Journal of Applied Geophysics 138 (2017) 210–219

Contents lists available at ScienceDirect

Journal of Applied Geophysics

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

Spatial filtering of electrical resistivity and slope intensity: Enhancement of spatial estimates of a soil property

Hocine Bourennane [a,*], Florent Hinschberger [b], Caroline Chartin [b,c], Sébastien Salvador-Blanes [b]

[a] URSOLS, INRA, 45075 Orléans, France
[b] Université François-Rabelais de Tours, EA 6293 GéHCO, Faculté des Sciences et Techniques, Laboratoire Géosciences et Environnement, Parc Grandmont, F-37200 Tours, France
[c] George Lemaître Centre for Earth and Climate Research, Earth & Life Institute, Université catholique de Louvain, Place Louis Pasteur 3, 1348 Louvain-la-Neuve, Belgium

A R T I C L E I N F O

Article history:
Received 27 May 2016
Received in revised form 27 January 2017
Accepted 28 January 2017
Available online 1 February 2017

Keywords:
Kriging spatial components
Multiple external drifts
Principal component analysis
Soil thickness
High resolution electrical resistivity
Slope intensity

A B S T R A C T

To best utilize the electrical resistivity data and slope intensity derived from a Digital Elevation Model, the kriging spatial components technique was applied to separate the nuggets and small- and large-scale structures for both resistivity and slope intensity data. The spatial structures in the resistivity and slope intensity data, which are poorly correlated with soil thickness (ST), are then filtered out prior to integrating the resistivity data and slope intensity into soil thickness estimation over a 12 ha area located in the south-western Parisian Basin (France). ST was measured at 650 locations over the study area by manual augering. Twenty percent of the observations (131 points) were randomly selected to constitute the validation dataset. The remaining 80% of the dataset (519 points) was used as the prediction dataset.

The resistivity data represent a set of 7394 measurement points for each of the three investigated depths over the study area. The methodology involves successively (1) a principal component analysis (PCA) on the electrical measurements and (2) a geostatistical filtering of the small-scale component and noise in the first component (PC1) of the PCA. The results show that the correlation between ST and PC1 is greatly improved when the small-scale component and noise are filtered out, and similarly, the correlation between ST and slope intensity is greatly improved once the geostatistical filtering is carried out on the slope data. Thus, the large scales of both slope intensity and the electrical resistivity's PC1 were used as external drifts to predict ST over the entire study area. This prediction was compared with ordinary kriging and kriging either with a large scale of slope intensity or with a large scale of the electrical resistivity's PC1 taken as an external drift. The first prediction of ST by ordinary kriging, which was considered as our reference, was also compared to those achieved by kriging using the raw secondary variables: PC1 and slope intensity as external drifts; slope intensity as an external drift; and PC1 as an external drift. The results indicate a reasonably low bias of prediction for all of the methods, in particular in the case of kriging using the large scales of both slope intensity and PC1 as external drifts. The root mean square error shows that kriging accounting for the large scales of two secondary exhaustive variables is the most accurate prediction method. The relative improvement of the accuracy is at least equal to 29% between the approach accounting for both large scale components of secondary attributes in the spatial estimates of ST and the other approaches of estimates considered in this study.

1. Introduction

Soil thickness is one of the most important input parameters for hydroecological models (e.g., Tesfa et al., 2009; Wahren et al., 2016). Soil thickness also provides an indication of the available water capacity and exerts major control on many physical and biological processes occurring in soils (e.g., Gessler et al., 1995; Diek et al., 2014; Akumu et al., 2016). Consequently, the accurate representation of soil thickness at scales relevant to these processes is increasingly important for use in distributed simulation models of hydrology and ecology. Soil thickness is highly variable spatially and is laborious, time consuming and difficult to practically measure, even for a modestly sized watershed (e.g., Dietrich et al., 1995; Afshar et al., 2016). Thus, there is a need for models that can predict the spatial pattern of soil thickness. Abundant and accessible ancillary information, such as electrical resistivity and attributes derived from a Digital Elevation Model, is widely used to improve the estimates of a sparse target variable, such as soil thickness (e.g., De Benedetto et al., 2012; Besalatpour et al., 2013; Mehnatkesh et al., 2013). Such approaches assume that relevant relationships exist between the target variable and ancillary variables. Unfortunately,

* Corresponding author at: Institut National de la Recherche Agronomique (INRA), France.
E-mail address: Hocine.Bourennane@inra.fr (H. Bourennane).

http://dx.doi.org/10.1016/j.jappgeo.2017.01.032
0926-9851/

good correlation between electrical resistivity measurements and soil properties, as observed in the laboratory or documented in the literature, may not be reproduced in field data because of differences in the measurement supports of soil properties and resistivity data (e.g., Chilès and Guillan, 1984; Ma and Royer, 1988; Bourgault, 1994; Coscia et al., 2012; Minsley et al., 2012; Davydenko and Grayver, 2014). Such a finding is often observed between the soil properties and terrain attributes derived from a Digital Elevation Model (e.g., Zhu and Lin, 2011; Kim and Zheng, 2011).

