Journal of Hydrology (2007) 334, 405–422



available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jhydrol



Uncertainty in predicting soil hydraulic properties at the hillslope scale with indirect methods

G.B. Chirico [a,*], H. Medina [b], N. Romano [a]

[a] Department of Agricultural Engineering, University of Naples Federico II, 80055 Portici, Naples, Italy
[b] Agrarian University of Havana, Havana, Cuba

Received 10 March 2006; received in revised form 3 October 2006; accepted 12 October 2006

KEYWORDS
Unsaturated soil;
Soil hydraulic properties;
Pedotransfer function;
Spatial uncertainty;
Sequential stochastic simulation

Summary Several hydrological applications require the characterisation of the soil hydraulic properties at large spatial scales. Pedotransfer functions (PTFs) are being developed as simplified methods to estimate soil hydraulic properties as an alternative to direct measurements, which are unfeasible for most practical circumstances. The objective of this study is to quantify the uncertainty in PTFs spatial predictions at the hillslope scale as related to the sampling density, due to: (i) the error in estimated soil physico-chemical properties and (ii) PTF model error. The analysis is carried out on a 2-km-long experimental hillslope in South Italy. The method adopted is based on a stochastic generation of patterns of soil variables using sequential Gaussian simulation, conditioned to the observed sample data. The following PTFs are applied: Vereecken's PTF [Vereecken, H., Diels, J., van Orshoven, J., Feyen, J., Bouma, J., 1992. Functional evaluation of pedotransfer functions for the estimation of soil hydraulic properties. Soil Sci. Soc. Am. J. 56, 1371–1378] and HYPRES PTF [Wösten, J.H.M., Lilly, A., Nemes, A., Le Bas, C., 1999. Development and use of a database of hydraulic properties of European soils. Geoderma 90, 169–185]. The two PTFs estimate reliably the soil water retention characteristic even for a relatively coarse sampling resolution, with prediction uncertainties comparable to the uncertainties in direct laboratory or field measurements. The uncertainty of soil water retention prediction due to the model error is as much as or more significant than the uncertainty associated with the estimated input, even for a relatively coarse sampling resolution. Prediction uncertainties are much more important when PTF are applied to estimate the saturated hydraulic conductivity. In this case model error dominates the overall prediction uncertainties, making negligible the effect of the input error.
© 2006 Elsevier B.V. All rights reserved.

* Corresponding author. Tel.: +39 081 2539415; fax: +39 081 2539412.
E-mail address: gchirico@unina.it (G.B. Chirico).

0022-1694/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhydrol.2006.10.024

Introduction

An effective representation of soil hydraulic properties and their spatial variability is of prime importance for hydrological studies, especially those applied to environmental and landuse planning problems. This is a difficult task for most practical circumstances, since soil properties show high spatial variability, which is difficult to assess by deterministic relationships. Accurate laboratory and field direct methods are currently available for measuring soil hydraulic properties at the core and plot scales, but they are costly and time consuming, thus turning out to be unfeasible when the analysis spans over larger land areas. A few studies have suggested to address this issue by integrating direct observations with Geostatistics (e.g. Burgess and Webster, 1980; Romano, 1993), but yet the experimental effort required is not acceptable for most practical circumstances.

These difficulties can be overcome by employing pedotransfer functions (PTFs) that estimate soil hydraulic properties from basic soil physical and chemical properties, more easy to measure. The majority of PTFs are regression equations derived from data collected during site-specific field campaigns. Several studies show that PTFs predict soil hydraulic properties at many positions over a region with acceptable precision compared to the cost of the investigations required for their application (e.g. Rawls and Brakensiek, 1989; Vereecken et al., 1992). More recent studies show that PTF prediction performance can be also improved by exploiting terrain data (Pachepsky et al., 2001; Romano and Palladino, 2002). The underlying idea moves from the consideration that terrain data, nowadays available at relatively low cost, may contain ancillary information on the soil variability, since topography exerts a significant control on hydrological, geomorphological and biological processes affecting soil formation (Jenny, 1980).

Despite these advantageous elements, one should ponder over the issues concerning the uncertainty associated to PTF predictions. Evaluating PTF prediction uncertainty can be fundamental for many hydrological applications, especially those whose results are employed to assess environmental risks and guide decision makers in planning mitigation strategies.

As in all model predictions, the overall PTF prediction uncertainty can be identified by analysing the propagation of three major sources of uncertainty (Sivapalan, 2003): (i) the input variables, (ii) the model structure, and (iii) the model parameters. The uncertainty in the input variables at the observation locations is explained by the measurement errors. As the prediction is accomplished further from the observation points, the uncertainty in the input variables due to local measurement errors is increased by the interpolation uncertainty.

Previous studies have addressed this issue, but principally focussed on the uncertainties in local PTF predictions. Vereecken et al. (1992) evaluated the errors in PTF predictions and their effects in the simulated soil water dynamics. Schaap and Leij (1998) considered the uncertainty in the PTF based on the accuracy of the soil database from which the PTF relations were defined and calibrated. Christiaens and Feyen (2001) used a Latin Hypercube sampling strategy to evaluate how the uncertainties in the predicted soil hydraulic properties (estimated with four different indirect methods) propagate into the output of a distributed hydrological model applied to a 1-km^2 catchment. Minasny et al. (1999) analysed the uncertainty in water retention prediction with Monte Carlo simulations and compared the effect that the input variables exerted on the uncertainty in the PTF model parameters. They showed that the uncertainty in the parameters is small compared to the uncertainty due to error in the input variables and this error affects the PTF predictions most significantly. Using a Latin Hypercube sampling strategy, Minasny and McBratney (2002) evaluated how measurement errors of the PTF input variables affects the uncertainty in PTF predictions and the soil water budget modelling. They concluded that small uncertainties in the input data can produce large uncertainty in the PTF predictions, and that the soil water budget is more sensitive to input uncertainties during dry periods rather than during wet periods.

From a practical perspective, it is also interesting to evaluate PTF prediction uncertainty at larger spatial scales, such as the scale of a hillslope or a small catchment, the scales at which PTFs are prime alternatives to direct measurements. This kind of analysis is consistent with the focus of many scientific programs, such as PUB (Prediction of Ungauged Basins) promoted by IAHS (Wagener et al., 2004), aiming at achieving advances in hydrological predictions for catchments lacking of direct observations, with emphasis on the estimation of predictive uncertainty and its subsequent reduction.

At the scale of a hillslope or a catchment, PTF input variables, i.e. basic soil physical and chemical properties, can be observed only in a limited number of locations. To make predictions at unsampled locations, PTFs have necessarily to be coupled with interpolation strategies. An issue being rarely addressed is the quantification of the effects exerted by the sampling density on the uncertainty in PTFs spatial predictions at the hillslope scale, taking the spatial variability of the input soil properties and PTF model error (due to both model structure and parameter uncertainty) into account. For this kind of analysis to be carried out, one needs a tool for estimating soil properties at unsampled locations and the corresponding probability distributions conditioned to a limited number of sample locations.

The techniques most commonly applied for assessing estimation uncertainty at unsampled locations relies on the kriging theory for spatially correlated and second order stationary random variables. Kriging looks for obtaining an unbiased estimation with the minimum estimation variance at each unsampled locations, by a linear combination of the observed values. The local uncertainty is defined by a Gaussian distribution centred on the kriging estimate and variance equal to the kriging variance. This uncertainty model is subjected to stringent hypotheses that not always are met in practice. Moreover, kriging reproduces a single pattern of the estimated variable, which is generally much smoother than a single possible realisation. The dispersion of the estimated values toward the extremes tends to be reduced, producing an asymmetry in the error distribution in the locations with low or high values. The spatial smoothing is also not uniform, being dependent on local data patterns (Goovaerts, 2000).

An alternative approach is the stochastic simulation (Goovaerts, 2000). Unlike kriging, stochastic simulations aim to produce many equi-probable realisations of an estimating variable, each of these reflecting as close as possible the spatial statistics (e.g. histograms, semivariograms) of the random field, rather than minimising a local error variance. Stochastic simulation is more suited for evaluating how the uncertainty on the spatial distribution of the input variables propagates through a non-linear model. By means of a set of equi-probable realisations, a non-parametric measure of the error can be derived as related to non-linear models. In the case of the PTFs, sets of multiple patterns, i.e. the soil physico-chemical properties, have to be generated. In principle, each input variable is not only spatially autocorrelated, but it can be also spatially correlated with other variables. Therefore, a set of equi-probable realisations of multiple variables has to reproduce the global multivariate statistics of the soil physico-chemical properties, including both their semivariograms and cross-variograms.

Using soil physical and hydraulic data collected along an experimental hillslope located in South Italy, the objective of the present study is to assess the prediction uncertainty of two different PTFs assuming that the soil spatial statistics are known, i.e. the effect of the bias in the estimated statistics induced by the different sampling density is not considered. The study does not consider the uncertainty due to the measurement errors. Measurement errors can be assumed spatially uncorrelated and therefore can be addressed by local scale analysis (Minasny and McBratney, 2002). The PTF spatial prediction is determined by a stochastic generation of soil patterns, conditioned to sampled points. The following section illustrates the overall experimental framework, including both the multivariate stochastic generation technique and the uncertainty analysis. Then the study site and the sample data are presented and analysed. The last section is devoted to the discussion and the conclusions.

Method for multivariate stochastic generations and uncertainty analysis

Fig. 1 describes the overall experimental framework, which is designed for the multivariate stochastic generations of soil physico-chemical properties and the uncertainty analysis carried out on the PTF spatial predictions. Data collected along an experimental hillslope transect, together with terrain attributes taken as ancillary information, are employed to identify the spatial structure of the PTF input variables. Hereafter, the PTF input variables are named ''basic soil properties'', while an ensemble of patterns representing the basic soil properties across the hillslope is named ''set of basic soil properties''. A multivariate geostatistical analysis is carried out to identify deterministic trends, semivariograms and cross-variograms of the basic soil properties. A multivariate stochastic simulation is then performed for generating high-resolution set of basic soil properties. This set is taken as an exhaustive representation of the soil

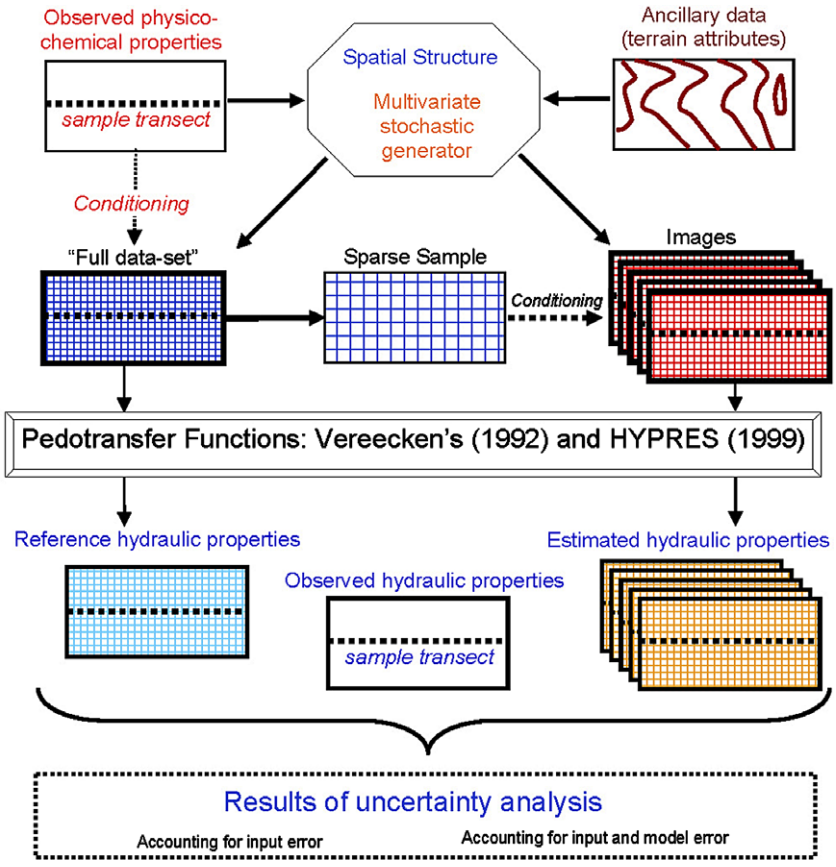


Figure 1 Schematic diagram describing the procedure applied for uncertainty analysis.

variability across the study hillslope and hereafter it is refereed to as the ''full dataset''. The full dataset is taken as the true representation of the hillslope soil variability and as reference data for the following uncertainty analysis.

