al.*, 1991], or by surface tillage in the upper horizons [*Bosch and West*, 1998]. Hence it appears that a rather uniform soil texture and a stable soil structure that does not undergo frequent disturbance may favor the development of a predictable spatial structure of saturated hydraulic conductivity. This interpretation applies to our study as well, because the permeability of the surface horizons, which have experienced disturbances such as the compaction by cattle treading, exhibit less spatial structure than deeper, unaffected horizons.

7. Conclusions

[67] We first summarize our results by answering the research questions posed in the introduction.

[68] 1. Pronounced trends of the mean permeability exist at every investigated soil depth, with particular importance for the shallow soil. Ignoring it would produce misleading results.

[69] 2. Robust estimation techniques are not required, since the log-transformed data do not exhibit extreme values in their univariate and bivariate distributions. This was further confirmed using the approach of the θ-statistic [*Lark*,

2000a] to assess the need for robust estimators. The arbitrary use of robust estimators incurs the risk of lower efficiency.

[70] 3. Anisotropic variations are not evident. The seeming anisotropy in the non-detrended data associates the trend component with an imaginary direction-dependent spatial structure.

[71] 4. The estimated covariance parameters depend on the estimation method and, in the case of least-squares approaches, also on the chosen experimental variograms.

[72] 5. The differences between the Matérn and standard models are rather small. It was not always possible to estimate the parameters of the Matérn function.

[73] 6. Fixing the nugget variance potentially modifies all parameters of the covariance function and may therefore have an impact on spatial prediction. This should be investigated further. In some cases, the estimated nugget variance appeared to be too small.

[74] 7. As to prediction, the model differences do not influence average values, e.g., the average probability that rainfall intensities exceed the soil permeability. The spatial arrangement, however, particularly the connectivity of zones with large or small *Ks*, depends on the model used, and local differences in kriging maps can be considerable.

[75] Generalizations regarding the spatial structure of saturated hydraulic conductivity as a function of attributes such as soil type, land use, and measurement techniques are currently impossible. Limited earlier interpretations and our results, however, suggest a dependence of its spatial structure on land use-controlled soil conditions.

[76] Our results permit some general recommendations. First, the discussion of a plausible estimate of the nugget variance emphasizes the advantage of a sampling design which allows for calculating the semivariance at short lags. Second, under bivariate normal conditions we encourage estimation by restricted maximum likelihood, because it does not require the arbitrary selection of lag classes. The mean of the θ-statistic for the REML models, which is very close to the expectation, and their plausible effective range further support their accuracy. If one chooses a robust estimation technique, the variogram model should only be fitted over half of the linear dimension of the maximum lag to avoid edge effects. Third, in the framework of REML estimation, we propose to use the Matérn correlation function when possible and to compare it to some standard model (e.g., spherical), e.g., by means of the Akaike information criterion.

[77] Figure 13 summarizes our recommendations for the geostatistical analysis of environmental data. This flowchart should be continuously updated to accommodate new developments.

[78] **Acknowledgments.** We thank the German Research Foundation (DFG) for funding the project (FOR 402, TP B3, El 255/4-1), the Fundación Científica San Francisco (Nature and Culture International (NCI)) for providing access to the study area, D. Hodapp, D. Reetz and C. Leiberg for field assistance, T. Peters for providing climatic data, and R.M. Lark for stimulating discussions. The comments of two anonymous reviewers helped to improve the manuscript considerably.

References

Amoozegar, A. (1989a), A compact, constant-head permeameter for measuring saturated hydraulic conductivity of the vadose zone, *Soil Sci. Soc. Am. J.*, *53*, 1356–1361.

Amoozegar, A. (1989b), Comparison of the Glover solution with the simultaneous equations approach for measuring hydraulic conductivity, *Soil Sci. Soc. Am. J.*, *53*, 1362–1367.

Amoozegar, A. (1993), Comments on "Methods for analyzing constant-head well permeameter data", *Soil Sci. Soc. Am. J.*, *57*, 559–560.

Barnes, R. J. (1991), The variogram sill and the sample variance, *Math. Geol.*, *23*, 673–678.

Bloeschl, G. (1999), Scaling issues in snow hydrology, *Hydrol. Proc.*, *13*, 2149–2175.

Bosch, D. D., and L. T. West (1998), Hydraulic conductivity variability for two sandy soils, *Soil Sci. Am. J.*, *62*, 90–98.

Box, G. E. P., and D. R. Cox (1964), An analysis of transformations, *J. R. Stat. Soc., Ser. B*, *26*, 211–246.

Bruijnzeel, L. A., and L. S. Hamilton (2000), Up in the clouds, decision time for cloud forests, *IHP Humid Tropics Progr. Ser.*, *13*, IHP-UNESCO, Paris.