It is also well-established that any soil property measured at specific locations in space is a result of several physical, chemical and biological processes. Some processes can operate only at small distances (e.g., biological activities), whereas others may act over larger distances (e.g., weathering of parent material). In geostatistical terminology, the combined effects of different sources of spatial variation produce variograms of the considered properties with nested structures, provided that they act at distinct spatial scales.

Thanks to methods such as factorial kriging (Matheron, 1982; Goovaerts, 1997; Wackernagel, 1998), the spatial components of a nested variogram can be estimated and mapped separately. This geostatistical technique was first used in geochemical exploration to distinguish anomalies (Sandjivy, 1984). The same technique was applied to image restoration, filtering and lineament enhancement by Ma and Royer (1988), whereas Oliver et al. (2000) used factorial kriging to separate short-range spatial components from long-range components in SPOT images. Van Meirvenne and Goovaerts (2002) applied factorial kriging (FK) to the filtering of synthetic aperture radar (SAR) images, strengthening relationships with land characteristics such as topography and land use. Goovaerts et al. (2005) applied this technique to detecting anomalies and patches on high spatial resolution hyperspectral imagery. In addition, numerous applications of factorial kriging (Goulard and Voltz, 1992; Goovaerts and Webster, 1994; Webster et al., 1994; Dobermann et al., 1995, 1997; Bocchi et al., 2000; Castrignano et al., 2000; Lin, 2002; Bourennane et al., 2003, 2004, 2012; de Fouquet et al., 2011; Milne et al., 2012; Allaire et al., 2012) have shown that approaches in which all sources of variation are mixed (e.g., correlation analysis, common principal component analysis and multi-linear regression) blur the real relationships among variables, as they average out distinct changes in the correlation structures occurring at different spatial scales and they included the micro-scale variations. The filtering of the different components often discloses interesting correlations between variables changing as a function of spatial scale. Such filtering leads to an enhancement of relationships between the studied variables. Small prediction errors of the target variable are expected as the correlation between a target variable and explanatory variable increases.

In this paper, the explanatory variables are electrical resistivity measurements and slope intensity. The choice of electrical resistivity measurements and slope intensity derived from a Digital Elevation Model for the enhancement of the spatial estimation of soil thickness was guided by the fact that several studies (e.g., Bourennane et al., 1996, 1998; Herbst et al., 2006; Besson et al., 2010) have shown that soil thickness is related to both electrical resistivity and slope intensity. A shallow soil is generally less conductive than a thicker soil, and high slope intensity is commonly associated with a shallow soil.

This paper investigates how spatial prediction can be improved by capitalizing on the better correlation between variables at specific spatial scales. The final objective of this paper is to confirm the generic aspect of the approach deployed here. In fact, in a previous paper (Bourennane et al., 2012), we used a similar approach to improve the spatial estimates of soil water content based on electrical resistivity data. The major differences between the two papers concern the pedological context, as well as the target variable, and the number of auxiliary variables (two instead of one) used for spatial estimates of the target variable.

The objective of this paper was examined through the spatial estimation of soil thickness (ST), punctually measured over 12 ha in the center of France, using as external drifts in kriging equations both electrical resistivity measurements (ARP) and slope intensity. Both external drifts were exhaustively measured over the study area. The basic idea is to replace the values of ARP and slope intensity in the kriging equations by the values of spatial components that are the most strongly correlated with the soil thickness. The performance of soil thickness mapping was analyzed using a validation dataset.

2. Materials and methods

2.1. Location of the study area and physiographic settings

The field study was carried out on a 12 ha southeast-facing hillslope located near the village of Seuilly (south-western Parisian Basin, 47°08.31′N, 0°10.97′E). The elevation of the study area ranges from 41 m to 80 m, and the slope length is approximately 600 m. The study area is composed of the following sedimentary bedrocks from downslope to upslope: Lower Turonian white chalks, Middle Turonian white chalks and Upper Turonian yellow sandy limestones (Alcaydé et al., 1989; Bellemlih, 1999). The main soils observed in the study area are calcaric Cambisols, epileptic calcaric Cambisols and colluvic Cambisols (Boutin et al., 1990; FAO, 1998).