Several algorithms have been proposed for carrying out stochastic simulations. Stochastic simulations can be unconditioned, i.e. they do not allow for any observed data, or conditioned so that they preserve the observed values in the sampled locations. The choice of one technique or the other depends chiefly on the objective of the stochastic simulation and, obviously, on the data availability. Since the major aim of this study is to quantify the uncertainty in PTF prediction due to both PTF model error and input estimation error from sparse samples, the ''full dataset'' is generated as conditioned to the values observed across the hillslope transect.

The conditional simulation was performed by employing the Gstat software package on S (R/S Plus) (Pebesma, 2004), which complements most of the features conceived in the original Gstat stand-alone program (Pebesma and Wesseling, 1998) with the graphical capabilities of S. The generation of independent realisations in the Gstat package is carried out from a sequential simulation algorithm (Gómez-Hernández and Journel, 1993). In this algorithm each node of the grid where generations take place is visited once following a random path. The new vector of values in a point is produced by randomly drawing a conditional distribution obtained from a co-kriging interpolation which employs a user-defined number of nearest sample data and previously simulated set of values. The application of a Gaussian simulation requires that the conditional probability distribution function of the variables of interest is a Gaussian multivariate random function. Therefore, each variable of interest must follow a normal distribution. If this is not the case, then the variable has to be transformed into a normally distributed variable prior to apply the sequential simulation algorithm.

Recent applications, involving the estimation of soil properties, also applied sequential simulation algorithms. Castrignanó and Buttafuoco (2004) evaluated the probability that soil water content in unsampled points did not exceed a critical threshold in a forested area. Zhao et al. (2005) assessed the uncertainty of spatial patterns of soil organic carbon density.

As depicted in Fig. 1, the full dataset is sampled and multivariate stochastic simulations are then performed to generate sets of basic soil properties conditioned to these sparse samples. Each generated set of basic soil properties is named as ''image'' of the ''full dataset''.

The full dataset and its images of basic soil properties are translated into patterns of soil hydraulic properties by applying a specific PTF. The full dataset provides the reference patterns of soil hydraulic properties. The images of the full dataset provide the estimated set of hydraulic properties. The comparison of this reference pattern to the estimated set of hydraulic properties enables the evaluation of the PTF prediction uncertainty.

The relative importance of PTF model structure and parameter uncertainty with respect to the input uncertainty, within the overall PTF prediction uncertainty, is evaluated by comparing the reference hydraulic properties and the estimated set of hydraulic properties to the observed hydraulic properties along the transect.

Experimental hillslope and soil dataset

The study area is a hillslope at the west side of the Fiumarella di Corleto River catchment in the Basilicata region, Southern Italy (see Fig. 2). The hillslope spans for about 2200 m, from the catchment ridge at 1200 m above mean sea-level to the right riverbank at 770 m above mean sea-level. The average slope is 16% and the average aspect is 100° clockwise to the north (south-east direction). From shoulder to toeslope of the study hillslope, the following main soil-landscape units have been identified: rounded ridges on clay and marl rocks, slopes on arenaceous clay flysch, and slightly unstable slopes on clay flysch. Landuse consisted mainly of broadleaf tree and conifer wood.

Undisturbed soil cores were sampled 50 m apart, along a linear transect 2200 m long, using steel cylinders (7.2 cm in inner diameter and 7.0 cm in length), at depths from 5 to 12 cm, running from the top of the hillslope to the river bank. Spatial coordinates of the sampled locations were determined with a GPS with an average accuracy of 4.5 m. Elevation was determined from a 1-m grid DEM recently obtained by a Laser-Scan technology for the entire Fiumarella catchment. This sample set is part of a larger dataset collected in the same area (Romano and Palladino, 2002). All cores were subjected to laboratory measurements to determine the particle-size distribution, organic carbon content (OC), and soil hydraulic characteristics. The cores were oven-dried at 105 °C to determine dry bulk density (BD). Total porosity was calculated from the measured oven-dry

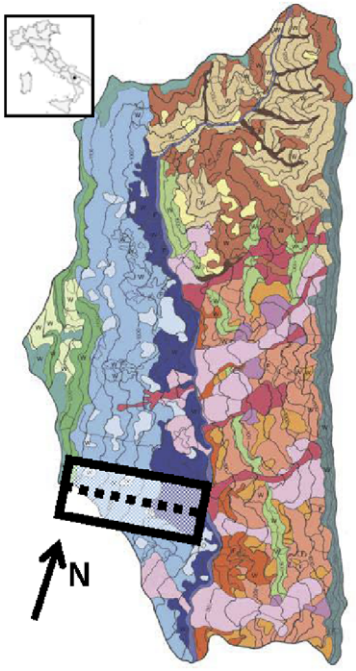


Figure 2 Soil-landscape map of the Fiumarella di Corleto catchment (Basilicata region, south Italy) with the study transect (dots) and hillslope (rectangle) highlighted.

bulk density and assuming a particle density of 2.65 g/cm^3. Organic carbon, OC, in soil was determined with the dichromate method, whereas organic matter content, OM, was calculated by multiplying the organic carbon content by 1.724. Sand (SA), silt (SI), and clay (CL) contents were expressed as percentage by mass of the fine-earth fraction (0.2 mm). Table 1 summarizes the statistics of the measured soil physical and chemical properties along the transect. Drying soil water retention data points were measured at several matric pressure heads using a specifically designed suction table apparatus (Romano and Palladino, 2002). To allow comparisons between PTF predicted and measured soil hydraulic properties, the water retention characteristics observed for each soil core are assumed to be described by van Genuchten's closed-form relationship:

$$\left(\frac{\theta-\theta_r}{\theta_0-\theta_r}\right)=[1+|\alpha h|^n]^{-m} \quad (1)$$

where θ (cm^3 cm^{-3}) is volumetric water content and h is matric pressure head (in cm). Parameter θ_0 is water content at $h=0$ (often assumed equal to the measured saturated water content, θ_s), and θ_r is the residual water content, i.e. the water content when h tends to $-\infty$. The van Genuchten parameters α (>0, in cm^{-1}), n ($n>1$), and m control the shape of the retention curve. The unknown parameters θ_0, θ_r, α, n and m were estimated by fitting Eq. (1) to the measured water retention points.The fit was performed with three hypotheses concerning the parameter m: (i) $m=1$, as assumed by Vereecken et al. (1989); (ii) $m=1-1/n$, according to the classic constraint for van Genuchten's closed-form relationship; (iii) no constraints for m. No significant differences were found in the quality of fit in the three cases. Therefore it was chosen to impose the classic constraint $m=1-1/n$ and to estimate the remaining unknown four parameters θ_0, θ_r, α, and n.

Saturated hydraulic conductivity, K_s (in cm day^{-1}), was measured on the undisturbed soil cores by the falling head method (Reynolds and Elrick, 2002).

Soil water content, θ, at h of -10, -100, and $-10{,}000$ cm as well as saturated hydraulic conductivity, K_s, are taken as target variables for the following PTF prediction uncertainty analysis. Table 2 summarizes the statistics of these variables as computed from the observed values along the hillslope transect.

Table 1 Statistics of the soil physico-chemical properties observed along the transect

Statistics	OM	SA	SI	CL	BD
$P_{0.95}$	0.97	38.79	50.01	43.69	1.38
$P_{0.05}$	0.60	14.18	36.62	16.80	1.04
μ	0.75	24.48	43.48	32.04	1.19
CV (%)	15.7	36.7	9.48	27.1	9.01

$P_{0.95}$ = 95th percentile; $P_{0.05}$ = 5th percentile; μ = mean; CV = coefficient of variation.
OM, organic matter (log(%)); SA, sand (%); SI, silt (%); CL, clay (%); BD, bulk density (g/cm^3).

Table 2 Statistics of the selected soil hydraulic properties observed along the transect

Statistics	θ_{-10}	θ_{-100}	$\theta_{-10,000}$	Log[K_s]
$P_{0.95}$	0.540	0.489	0.292	2.73
$P_{0.05}$	0.418	0.325	0.167	1.10
μ	0.482	0.418	0.240	1.99
CV (%)	8.32	12.6	16.5	32.2

$P_{0.95}$ = 95th percentile; $P_{0.05}$ = 5th percentile; μ = mean; CV = coefficient of variation; θ_h, soil water content (cm^3/cm^3) at $h=-10$, -100, and $-10{,}000$ cm soil matric pressure head (cm^3/cm^3); Log[K_s] decimal logarithm of the saturated hydraulic conductivity (Log(cm/d)).

Analysis of the spatial structure of the basic soil properties

Indirect estimation of soil hydraulic properties via pedotransfer functions requires soil physical and chemical properties as input. They are referred to in this paper as the ''basic soil properties''. The properties most commonly employed are bulk density (BD), percentages of sand (SA), silt (SI), and clay (CL), mostly according to the USDA classification, and organic matter content (OM). As said before, an ensemble of patterns representing the basic soil properties across the hillslope is a ''set of basic soil properties''.

To generate sets of basic soil properties, a model of their spatial variability has to be estimated. For multivariate geostatistical simulations, means, variances, semivariograms and cross-variograms of the simulated variables are required. These statistics are assumed to be not dependent upon sampling location (condition for second order stationarity). However, the hypothesis of second order stationarity does not generally apply to basic soil properties as they are. Basic soil properties exhibit significant trends in the structure of their spatial variability. These trends can be explained by the effects of those environmental factors influencing soil formation (Jenny, 1980). Different environmental factors control soil variability at different spatial scales (Romano and Chirico, 2004) and several studies have shown that basic soil properties are significantly correlated with terrain attributes at the hillslope scale (e.g. Brubacker et al., 1994; Florinsky and Arlashina, 1998). Romano and Palladino (2002) found significant correlation between soil water retention values and terrain attributes along different hillslope transects of the Fiumarella di Corleto River. This suggested performing a preliminary correlation analysis between basic soil properties and terrain attributes so as to identify the dominant deterministic spatial trends in the basic soil properties.