Buttle, J. M., and D. A. House (1997), Spatial variability of saturated hydraulic conductivity in shallow macroporous soils in a forested basin, *J. Hydrol.*, *203*, 127–142.

Cambardella, C. A., T. B. Moorman, J. M. Novak, T. B. Parkin, D. L. Karlen, R. F. Turco, and A. E. Konopka (1994), Field-scale variability of soil properties in central Iowa soils, *Soil Sci. Soc. Am. J.*, *58*, 1501–1511.

Chilès, J.-P., and P. Delfiner (1999), *Geostatistics: Modeling Spatial Uncertainty*, John Wiley & Sons, New York.

Cressie, N. A. C. (1993), *Statistics for Spatial Data*, John Wiley & Sons, New York.

Cressie, N., and D. Hawkins (1980), Robust estimation of the variogram, *Math. Geol.*, *12*, 115–125.

Dowd, P. A. (1984), The variogram and kriging: Robust and resistant estimators, in *Geostatistics for Natural Resources Characterization*, edited by G. Verly et al., pp. 91–106, D. Reidel, Dordrecht.

Duffera, M., J. G. White, and R. Weisz (2007), Spatial variability of southeastern U.S. coastal plain soil physical properties: Implications for site-specific management, *Geoderma*, *137*, 327–339.

Elrick, D. E., and W. D. Reynolds (1992), Methods for analyzing constant-head well permeameter data, *Soil Sci. Soc. Am. J.*, *56*, 320–323.

Fisher, R. A. (1972), *Statistical Methods for Research Workers*, Oliver and Boyd, Edinburgh.

Gascuel-Odoux, C., and P. Boivin (1994), Variability of variograms and spatial estimates due to soil sampling: A case study, *Geoderma*, *62*, 165–182.

Geiler, H., F. Aschenbrenner, H. S. Dengel, A. Donsbach, R. Ostermann, K. Knoblich, and W. Maurer (1997), Räumliche Variabilität und Verteilung von Schwermetallen in den Böden einer 1km^2 großen Modellfläche im Siegerland, *Z. Pflanzenernähr. Bodenkd.*, *160*, 603–612.

Genton, M. G. (1998), Highly robust variogram estimation, *Math. Geol.*, *30*, 213–221.

Goovaerts, P. (1997), *Geostatistics for Natural Resources Evaluation*, Oxford Univ. Press, New York.

Heuvelink, G. B. M., and R. Webster (2001), Modelling soil variation: Past, present, and future, *Geoderma*, *100*, 269–301.

Iqbal, J., J. A. Thomasson, J. N. Jenkins, P. R. Owens, and F. D. Whisler (2005), Spatial variability analysis of soil physical properties of alluvial soils, *Soil Sci. Soc. Am. J.*, *69*, 1338–1350.

Lark, R. M. (2000a), A comparison of some robust estimators of the variogram for use in soil survey, *Eur. J. Soil Sci.*, *51*, 137–157.

Lark, R. M. (2000b), Estimating variograms of soil properties by the method-of-moments and maximum likelihood, *Eur. J. Soil Sci.*, *51*, 717–728.

Lark, R. M. (2002), Modelling complex soil properties as contaminated regionalized variables, *Geoderma*, *106*, 173–190.

Lark, R. M., B. R. Cullis, and S. J. Welham (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.

Leuangthong, O., and C. V. Deutsch (2004), Transformation of residuals to avoid artifacts in geostatistical modelling with a trend, *Math. Geol.*, *36*, 287–305.

Lophaven, S., J. Carstensen, and H. Rootzen (2002), Methods for estimating the semivariogram, in *Symposium i Anvendt Statistik*, pp. 128–144, Institut for Informationsbehandling, Handelshøjskolen i Århus.

Mallants, D., B. P. Mohanty, D. Jacques, and J. Feyen (1996), Spatial variability of hydraulic properties in a multi-layered soil profile, *Soil Sci.*, *161*, 167–181.

Marchant, B. P., and R. M. Lark (2004), Estimating variogram uncertainty, *Math. Geol.*, *36*, 867–898.

Marchant, B. P., and R. M. Lark (2007), The Matérn variogram model: Implications for uncertainty propagation and sampling in geostatistical surveys, *Geoderma*, *140*, 337–345.

Matheron, G. (1962), *Traité de Géostatistique Appliqué, Tome 1*, Memoires du Bureau de Recherches Géologiques et Minières, Paris.

Minasny, B., and A. B. McBratney (2005), The Matérn function as a general model for soil variograms, *Geoderma*, *128*, 192–207.

Mohanty, B. P., and Z. Mousli (2000), Saturated hydraulic conductivity and soil water retention properties across a soil-slope transition, *Water Resour. Res.*, *36*, 3311–3324.