2.2. Data acquisition

2.2.1. Topography

Two DGPSs (Trimble ® ProXRS) were used as a base and a mobile recorder. The coordinates and elevations of 1550 points were obtained by post-treatment of the data. A Digital Elevation Model (DEM) was estimated over a two-meter grid. Finally, topographic attributes, such as the slope intensity (Fig. 1a), were derived from the DEM through the algorithms implemented in the GIS software ArcGis 9.3.1®

2.2.2. Soil thickness

Soil thickness was measured by manual augering at 650 different locations over the study area. The soil thickness was defined as the summation of the A and B horizons, i.e., the depth of the upper saprolite limit. Differentiation between the B and C horizons was relatively easy because C horizons are often white and the transition is sharp. Two sampling schemes were established to consider the short-distance variability of soil thickness, especially when associated with linear anthropogenic landforms (Bollinne, 1971; Macaire et al., 2002; Salvador-Blanes et al., 2006; Chartin et al., 2011). The two sampling schemes were defined as follows. A total of 502 soil augerings were concentrated on the more relevant linear landforms (lynchets and undulations) observed in the study area (Chartin et al., 2011; Bourennane et al., 2014). An additional 148 soil augerings were performed to measure the soil thickness variation over all the study area. For that purpose, a point was sampled randomly in each square of a 25 m × 25 m grid over the entire study area (Fig. 1b). Twenty percent of the observations (131 points) were randomly selected to constitute a validation dataset. The remaining 80% of the dataset (519 points) was used as a prediction dataset.

2.2.3. Electrical resistivity measurements

An Automatic Resistivity Profiling (ARP) survey was conducted within the survey site. The device used involves three arrays. Each array is composed of four wheels that are metallic probes; two are used to pass current into the soil, and the other two are used to record the electrical potentials of the soil. The spacing between the current probes and the potential probes is 0.5 m for the first array, 1.0 m for the second array, and 2.0 m for the third array. Thus, for each measurement point, three apparent resistivity values (namely, ARP1, ARP2 and ARP3) were computed, corresponding to three proxy depths of

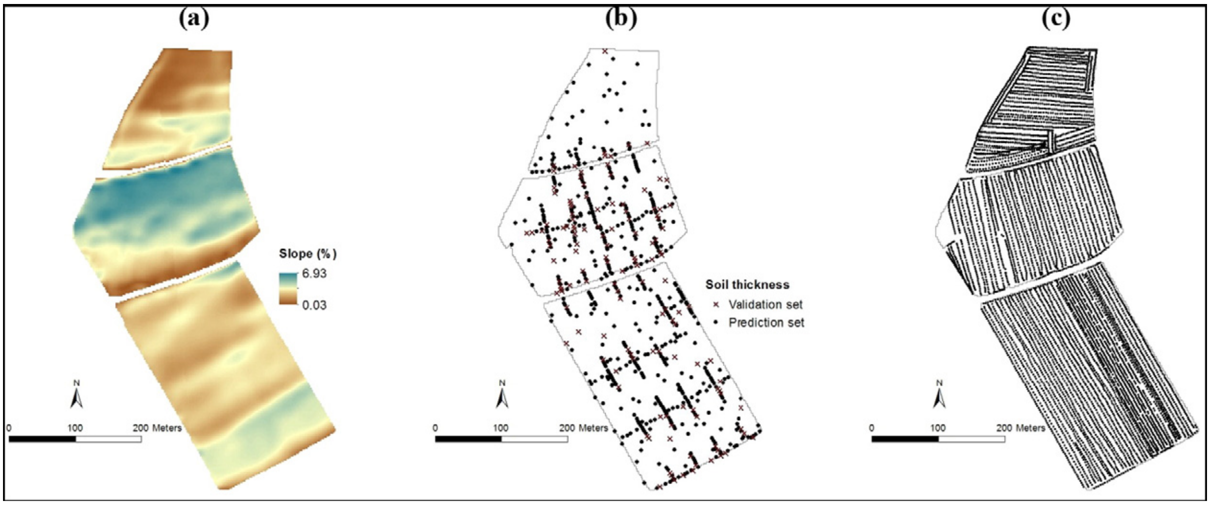

Fig. 1. Global sampling pattern for the target variable and secondary attributes: (a) slope intensity derived from DEM; (b) soil thickness; (c) Automatic Resistivity Profiling (ARP) measurements.

investigation (approximately 0–0.5 m, 0–1 m and 0–1.7 m, respectively). Data were collected continuously along the profiles and georeferenced using a DGPS positioning system. The average measurement interval along the profiles was 0.2 m, whereas the spacing between the profiles was much larger (approximately 6 m). The measured resistivity values were filtered along each profile using a 1D median moving window filter to increase the signal/noise ratio. In total, approximately 7500 ARP resistivity values (Fig. 1c) within the prospected area were recorded.