Terrain attributes have been calculated from the available 1-m grid DEM, smoothed with 9-m $\times$ 9-m moving window, given the accuracy of 4.5 m with which the sample locations were surveyed. A numerical scheme built on a 3×3 centred grid (Moore et al., 1993) has been employed for calculating the following primary terrain attributes: slope (Sl = tan(β), where β is the slope angle), aspect (As = cos(ϕ), where ϕ is the angle clockwise from north), plane curvature (PlCv), profile curvature (PrCv), curvature (Cv).

Specific contributing area (SCA) has been calculated using the terrain analysis software suite provided by Tarboton (1997). This software utilizes a specifically designed algorithm named D∞, that among other things identifies lateral connectivity after spurious pits are removed from the grid DEM. We compute the wetness index (WI) as the ratio of Sl to SCA, in order to avoid the occurrence of an indefinite form (infinite number) at locations with zero slope. Slope, organic matter, and the wetness index have been log-transformed in order to satisfy the normality condition, prior calculating the linear correlation coefficient. Table 3 summarizes the statistics of the computed terrain attributes.

Table 4 shows the linear correlation coefficients calculated among the selected terrain attributes and the basic soil properties. Correlation values with a significance level equal to 0.01 are given in bold. The slope explains 30% of the clay variability and 22% of the sand variability, whereas elevation explains 15% of the bulk density variability. Silt exhibits strong correlation with plane curvature (10%), wetness index (20%), and specific contributing area (15%). A Spearman correlation test has been performed to assess whether the linear correlation coefficients are biased by the limited number of sample data. As listed in Table 5, the results are similar, thus confirming the robustness of the data reported in Table 4, as well as the linearity of the trends between auxiliary data and analysed properties.

A correlation analysis has been also applied to assess possible linear trends among the basic soil properties and relevant results are summarized in Tables 4 and 5. The most interesting result is that clay and sand exhibit a strong linear correlation (80%).

Based on the results of the correlation analyses illustrated above, the spatial variability of the basic soil properties has been modelled as follows. A linear regression model has been adopted to describe part of the clay content spatial variability as deterministic function of the terrain slope. Similarly, a linear regression model has been adopted to explain part of the spatial variation of the bulk density as deterministic function of the elevation. It is also assumed that the spatial variation of the sand content can be fully described by a linear regression model with respect to the clay content. Consequently, the silt content, which is the complement to one with respect to the clay and the sand contents, is also totally explained by a linear function of the clay content. The following three variables have been examined within a stochastic framework: (i) the regression residuals of clay content with respect to slope angle; (ii) the regression residuals of bulk density with respect to elevation above sea-level; and (iii) the decimal logarithm of organic matter content. These variables satisfy the normality

Table 3 Statistics of the terrain attributes computed at the soil sampling location along the transect

	$P_{0.95}$	$P_{0.05}$	μ	St. Dev. (%)
El	1183.66	827.40	1011.02	118.89
As	332.30	9.40	91.58	96.31
Sl	1.74	1.07	1.39	0.21
PrCv	2.41	−3.13	0.08	1.77
PlCv	2.82	−3.22	0.04	1.99
Cv	4.30	−4.76	−0.04	3.33
SCA	403.44	1.00	73.85	153.09
WI	1.44	−1.36	0.11	0.87

El, elevation (m a.s.l.); As, aspect (cos(ϕ)); Sl, slope (log(tan(β)%)); PrCv, profile curvature (1/m); PlCv, plane curvature (1/m); Cv, curvature (1/m); SCA, specific contributing area (m); WI (log(tan(β)%/SCA)).

Table 4 Pearson (parametric) correlation between soil properties and terrain attributes

	OM	SA	SI	CL	BD
El	−0.269	−0.073	−0.281	0.139	**0.386**
As	0.287	0.218	0.010	−0.182	−0.265
Sl	0.083	**0.470**	0.124	**−0.555**	−0.190
PrCv	−0.191	0.159	−0.098	−0.099	0.189
PlCv	0.219	0.168	**0.327**	−0.220	−0.227
Cv	0.180	−0.058	0.251	−0.008	−0.162
SCA	−0.158	0.154	**−0.387**	−0.094	−0.009
WI	0.259	−0.162	**0.440**	−0.059	**−0.308**
OM	1.000	**0.362**	0.299	**−0.397**	**−0.462**
SA		1.000	−0.115	**−0.883**	−0.047
SI			1.000	−0.195	**−0.429**
CL				1.000	0.120
BD					1.000

In bold, correlation values with a significance level equal to 0.01. The marked cells indicate the combination of variables for which linear regression model is employed.
OM, organic matter (log(%)); SA, sand (%); SI, silt (%); CL, clay (%); BD, bulk density (g/cm^3); El, elevation (m a.s.l.); As, aspect (cos(ϕ)); Sl, slope (log(tan(β)%)); PrCv, profile curvature (1/m); PlCv, plane curvature (1/m); Cv, curvature (1/m); SCA, specific contributing area (m); WI (log(tan(β)%/SCA)).

Table 5 Spearman (non-parametric) correlation between soil properties and terrain attributes

	OM	SA	SI	CL	BD
El	−0.161	−0.066	−0.262	0.175	**0.407**
As	0.038	**0.505**	0.068	**−0.592**	−0.159
Sl	−0.311	0.118	−0.156	−0.027	0.183
PrCv	0.210	0.113	**0.300**	−0.201	−0.161
PlCv	0.211	−0.042	0.270	−0.091	−0.120
Cv	0.337	0.263	−0.032	−0.156	−0.141
SCA	0.390	−0.129	**0.486**	−0.079	**−0.337**
WI	−0.301	0.211	**−0.332**	−0.141	0.016
OM	1.000	**0.300**	0.310	**−0.333**	**−0.421**
SA		1.000	−0.093	**−0.870**	−0.040
SI			1.000	−0.212	**−0.398**
CL				1.000	0.111
BD					1.000

In bold, correlation values with a significance level equal to 0.01.
OM, organic matter (log(%)); SA, sand (%); SI, silt (%); CL, clay (%); BD, bulk density (g/cm^3); El, elevation (m a.s.l.); As, aspect (cos(ϕ)); Sl, slope (log(tan(β)%)); PrCv, profile curvature (1/m); PlCv, plane curvature (1/m); Cv, curvature (1/m); SCA, specific contributing area (m); WI (log(tan(β)%/SCA)).

condition required by multi-Gaussian simulations. Plots a, b, and c of Fig. 3 show the relevant experimental semivariograms. A structure of spatial variability is detected in the cross-correlation between residual of clay content and the decimal logarithm of organic matter content. Fig. 3d shows the corresponding experimental cross-variogram. No significant structure was found in the remaining two combinations of selected variables. The sample variograms of Fig. 3, particularly the one of the logarithm of organic matter content, show some hole-effect for which there is no evident explanation. As an in-depth understanding of the observed spatial variations is not the scope of this paper (only part of the entire dataset collected up to now), the hole-effect is neglected in the following analysis, treated as noise of the study dataset. The experimental semivariogram and cross-variogram have been interpolated with spherical models. Clay content residuals and the logarithmic transform of organic matter content show a range of 500 m, while bulk density residuals show a range of 250 m. It is important to note that the semivariograms and cross-variograms of the two cross-correlated variables, clay content residuals and the logarithmic of the organic matter content, are forced to have the same range. This restriction is due to the co-simulation algorithm implemented in the Gstat software, which requires the covariance model of two correlated random fields to be positive definite. This assumption is common to many Gaussian co-simulation algorithms, although less restrictive models have been recently proposed in literature (Oliver, 2003).

Multivariate stochastic simulation of sets of basic soil properties

Following the results illustrated in the previous section, sets of basic soil properties can be generated by multivariate stochastic simulation of the three selected variables (respectively clay content regression residuals, bulk density regression residuals and logarithmic of the organic matter content) and by adding to these three simulated variables the deterministic components explained by the terrain attributes.

Twenty sets of basic soil properties have been firstly simulated conditioned to the sample data collected along the transect. As depicted in Fig. 4, the sample nodes of these patterns consist of the transect samples plus a combination of two sample grids, with size of 50 m and 100 m, respectively, not-centred one on the other. The grid nodes are 880 from the 50-m grid and 220 from the 100-m grid. All these simulated patterns show a good approximation to the targeted spatial structure. The one showing the best approximation has been considered as an exhaustive set of basic soil properties. This set is named hereafter as the full dataset.

The ''full dataset'' has been then sampled in the nodes of the 50 m grid, with three different sampling densities: 50 m, 100 m and 200 m. These sample densities have been chosen in order to keep the spacing among the sample points lower than the minimum correlation length observed in the experimental data (see Fig. 3). The set of basic soil

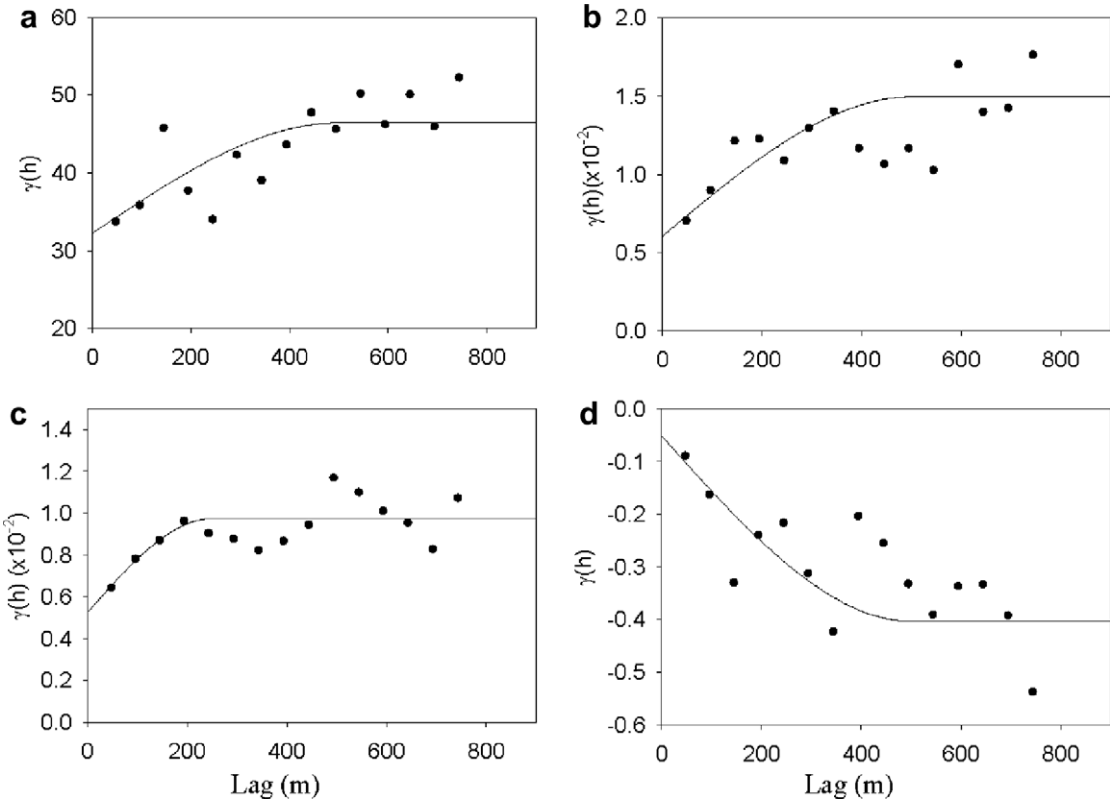


Figure 3 Semivariograms of (a) residual of the clay contents $(CL\%)^2$, (b) logarithmic of the organic matter $(\log(OM\%))^2$ and (c) residuals of bulk density $(BDg/cm^3)^2$; (d) represents the cross-semivariogram of the residual of clay and logarithmic of the organic matter (CL%log(OM%)).