Mohanty, B. P., R. S. Kanwar, and R. Horton (1991), A robust-resistant approach to interpret spatial behavior of saturated hydraulic conductivity of a glacial till soil under no-tillage system, *Water Resour. Res.*, *27*, 2979–2992.

Motzer, T., N. Munz, M. Kuppers, D. Schmitt, and D. Anhuf (2005), Stomatal conductance, transpiration and sap flow of tropical montane rain forest trees in the southern Ecuadorian Andes, *Tree Physiol.*, *25*, 1283–1293.

Pardo-Igúzquiza, E. (1998), Maximum likelihood estimation of spatial covariance parameters, *Math. Geol.*, *30*, 95–108.

Pebesma, E. J. (2004), Multivariable geostatistics in S: The gstat package, *Comput. Geosci.*, *30*, 683–691.

R Development Core Team (2004), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

Reynolds, W. D., and W. D. Zebchuk (1996), Hydraulic conductivity in a clay soil: Two measurement techniques and spatial characterization, *Soil Sci. Soc. Am. J.*, *60*, 1679–1685.

Ribeiro, P. J., and P. J. Diggle (2001), GeoR: A package for geostatistical analysis, *R-NEWS*, 1(2). (Available at http://cran.r-project.org/doc/Rnews)

Rollenbeck, R., J. Bendix, P. Fabian, J. Boy, H. Dalitz, P. Emck, M. Oesker, and W. Wilcke (2007), Comparison of different techniques for the measurement of precipitation in tropical montane rain forest regions, *J. Atmos. Ocean. Technol.*, *24*, 156–168.

Rousseeuw, P. J., and C. Croux (1993), Alternatives to the median absolute deviation, *J. Am. Stat. Assoc.*, *88*, 1273–1283.

Russo, D., I. Russo, and A. Laufer (1997), On the spatial variability of parameters of the unsaturated hydraulic conductivity, *Water Resour. Res.*, *33*, 947–956.

Schrumpf, M., G. Guggenberger, C. Valarezo, and W. Zech (2001), Tropical montane rain forest soils, development and nutrient status along an altitudinal gradient in the south Ecuadorian Andes, *Die Erde*, *132*, 43–60.

Shukla, M. K., B. K. Slater, R. Lal, and P. Cepuder (2004), Spatial variability of soil properties and potential management classification of a chernozemic field in lower Austria, *Soil Sci.*, *169*, 852–860.

Skøien, J. O., and G. Bloeschl (2006), Scale effects in estimating the variogram and implications for soil hydrology, *Vadose Zone J.*, *5*, 153–167.

Sobieraj, J. A., H. Elsenbeer, R. M. Coelho, and B. Newton (2002), Spatial variability of soil hydraulic conductivity along a tropical rainforest catena, *Geoderma*, *108*, 79–90.

Sobieraj, J. A., H. Elsenbeer, and G. Cameron (2004), Scale dependency in spatial patterns of saturated hydraulic conductivity, *Catena*, *55*, 49–77.

Soil Survey Staff (1999), *Soil Taxonomy—A Basic System of Soil Classification for Making and Interpreting Soil Surveys*, 2nd ed., US Government Printing Office, Washington, D. C.

Stein, M. L. (1999), *Interpolation of Spatial Data: Some Theory for Kriging*, Springer, New York.

Unlu, K., D. R. Nielsen, J. W. Biggar, and F. Morkoc (1990), Statistical parameters characterizing the spatial variability of selected soil hydraulic properties, *Soil Sci. Soc. Am. J.*, *54*, 1537–1547.

Webster, R., and M. A. Oliver (1992), Sample adequately to estimate variograms of soil properties, *J. Soil Sci.*, *43*, 177–192.

Webster, R., and M. A. Oliver (2001), *Geostatistics for Environmental Scientists*, John Wiley & Sons, Chichester, U.K.

Western, A. W., and G. Bloeschl (1999), On the spatial scaling of soil moisture, *J. Hydrol.*, *217*, 203–224.

Wilcke, W., H. Valladarez, R. Stoyan, S. Yasin, C. Valarezo, and W. Zech (2003), Soil properties on a chronosequence of landslides in montane rain forest, Ecuador, *Catena*, *53*, 79–95.

H. Elsenbeer and N. K. Hartmann, Institute of Geoecology, University of Potsdam, Karl-Liebknecht-Strasse 24-25, D-14476 Potsdam, Germany.

E. Zehe, Institute of Water and Environment, Technische Universitaet Muenchen, Arcisstrasse 21, D-80333 Munich, Germany.

B. Zimmermann, Smithsonian Tropical Research Institute, Center for Tropical Forest Science, Apartado Postal 0843-03092, Balboa, Ancón, Panama. (zimmermannb@si.edu)