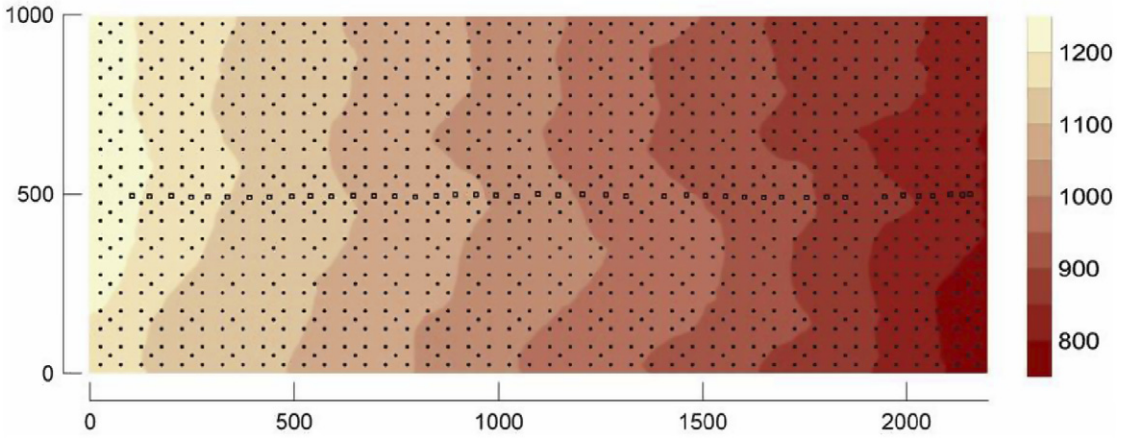


Figure 4 Sample nodes represented on a contour map of the elevation (m a.s.l., as in colour scale). The squares are the sample locations along the experimental transect.

properties, simulated conditionally to these sparse samples, is named as the ''image'' of the ''full dataset''. The images of the full data-set are generated by assuming that the spatial structure of the simulated variables is known and always equal to that defined for generating the full data-set. In other words, the effect of different sampling strategies on the estimated variograms (e.g. Western and Blöschl, 1999) is not included in the following uncertainty analysis.

Images of the full data-set have been simulated conditioned to three chosen sample networks. The simulated patterns include the transect locations plus: 265 nodes of the 100 m grid in the case of the 50 m sampling density; 660 nodes of the remaining 50 m grid nodes in the case of the 100 m sampling density; 825 nodes of the residual 50 m grid nodes for the 200 m sampling density. For each sampling density, 100 sets of multiple patterns of soil properties (clay regression residuals, bulk density regression residuals and organic matter) have been simulated conditioned to the corresponding data sampled from the full data-set. At the end, three sets of 100 multiple patterns of clay regression residuals, bulk density residuals and organic matter, have been generated, covering all sample nodes of the exhaustive data set. All these patterns show good agreement with the imposed spatial structure.

All these images have been employed for defining the basic properties required as input for PTF regression models. Clay and organic matter contents have been computed by adding to the corresponding stochastic residuals the deterministic component resulting from the application of the regression equations with the terrain variable. Fig. 5 shows an example of simulated pattern of clay. Patterns of silt and sand have been derived by linear regression with the patterns of clay.

In order to assess the uncertainty in the estimation of the basic soil properties (i.e. uncertainty of the input to the PTF) at the unsampled locations, the differences between the patterns of each simulated variables (respectively clay regression residuals, bulk density regression residuals and organic matter) and the corresponding patterns of the full dataset have been evaluated with the following statistics:

$$\mathrm{ME} = \frac{1}{N}\sum_{k=1}^{N}(v_{sk} - v_{fk}) \tag{2}$$

$$\mathrm{RMSE} = \left[\frac{1}{N}\sum_{k=1}^{N}(v_{sk} - v_{fk})^2\right]^{1/2} \tag{3}$$

where N is number of locations across the hillslope in which the soil property has been simulated, v_{sk} is the simulated soil property value in the kth location, v_{fk} is the corresponding value in the full data-set. ME is an indicator of accuracy of the single simulation, revealing the presence of the posi-

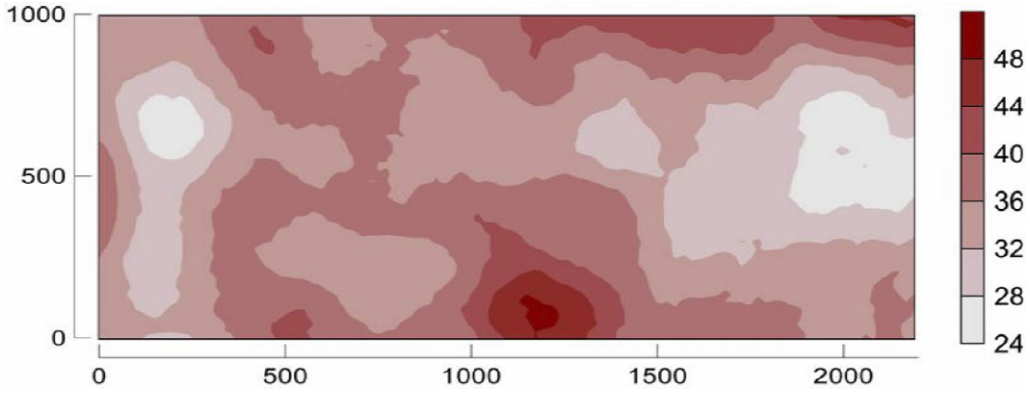


Figure 5 Example of generated pattern of clay (colour scale representing clay content in %).

tive or negative simulation bias, i.e. an average tendency for overestimation (if positive) or underestimation (if negative). RMSE is a measure of the overall performance of the single simulation, including both accuracy and precision error. The accuracy can be evaluated by the standard deviation of the simulation error:

$$\mathrm{SDE} = \left[\frac{1}{N-1}\sum_{i=1}^{N}(v_{sk} - v_{fk} - \mathrm{ME})^2\right]^{1/2} \quad (4)$$

whereas RMSE can be expressed as a combination of SDE and ME, as follows:

$$\mathrm{RMSE}^2 = \mathrm{ME}^2 + \frac{N-1}{N}\mathrm{SDE}^2 \quad (5)$$

For large N (as in this case), $\mathrm{RMSE}^2 \approx \mathrm{ME}^2 + \mathrm{SDE}^2$. Therefore, the ratio $\mathrm{ME}^2/\mathrm{RMSE}^2$ can be viewed as the relative contribution of the accuracy to the total simulation error within a given simulated pattern. The distributions of the 100 ME and RMSE values (one for each simulated image of the full dataset) is normal. The means (μ(ME), μ(RMSE)) and the standard deviations (σ(ME), σ(RMSE)) of these distributions are listed in Table 6a. These statistics provide an estimation of the relative contribution of the expected accuracy to the expected overall simulation error at the hillslope scale, after multiple simulations of spatial patterns of soil properties. In case of large N values, the following relation can be easily retrieved from Eq. (5):

$$\mu(\mathrm{RMSE})^2 + \sigma(\mathrm{RMSE})^2 \approx \mu(\mathrm{ME})^2 + \sigma(\mathrm{ME})^2 + \mu(\mathrm{SDE})^2 + \sigma(\mathrm{SDE})^2 \quad (6)$$

The stochastic images show a very low simulation bias (see μ(ME)). This tends slightly to increase toward negative values as the sampling resolution density decreases, except for clay content from 100 m to 200 m where the bias slightly decreases, albeit with larger dispersions (see σ(ME)). The simulation bias is at least one order of magnitude smaller of the average simulation error (μ(RMSE)). The statistics μ(ME), σ(ME), and σ(RMSE) are small compared to μ(RMSE). Therefore, following Eq. (5), μ(RMSE) can be considered as an indicator of the precision of the simulation. The soil property values are simulated in the unsampled locations with a precision generally one order of magnitude larger than the precision in laboratory measurements and four to five times larger than the precision in field measurements. As a reference, Table 6b summarizes the precision of laboratory measurement for clay, silt and bulk density as described in Minasny and McBratney (2002).

Table 6b Precision in laboratory and field measurements of soil physico-chemical properties (from Minasny and McBratney, 2002)

	CL	SI	BD
Laboratory	0.5	0.5	0.025
Field	2.7	4.0	0.200

SI, silt (%); CL, clay (%); BD, bulk density (g/cm^3).

Evaluation of PTF prediction uncertainty

Patterns of soil hydraulic properties estimated with indirect methods

The sets of basic soil properties are transformed into sets of soil hydraulic properties using the following indirect methods: (1) the HYPRES pedotransfer function, developed from European soil data (Wösten et al., 1999) and (2) the Vereecken's pedotransfer function (Vereecken et al., 1992). These two PTFs have been mainly chosen as illustrative test cases, among the most commonly employed PTFs in Europe. No consideration has been done concerning the performance of these PTFs for the study dataset, since the major aim of the present study is to speculate on the uncertainty being associated to the use of a certain published PTF, rather than comparing the performances of different PTFs.

The HYPRES pedotransfer relations predict the water retention function of soils indirectly by estimating parameters α and m of van Genuchten's equation (see Eq. (1)), while setting θ_r at 0.01 (except for coarse soils, where θ_r is set at 0.25) and $m = 1 - 1/n$. Vereecken's pedotransfer function also adopts van Genuchten's model for the retention function, but assumes $m = 1$.

The full dataset and the corresponding stochastic images of basic soil properties are employed to calculate spatial patterns of the corresponding van Genuchten soil water retention parameters, with Vereecken's and HYPRES PTF. These patterns are then employed to calculate patterns of soil water content, θ_h, at matric pressure head, h, respectively equal to -10 cm, -100 cm and $-10{,}000$ cm. The two pedotransfer functions have been also employed to estimate patterns of saturated hydraulic conductivity, K_s. HYPRES and Vereecken's PTF equations employed for estimating van Genuchten soil water retention parameters and the saturated hydraulic conductivity are listed in Table 7.

Table 6a Errors statistics of the simulated soil properties at the hillslope scale: mean (μ) and standard deviation (σ) of the mean error (ME), difference between generated value and exhaustive data set and the root mean square error (RMSE)

Sampling resolution	50 m			100 m			200 m		
	OM	CL	BD	OM	CL	BD	OM	CL	BD
μ(ME)	−0.006	0.152	−0.003	0.053	−2.119	−0.005	−0.086	−1.128	−0.008
σ(ME)	0.072	0.416	0.006	0.098	0.514	0.007	0.142	0.784	0.012
μ(RMSE)	1.573	8.258	0.120	1.603	10.478	0.126	1.765	10.700	0.134
σ(RMSE)	0.060	0.289	0.004	0.053	0.321	0.003	0.068	0.332	0.004

OM, organic matter (%); CL, clay (%); BD, bulk density (g/cm^3).

Table 7 HYPRES and Vereecken's PTF equations employed for estimating van Genuchten soil water retention parameters and the saturated hydraulic conductivity

HYPRES PTF
$\theta_0 = 0.7919 + 0.001691 * CL - 0.29619 * BD - 0.000001491 * SI^2 + 0.0000821 * OM^2 + 0.02427 * CL^{-1} + +0.01113 * SI^{-1} + 0.01472 * \ln(SI) - 0.0000733 * OM * CL - 0.000619 * BD * CL - 0.001183 * BD * OM + -0.0001664 * topsoil * SI$
$\ln(\alpha) = -14.96 + 0.03135 * CL + 0.0351 * SI + 0.646 * OM + 15.29 * BD - 0.192 * topsoil - 4.671 * BD^2 + -0.000781 * CL^2 - 0.00687 * OM^2 + 0.0449 * OM^{-1} + 0.0663 * \ln(SI) + 0.1482 * \ln(OM) - 0.04546 * BD * SI + -0.4852 * BD * OM + 0.00673 * topsoil * CL$
$\ln(n-1) = -25.23 - 0.02195 * CL + 0.0074 * SI - 0.1940 * OM + 45.5 * BD - 7.24 * BD^2 + 0.0003658 * CL^2 + +0.002885 * OM^2 - 12.81 * BD^{-1} - 0.1524 * SI^{-1} - 0.01958 * OM^{-1} - 0.2876 * \ln(SI) - 0.0709 * \ln(OM) - 44.6 * \ln(BD) + -0.02264 * BD * CL + 0.0896 * BD * OM + 0.00718 * topsoil * CL$
$\ln(K_s) = 7.755 + 0.0352 * SI + 0.93 * topsoil - 0.967 * BD^2 - 0.000484 * CL^2 - 0.000322 * SI^2 + 0.001 * SI^{-1} + -0.0748 * OM^{-1} - 0.643 * \ln(SI) - 0.01398 * BD * CL - 0.1673 * BD * OM + 0.02986 * topsoil * CL + -0.03305 * topsoil * SI$
Vereecken's PTF
$\theta_r = 0.015 + 0.005\ CL + 0.014\ OM$
$\theta_s = 0.81\ 0.283\ BD + 0.001\ CL$
$\ln(\alpha) = -2.486 + 0.025\ SA - 0.351\ OM - 2.617\ BD - 0.023\ CL$
$\ln(n) = 0.053 - 0.009\ SA - 0.013\ CL + 0.00015(SA)^2$
$\ln(K_s) = 20.62 - 0.96 * \ln(CL) - 0.66 * \ln(SA) - 0.46 * \ln(OM) - 8.43 * BD$

θ_0, α, n = parameters in van Genuchten equation (1); CL = percentage of clay; SI = percentage of silt; OM = percentage of organic matter; BD = bulk density (g/cm^3); topsoil is a qualitative variable having the value of 1 or 0. θ_r is equal to 0.025 for coarse soils and to 0.01 for other soils. K_s is the saturated hydraulic conductivity in (cm/d).

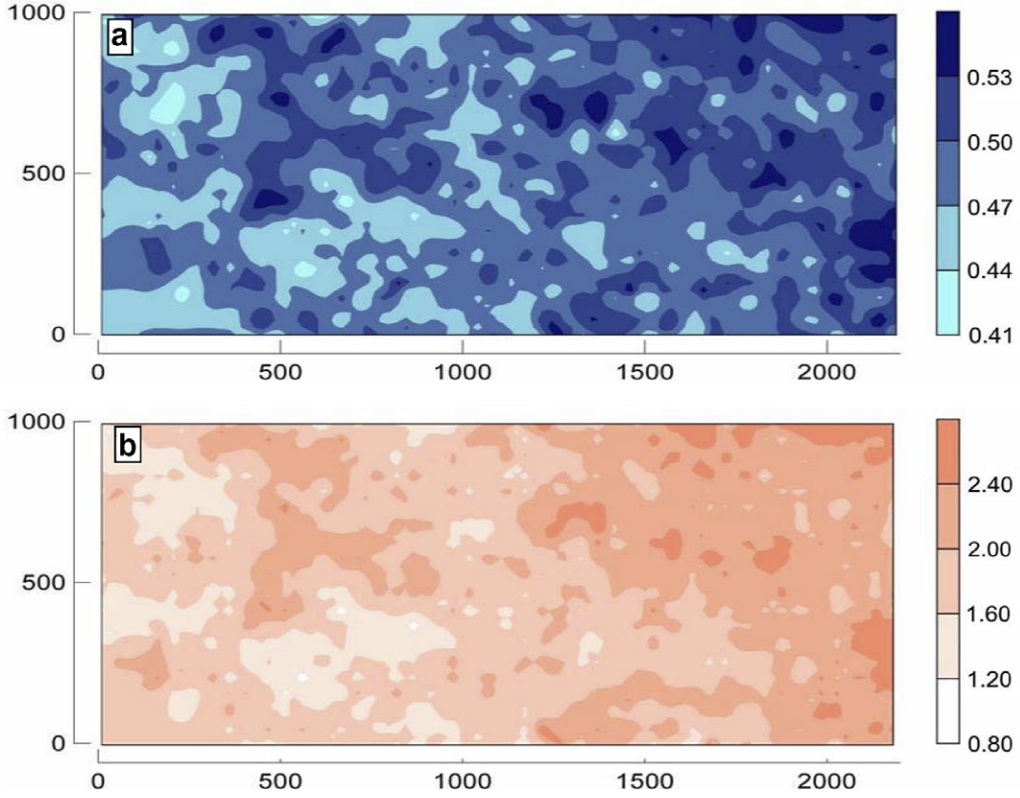


Figure 6 Example of simulated maps of: (a) soil water content (cm^3 cm^{-3}) at $h = -10$ cm, and (b) saturated hydraulic conductivity (cm day^{-1}), estimated with Vereecken et al.'s (1989) pedotransfer function.

Fig. 6a and b show an example of simulated pattern of soil water content at −10 cm matric pressure (i.e. θ_{-10}) and saturated soil hydraulic conductivity, K_s, respectively, both estimated with Vereecken's pedotransfer function.

PTF prediction uncertainty due to the uncertainty in input data

This section illustrates how the uncertainty in the estimation of the basic soil properties at unsampled locations across the hillslope (summarized in Table 6a) propagates into the PTF predictions. In order to evaluate this uncertainty, the patterns of soil hydraulic properties estimated from the images of the full dataset have been compared to those derived from the full dataset, which are taken as reference patterns. Discrepancy between these patterns has been evaluated by calculating the statistics ME and RMSE, according to Eqs. (2) and (3). The distributions of these two statistics are plotted in Figs. 7–9. Each distribution is made of 100 values, corresponding to the number of images of basic soil properties generated for each sampling

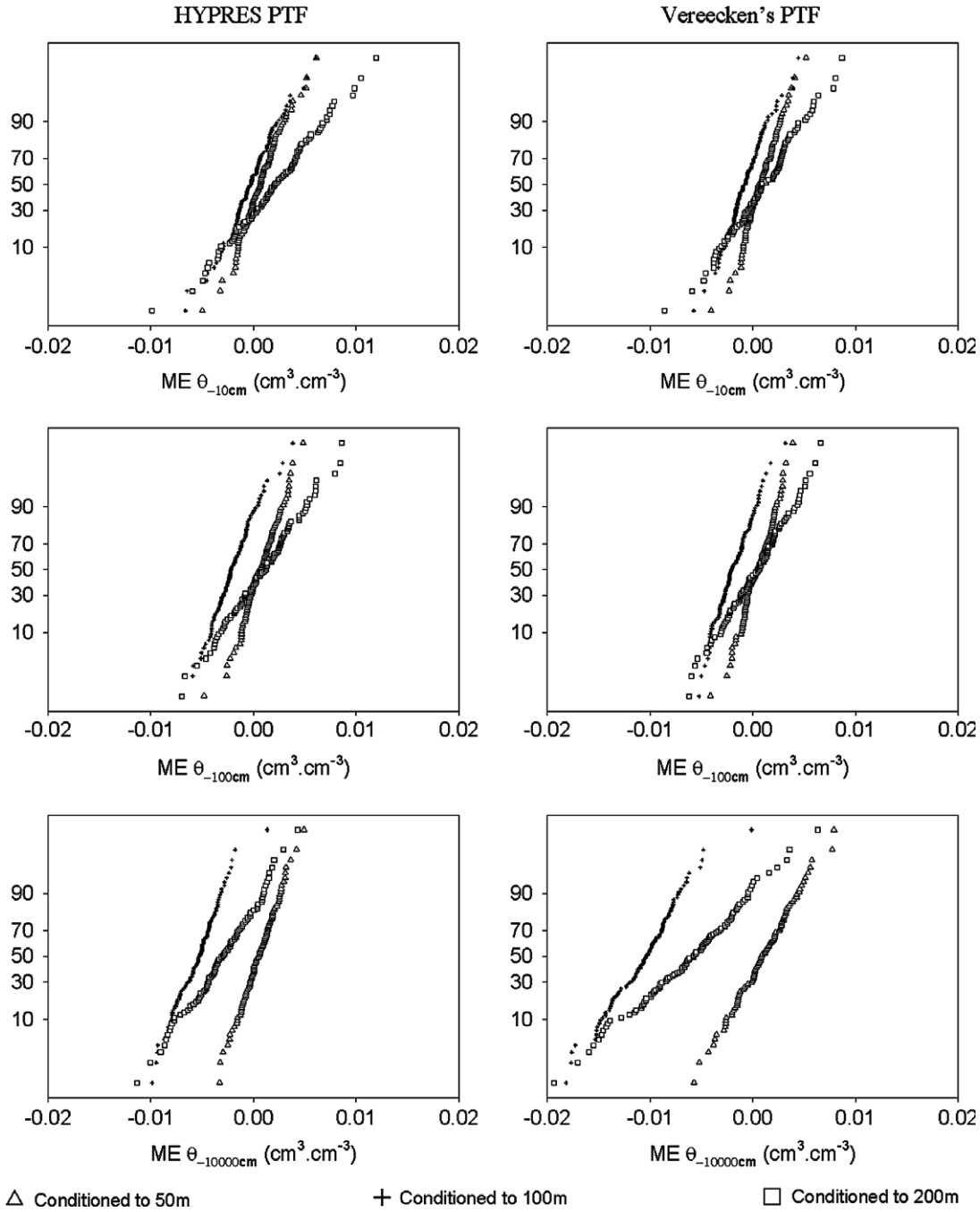


Figure 7 Normal probability plot of the mean error (ME, difference between values predicted from the simulated soil properties and the exhaustive data set) of the soil water retention θ_h at matric pressure head, h, respectively equal to −10 cm, −100 cm and −10,000 cm, predicted with HYPRES (left column) and Vereecken's PTF (right column) at the hillslope scale.

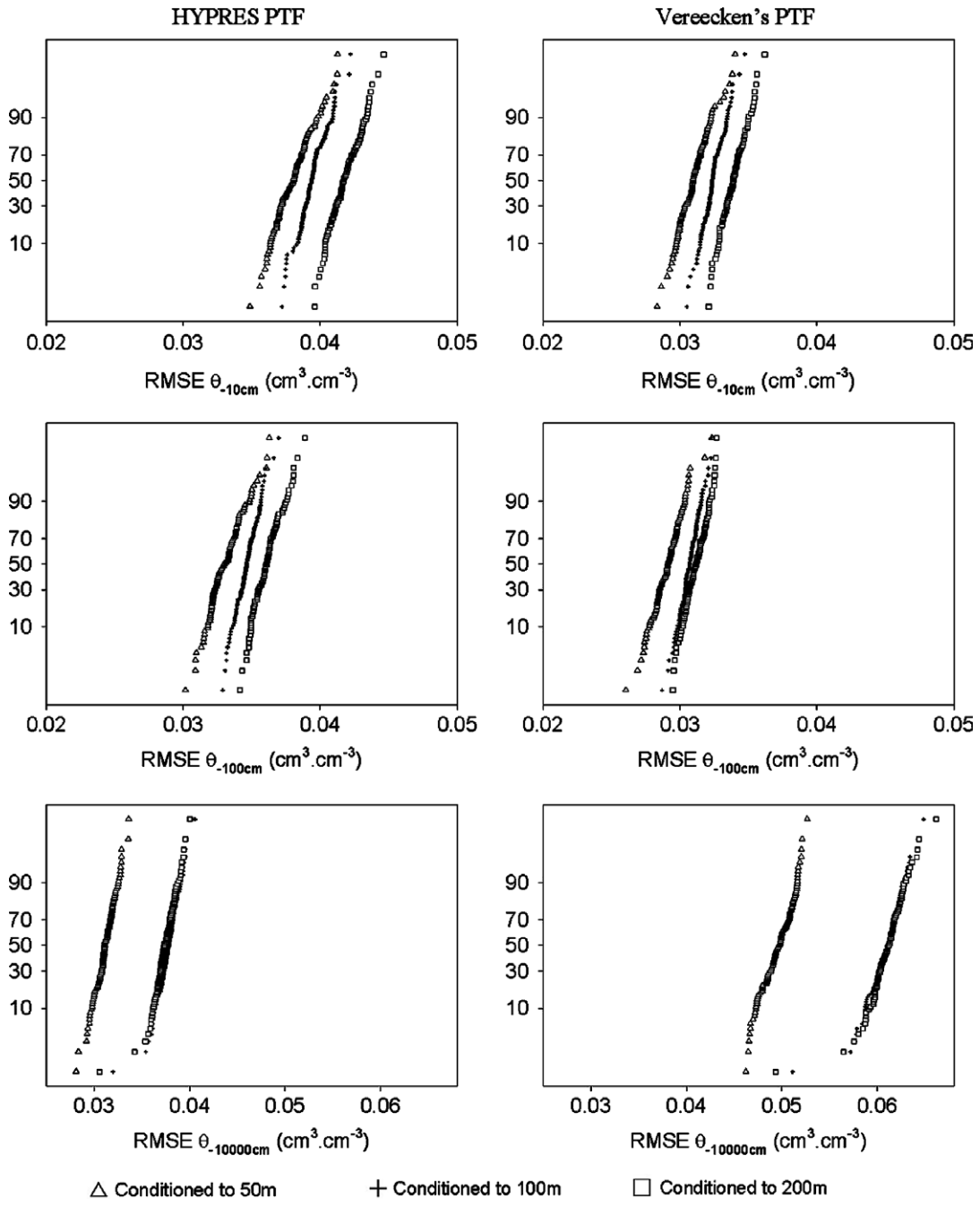


Figure 8 Normal probability plot of the root mean square error (RMSE) of the soil water retention of the soil water retention θ_h at matric pressure head, h, respectively equal to -10 cm, -100 cm and $-10{,}000$ cm, predicted with HYPRES (left column) and Vereecken's PTF (right column) at the hillslope scale.

density. Following the discussion in section ''Multivariate stochastic simulation of sets of basic soil properties'', ME is an indicator of the accuracy of the prediction, while RMSE is an indicator of the overall prediction error, including both prediction accuracy and precision. Both ME and RMSE are normally distributed for all cases, except for the saturated hydraulic conductivity estimated by Vereecken's PTF.

For what concerns the predicted soil water retention values, the means of the ME distributions are very close to zero (in absolute value not larger than 0.01 $cm^3\ cm^{-3}$), while the means of RMSE distributions are 0.03—0.04 $cm^3\ cm^{-3}$, except for $\theta_{-10,000}$ predicted by Vereecken's PTF, whose RMSE means are 0.05—0.06 $cm^3\ cm^{-3}$. The standard deviations of ME and RMSE are described by the slopes of their distributions. The lower this slope, the higher is the dispersion of the values around the mean, the higher is standard deviation. ME standard deviations are of the same order of magnitude of the corresponding means, while RMSE standard deviations are even smaller than the corresponding means. In summary, the results represented in Figs. 7 and 8 show that PTF prediction performances are fairly high, both in terms of accuracy and precision, for most practical applica-

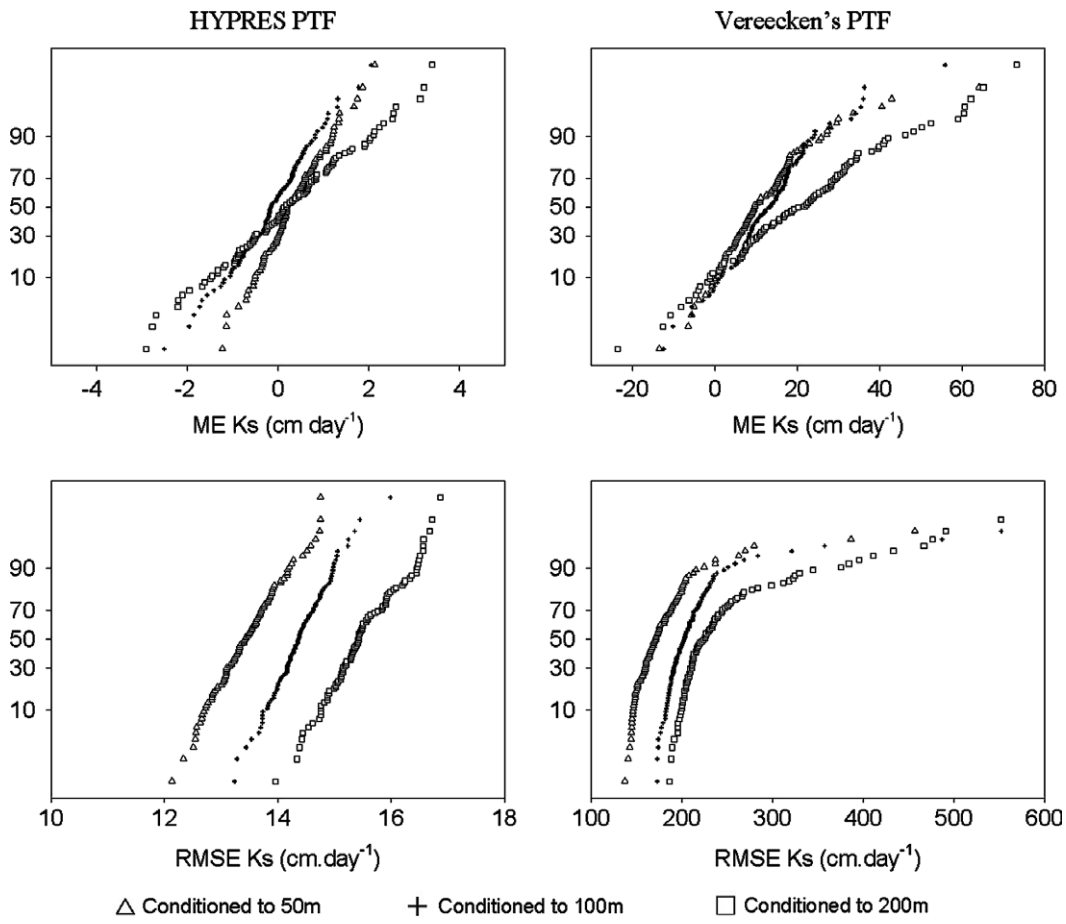


Figure 9 Normal probability plot of the error statistics of the saturated hydraulic conductivity (K_s) predicted with HYPRES (left column) and Vereecken's PTF (right column) at the hillslope scale. Top row describes the mean error (ME, difference between values predicted from the simulated soil properties and the exhaustive data set). The bottom row shows the root mean square error (RMSE).

tions. With a view to the major aim of this study, we argue that the PTF water retention prediction uncertainty, due to the input uncertainty only, is relatively small. The uncertainty of HYPRES PTF is slightly smaller than Vereecken's PTF at $\theta_{-10,000}$. HYPRES PTF sets $\theta_r = 0.01$ and thus it does not allow for significant variability in soil water contents at the lowest matric pressure heads.

Saturated hydraulic conductivity, K_s, as predicted by HYPRES and Vereecken's PTFs, is characterised by a very different sensitivity with respect to the uncertainty in input data, as shown in Fig. 9 by the relevant ME and RMSE distributions (note that these distributions are plotted by adopting different scales on the *x*-axes). HYPRES PTF predictions shows mean values of ME very close to zero and mean values of RMSE between 13 and 15 cm day^{-1}. Instead, Vereecken's PTF shows a mean bias of 10 cm day^{-1} and an average RMSE ranging from 150 cm day^{-1} to 200 cm day^{-1}. Vereecken's PTF RMSE distribution shows also a strong asymmetry toward higher values. In summary, the input uncertainty does not affect HYPRES PTF predictions accuracy K_s, rather its prediction precision. The input uncertainty is largely more significant for Vereecken's PTF, both in terms of precision and accuracy. This different behaviour is chiefly to be attributed to the structure of the Vereecken's PTF for estimating K_s (see Table 7), which makes it more sensitive to the uncertainty in the bulk density than HYPRES PTF.

PTF prediction uncertainty due to the uncertainty in input data and regression model

PTF prediction errors can occur even if input variables are ideally measured with no error. A source of uncertainty that can potentially affect hydraulic predictions from a PTF is the error in the regression model, which is made of the model structure limitations and actual errors in the regression coefficient values. One should always allow for an increase in the prediction uncertainty when using an indirect estimation method (a PTF in this study) as a simpler alternative to direct measurements. In a general framework, the uncertainty due to the model error can be smaller or larger than the uncertainty in the input variables. Knowledge of the relative importance of these two sources of uncertainty can turn out to be extremely useful for those who have to plan soil hydraulic characterisations over relatively large areas with limited resources available. It may not be worth to improve the sampling resolutions to reduce the input uncertainty beyond the model structure and parameter uncertainty.

The model error, i.e. the PTF prediction error in the case of ''correctly'' measured input, can be estimated by comparing the PTF predictions of the soil hydraulic properties performed with the measured basic soil properties (indirect estimations) to the direct measurements of the soil hydraulic properties along the study transect. This is possible

since, as illustrated in section ''Experimental hillslope and soil dataset'', both basic soil properties and soil hydraulic properties have been directly measured on the undisturbed soil cores collected along the study transect.

On the other hand, the PTF prediction uncertainty due to the input estimation uncertainty can be estimated in the same sampling locations of the study transect, by using as input data the basic soil properties simulated within the 100 images of the full dataset. The dispersion of the 100 PTF estimations at the each of the sample locations of the study transect is an indicator of the PTF prediction uncertainty due to the input estimation uncertainty.

Figs. 10 and 11 show direct measurements and indirect estimations along the transect. The same figures depict the 0.05 and 0.95 percentiles of the hydraulic properties estimated from the images of the full dataset for the three different sampling resolutions.

Table 8 shows the standard deviation of the direct measurements of the soil hydraulic properties, of the indirect estimation (i.e. PTF prediction from measured basic soil properties) and of the 0.5 percentile of PTF predictions from the 100 generated sets of basic soil properties at the transect points. According to a two-tailed *F*-test at a significance level $p = 0.05$, the standard deviations of the indirect method predictions are significantly smaller than those of the direct measurements, except for θ_{-10} estimated with HYPRES PTF and for $\theta_{-10,000}$ estimated with both HYPRES and Vereecken's PTF. The standard deviations of the 0.5

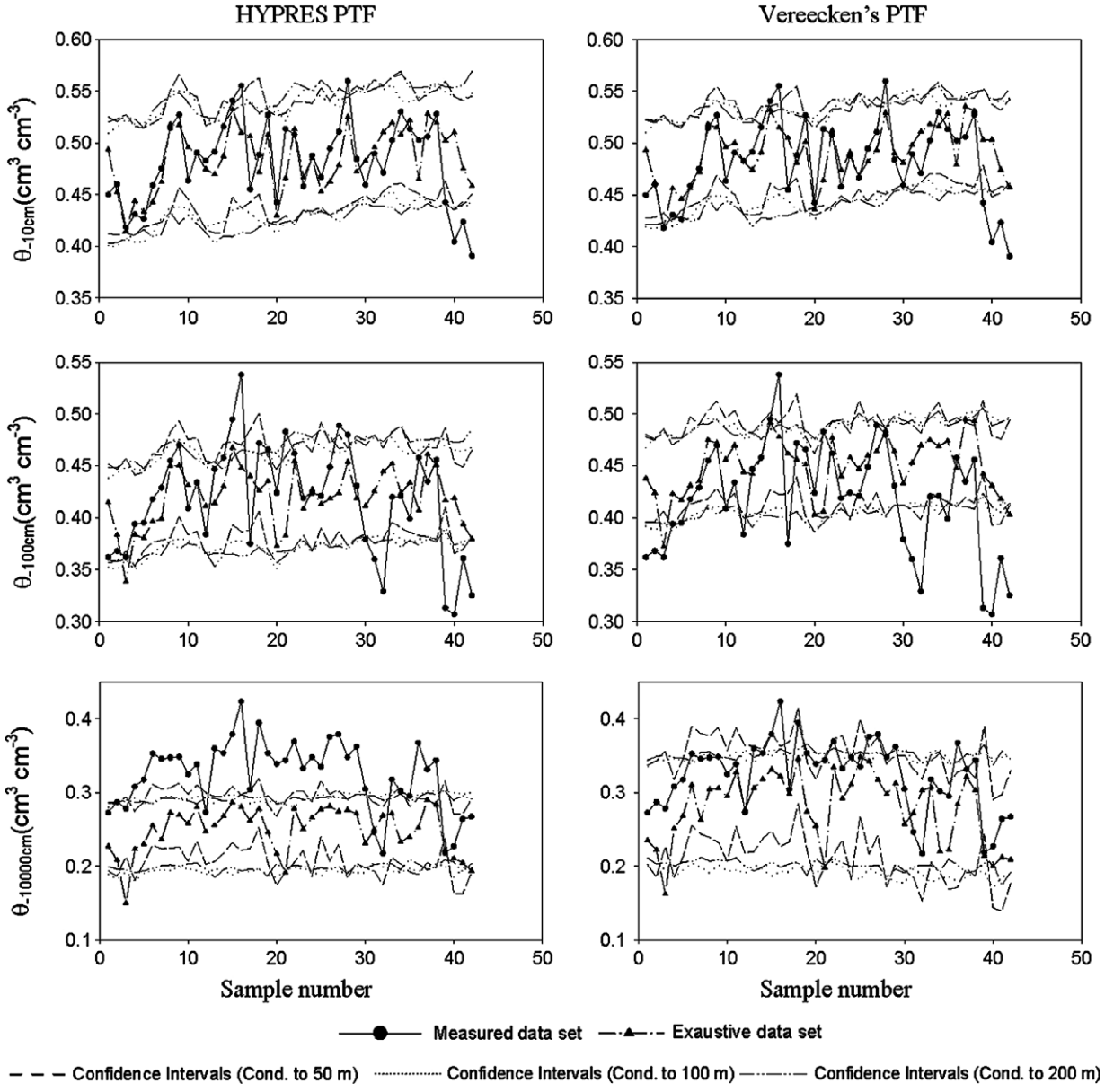


Figure 10 Evaluation of the PTF prediction uncertainty of soil water retention θ_h at matric pressure head, h, respectively equal to -10 cm, -100 cm and -10000 cm, accounting for both model and input error, by comparing observed and simulated soil hydraulic properties along the study transect.

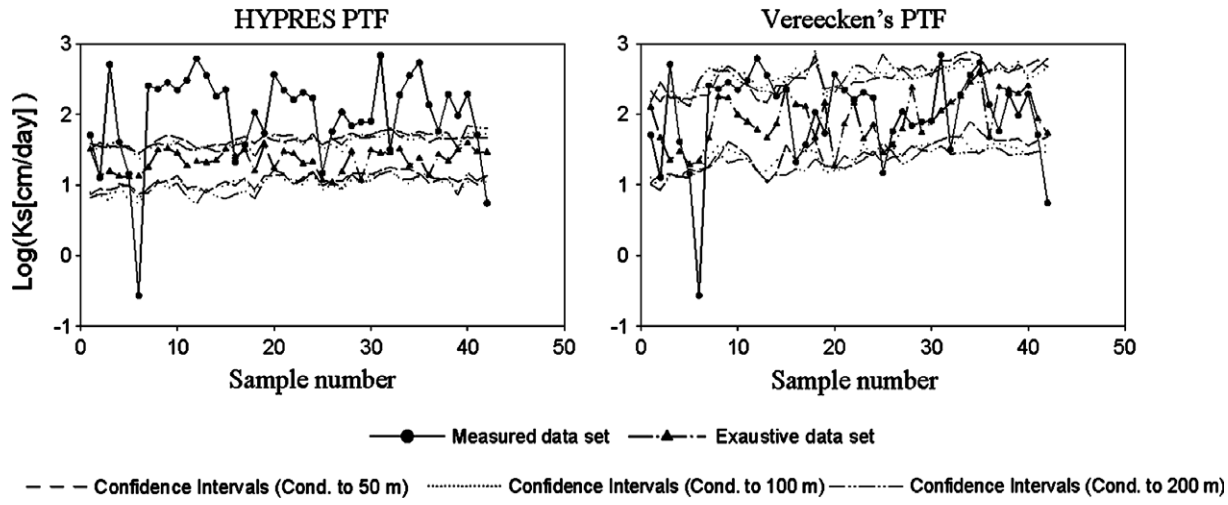


Figure 11 Evaluation of the PTF prediction uncertainty of the saturated soil hydraulic conductivity (K_s) accounting for both model and input error, by comparing observed and simulated soil hydraulic properties along the study transect.

percentile of the PTF predictions based of the generated values are even smaller. According to the *F*-test, only the standard deviation of $\theta_{-10,000}$ obtained from patterns conditioned to the 50 m grid is not significantly different from the observed one. No significant differences have been verified among the different sampling strategies.

PTF prediction error in the case of measured input is estimated by the mean error (ME) and the mean absolute error (MAE):

$$ME = \frac{1}{n}\sum_{k=1}^{n}(v_{ik} - v_{dk}) \quad (7)$$

$$MAE = \frac{1}{n}\sum_{k=1}^{n}|v_{ik} - v_{dk}| \quad (8)$$

where v_{ik} is the indirect estimation and v_{dk} the direct measurement of the variable of interest (either the soil water retention for a given matric pressure head or the saturated hydraulic conductivity) at the *k*th location of the *n* transect samples.

The MAE has been preferred to the RMSE, since RMSE, which includes the squared differences, can be significantly biased by outliers in small sample sets, such in this case. ME is an indicator of the PTF accuracy (mean bias), while MAE is an indicator of the overall prediction error (including both accuracy and precision). Both statistics are representative of the PTF prediction performance at the hillslope (transect) scale, in the case basic soil properties (the input data) are known.

Both PTFs provides fairly accurate estimation of the soil water retention function (see rows of ME statistics in Table 9). For the available dataset, HYPRES PTF seems more accurate than Vereecken's PTF at higher pressure heads (namely, h at -10 cm and -100 cm). At lower pressure heads ($h = -10{,}000$ cm), both PTFs underestimate the measured soil water content. In this case, HYPRES PTF

Table 8 Standard deviation of the soil hydraulic properties estimated from direct measurement and indirect estimation (i.e. PTF estimation from observed basic soil properties) compared to the standard deviation of the 0.5 percentile of the soil hydraulic properties predicted by the PTF from the generated patterns of soil hydraulic properties

		θ_{-10}	θ_{-100} ($cm^3\ cm^{-3}$)	$\theta_{-10,000}$	K_s ($cm\ day^{-1}$)
	Direct	0.0401	0.0526	0.0396	170.0
HYPRES	Indirect	**0.0297**	0.0277	**0.0312**	7.8
	$p_{0.5}$ 50 m	0.0161	0.0142	0.0171	4.1
	$p_{0.5}$ 100 m	0.0137	0.0097	0.0063	4.8
	$p_{0.5}$ 200 m	0.0131	0.0099	0.0056	*6.5*
Vereecken	Indirect	0.0273	0.0271	**0.0470**	87.6
	$p_{0.5}$ 50 m	0.0145	0.0134	**0.0332**	43.4
	$p_{0.5}$ 100 m	0.0132	0.0092	0.0098	33.6
	$p_{0.5}$ 200 m	0.0123	0.0076	0.0095	32.3

θ_h is the soil water content at matric pressure head, h, respectively equal to -10 cm, -100 cm and $-10{,}000$ cm. K_s is the saturated soil hydraulic conductivity. The values in bold indicate that the null hypothesis of equal variance between predicted and observed values cannot be rejected according to a two-tailed *F*-test at $p = 0.05$.

Table 9 Mean PTF prediction error compared to the mean PTF prediction standard deviations due to the input uncertainty at the transect locations

		θ_{-10}	θ_{-100}	$\theta_{-10,000}$	K_s
HYPRES	ME	0.003	0.001	−0.074	−156.9
	ME/$\sigma_{50\,m}$ (−)	0.112	0.062	−3.961	−19.5
	ME/$\sigma_{100\,m}$ (−)	0.112	0.062	−3.565	−20.2
	ME/$\sigma_{200\,m}$ (−)	0.103	0.058	−3.704	−19.0
Vereecken	ME	0.008	0.032	−0.044	−66.7
	ME/$\sigma_{50\,m}$ (−)	0.434	1.870	−1.455	−0.8
	ME/$\sigma_{100\,m}$ (−)	0.419	1.793	−1.330	−0.8
	ME/$\sigma_{200\,m}$ (−)	0.397	1.797	−1.364	−0.8

ME, mean error (difference between indirect estimations and direct measurements) of the soil water content θ_h ($cm^3\,cm^{-3}$) at different matric pressure head (h = −10, −100, and −10,000 cm) and the saturated hydraulic conductivity, K_s (cm day^{-1}); σ_x standard deviation of the PTF predictions due to input uncertainty from different sampling resolutions across the hillslope (x = 50, 100, and 200 m).

shows a larger bias than Vereecken's PTF. On the other hand, both HYPRES and Vereecken's PTF underestimate saturated hydraulic conductivity, K_s, although Vereecken's PTF bias is twice smaller than HYPRES PTF bias. It is interesting to note that nearly all the observed soil water retention and saturated hydraulic conductivity values fall within the 90% confidence interval of the values predicted by Vereecken's PTF.

HYPRES and Vereecken's PTFs provide similar MAE values for predicted water retentions and saturated hydraulic conductivities, as shown in Table 10 (see rows of MAE statistics). MAE values for water retention are still comparable to common errors attached to direct laboratory or field measurements. To this point, it is interesting to note that in the literature reported errors for measured soil water retention data range from 0.01 to 0.1 $cm^3\,cm^{-3}$ (Minasny et al., 2004; Vrugt et al., 2003).

By comparing ME values to MAE values, one can note that HYPRES PTF for soil water retention function is mainly affected by precision problems at the matric heads of −10 and −100 cm, whereas by accuracy problems at the matric head of −10,000 cm. HYPRES PTF for saturated hydraulic conductivity is also mainly affected by accuracy problems.

In the case of Vereecken's PTF, ME and MAE values of the soil water retention are similar, i.e. Vereecken's PTF model is essentially affected by accuracy problems. The MAE value of soil hydraulic conductivity is twice the corresponding ME absolute value, thus both accuracy and precision affect the overall Vereecken's PTF model prediction error of the saturated hydraulic conductivity.

The ratios of ME and MAE values to the corresponding standard deviations of the PTF estimations in the case of unknown input are computed for evaluating the relative importance of the PTF model prediction error versus the effect of the input uncertainty in the overall PTF prediction uncertainty. These statistics are also listed in Tables 9 and 10.

HYPRES PTF model bias of the predicted soil water retention at −10 cm and −100 cm pressure head is negligible compared to the input uncertainty, given that ME is always less than 50% the standard deviation of the predictions due the input uncertainty. At −10,000 cm pressure head, HYPRES PTF model bias is the most important source of uncertainty. For Vereecken's PTF, instead, the model bias is of the same order of magnitude of the input uncertainty.

Table 10 Mean PTF prediction absolute error compared to the mean PTF prediction standard deviations due to the input uncertainty at the transect locations

		θ_{-10}	θ_{-100}	$\theta_{-10,000}$	K_s
HYPRES	MAE	0.024	0.037	0.077	158.9
	MAE/$\sigma_{50\,m}$ (−)	1.071	1.929	4.109	19.8
	MAE/$\sigma_{100\,m}$ (−)	1.074	1.916	3.698	20.5
	MAE/$\sigma_{200\,m}$ (−)	0.983	1.809	3.842	19.3
Vereecken	MAE	0.022	0.043	0.050	124.2
	MAE/$\sigma_{50\,m}$ (−)	1.207	2.523	1.657	1.5
	MAE/$\sigma_{100\,m}$ (−)	1.164	2.420	1.515	1.5
	MAE/$\sigma_{200\,m}$ (−)	1.104	2.426	1.554	1.4

MAE, mean absolute error (difference between indirect estimations and direct measurements) of the soil moisture θ_h ($cm^3\,cm^{-3}$) at different matric pressure heads (h = −10, −100 and −10,000 cm) and the saturated hydraulic conductivity (cm day^{-1}); σ_x standard deviation of the PTF predictions due to input uncertainty from different sampling resolutions across the hillslope (x = 50, 100 and 200 m).

For both PTFs, the input uncertainty is less important than the corresponding model bias in the prediction of the saturated hydraulic conductivity. This is particularly true for HYPRES PTF, which ME is 20 times the standard deviation of the predictions due the input uncertainty.

The ratios of the MAE values to the corresponding standard deviations confirm that for both HYPRES and Vereecken's PTFs the overall model prediction error, including accuracy and precision, is as much as or even more important than the input uncertainty.

Conclusions

Although a Monte Carlo approach is computationally intensive, the stochastic simulation employed in the present study has proved to be an efficient tool for assessing quantitatively error propagation through non-linear models, such as pedotransfer functions. This issue has cross-implication when tackling problems of predicting soil hydraulic values at unsampled locations and upscaling local information to larger areas.

With specific reference to practical directions when planning field campaigns, outcomes of this study suggest that published PTFs provide accurate and precise estimates of the soil water retention characteristic at the hillslope scale even for a relatively coarse sampling resolution of basic soil hydraulic properties. Prediction errors due to input uncertainty are comparable to common errors attached to direct laboratory or field measurements.

This study also sheds light on the issue of uncertainty contributions. An interesting feature is that PTF prediction errors appear on average affected by the model structural error as much as or even more than by the uncertainties associated with the input data. This means that it may not worth measuring basic soil properties at high spatial resolution whether the target is to improve the PTF predictions. An increase in the spatial characterisation for the basic soil properties should be coupled with an improvement of the PTF model structure or parameterisation for the study area.

The PTFs considered here show worse level of performance when predicting saturated hydraulic conductivity. In this case the model error is much more significant than the input uncertainty, even at very high sample resolution.

This study will be extended by evaluating the propagation of the input uncertainty through the PTFs in a functional context (e.g. Vereecken et al., 1992), i.e. by examining the uncertainty in the outcome of a simulation model for a specific application, extending the strategy adopted by previous authors for evaluating PTF predictions at the point scale.

Acknowledgements

This study has been partly supported by P.O.N. project ''AQUATEC — New technologies of control, treatment, and maintenance for the solution of water emergency''. Hanoi Medina contributed to this paper during a visiting period at the University of Napoli Federico II, with a fellowship from the Italian—Latin American Institute. The final version of this paper benefited from the constructive comments of two anonymous reviewers and the thoughtful appraisal by the Editor Georges Vachaud.

References

Brubacker, S.C., Jones, A.J., Frank, K., Lewis, D.T., 1994. Regression models for estimating soil properties by landscape position. Soil Sci. Soc. Am. J. 58, 1763—1767.

Burgess, T.M., Webster, R., 1980. Optimal interpolation and isarithmic mapping of soil properties: I. The semi-variogram and punctual kriging. J. Soil Sci. 31, 315—331.

Castrignanó, A., Buttafuoco, G., 2004. Geostatistical stochastic simulation of soil water content in a forested area of South Italy. Biosyst. Eng. 87 (2), 257—266.

Christiaens, K., Feyen, J., 2001. Analysis of uncertainties associated with different methods to determine soil hydraulic properties and their propagation in the distributed hydrological MIKE SHE model. J. Hydrol. 246, 63—81.

Florinsky, I.V., Arlashina, H.A., 1998. Quantitative topographic analysis of gilgai soil morphology. Geoderma 82, 359—380.

Gómez-Hernández, J.J., Journel, A.G., 1993. Joint sequential simulation of multiGaussian fields. In: Soares, A. (Ed.), Geostatistics Tróia. Kluwer, Dordrecht, pp. 85—94.

Goovaerts, P., 2000. Estimation or simulation of soil properties? An optimization problem with conflicting criteria. Geoderma 97, 165—186.

Jenny, H., 1980. The Soil Resource, Origin and Behaviour. Springer-Verlag, New York.

Minasny, B. et al, 2004. Neural networks prediction of soil hydraulic functions for alluvial soils using multistep outflow data. Soil Sci. Soc. Am. J. 68, 417—429.

Minasny, B., McBratney, A.B., 2002. Uncertainty analysis for pedotransfer functions. Eur. J. Soil Sci. 53, 417—429.

Minasny, B., McBratney, A.B., Bristow, K.L., 1999. Comparison of different approaches to the development of pedotransfer functions for water retention curves. Geoderma 93, 225—253.

Moore, I.D., Gessler, P.E., Nielsen, G.A., Peterson, G.A., 1993. Soil attribute prediction using terrain analysis. Soil Sci. Soc. Am. J. 57, 443—452.

Oliver, D.S., 2003. Gaussian co-simulation: modeling of the cross-covariance. Math. Geol. 35 (4), 181—197.

Pachepsky, Y.A.J.T.D., Timlin, D.J., Rawls, W.J., 2001. Soil water retention as related to topographic variables. Soil Sci. Soc. Am. J. 65, 1787—1795.

Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. Comput. Geosci. 30, 683—691.

Pebesma, E.J., Wesseling, C.G., 1998. Gstat, a program for geostatistical modelling, prediction and simulation. Comput. Geosci. 24 (1), 17—31.

Rawls, W.J., Brakensiek, D.L., 1989. Estimation of soil water retention and hydraulic properties. In: Morel-Seytoux, H.J. (Ed.), Unsaturated Flow in Hydrologic Modeling — Theory and Practices, NATO ASI Series. Kluwer, Dordrecht.

Reynolds, W.D., Elrick, D.E., 2002. Falling head soil core (tank) method. In: Dane, J.H., Topp, G.C. (Eds.), Methods of Soil Analysis — Part 4: Physical Methods. Soil Science Society of America, Madison, WI, USA.

Romano, N., 1993. Use of an inverse method and geostatistics to estimate soil hydraulic conductivity for spatial variability analysis. Geoderma 60, 169—186.

Romano, N., Chirico, G.B., 2004. The role of terrain analysis in using and developing pedotransfer functions. In: Pachepsky, Y.A., Rawls, W.J. (Eds.), Development of Pedotransfer Functions in

Soil Hydrology, Developments in Soil Science. Elsevier, Amsterdam, pp. 273–294, Chapter 16.

Romano, N., Palladino, M., 2002. Prediction of soil water retention using soil physical data and terrain attributes. J. Hydrol. 265, 56–75.

Schaap, M.G., Leij, F.J., 1998. Database-related accuracy and uncertainty of pedotransfer functions. Soil Sci. 163, 765–779.

Sivapalan, M., 2003. IAHS decade on predictions in ungauged basins (PUB). Hydrol. Sci. J. 48 (6), 857–880.

Tarboton, D.G., 1997. A new method for the determination of flow directions and upslope areas in grid digital elevation models. Water Resour. Res. 33 (2), 309–319.

Vereecken, H., Maes, J., Darius, P., 1989. Estimating the soil moisture retention characteristics from texture, bulk density and carbon content. Soil Sci. 148, 389–403.

Vereecken, H., Diels, J., van Orshoven, J., Feyen, J., Bouma, J., 1992. Functional evaluation of pedotransfer functions for the estimation of soil hydraulic properties. Soil Sci. Soc. Am. J. 56, 1371–1378.

Vrugt, J.A., Bouten, W., Gupta, H.V., Hopmans, J.W., 2003. Toward improved identifiability of soil hydraulic parameters: on the selection of a suitable parametric model. Vadose Zone J. 2, 98–113.

Wagener, T. et al, 2004. Predictions in ungauged basins as a catalyst for multidisciplinary hydrology. Eos 85 (44), 451–457.

Western, A.W., Blöschl, G., 1999. On the spatial scaling of soil moisture. J. Hydrol. 217 (3-4), 203–224.

Wösten, J.H.M., Lilly, A., Nemes, A., Le Bas, C., 1999. Development and use of a database of hydraulic properties of European soils. Geoderma 90, 169–185.

Zhao, Y.C., Shi, X.Z., Yu, D.S., Wang, H.J., Sun, W.X., 2005. Uncertainty assessment of spatial patterns of soil organic carbon density using sequential simulation, a case study Hebei, province, China. Chemosphere 59 (11), 1527–1535.