2.3. Geostatistical methodology

2.3.1. Spatial filtering by factorial kriging

In geostatistics, the variogram enables the building of estimations and simulations by capturing the spatial correlation inherent to a dataset. Factorial kriging is a variogram-based filtering technique developed by Matheron (1982). It relies on a simple additive model, where the spatial variable under study is modeled by a random function, Z($\mathbf{x}$), which is separated into terms of independent factors:

$$Z(\mathbf{x}) = Z_1(\mathbf{x}) + Z_2(\mathbf{x}) + \ldots. \tag{1}$$

Noise attenuation issues can be easily handled in the framework of this model, as far as the noise part of a dataset can be considered independent of a complementary signal part:

$$Z(\mathbf{x}) = Z_{\mathrm{NOISE}}(\mathbf{x}) + Z_{\mathrm{SIGNAL}}(\mathbf{x}) \tag{2}$$

In such a way, factorial kriging, by estimating $Z_{\mathrm{SIGNAL}}(\mathbf{x})$, allows the filtering out of the noisy component of a dataset.

The spatial correlation structure of the dataset values can be described by a nested variogram model, which can be written as

$$\gamma(\mathbf{h}) = \sum_{l=0}^{L} \gamma_l(\mathbf{h}) = \sum_{l=0}^{L} b_l \gamma_l(\mathbf{h}) \qquad \text{with } b_l \geq 0 \tag{3}$$

The variogram is thus modeled as the sum of $L+1$ basic variograms (e.g., Goovaerts, 1997), each corresponding to a distinct spatial structure:where $b_l$is the variance of the corresponding basic variogram model $\gamma_l(\mathbf{h})$and $\mathbf{h}$ is a vector (distance) separating any pair of measures. The variance corresponding to $l=0$ is called a nugget and represents the spatially unstructured part of the total variance. Based on the linear model of regionalization (3), the random function $Z(\mathbf{x})$ can be decomposed into a sum of $(L+1)$ independent random functions, called spatial components, and its local mean m($\mathbf{x}$):

$$Z(\mathbf{x}) = \sum_{l=0}^{L} Z_l(\mathbf{x}) + m(\mathbf{x}) \tag{4}$$

where $Z_l(\mathbf{x})$ is the lth spatial component corresponding to the variogram model $\gamma_l(\mathbf{h})$.

Each spatial component is thus individually mapped by filtering out the other components. The estimator of the lth spatial component of variable Z at location $\mathbf{x}_0$ is

$$Z_i^*(\mathbf{x}_0) = \sum_{\alpha=1}^{n} w_{\alpha,l} Z(\mathbf{x}_\alpha) \tag{5}$$

with n the number of observations around $\mathbf{x}_0$ involved in the estimation and each observation receiving a weight $w_{\alpha,l}$.

Table 1
Summary statistics of the target variable (soil thickness) and the secondary attributes (slope intensity and ARP measurements).

Variable	Unit	Number	Mean	Variance	Min	Max	g_1^*	D^{**}	p-Value
Soil thickness: calibration set	m	519	0.58	0.08	0.22	1.80	1.46	0.154	<0.0001
Soil thickness: validation set	m	131	0.70	0.15	0.25	2.23	1.48	0.167	0.001
ARP1***	Ω·m	7394	33.51	129.64	5.37	138.41	1.26	0.098	<0.0001
ARP2	Ω·m	7394	24.69	96.09	2.62	129.51	1.11	0.063	<0.0001
ARP3	Ω·m	7394	33.97	259.81	3.44	119.61	0.77	0.065	<0.0001
Slope intensity	%	31,998	3.15	2.13	0.03	6.93	0.31	0.067	<0.0001

g_1^*: skewness; D^{**}: Kolmogorov-Smirnov statistic.
ARP1***: Automatic Resistivity Profiling of the first depth of investigation.

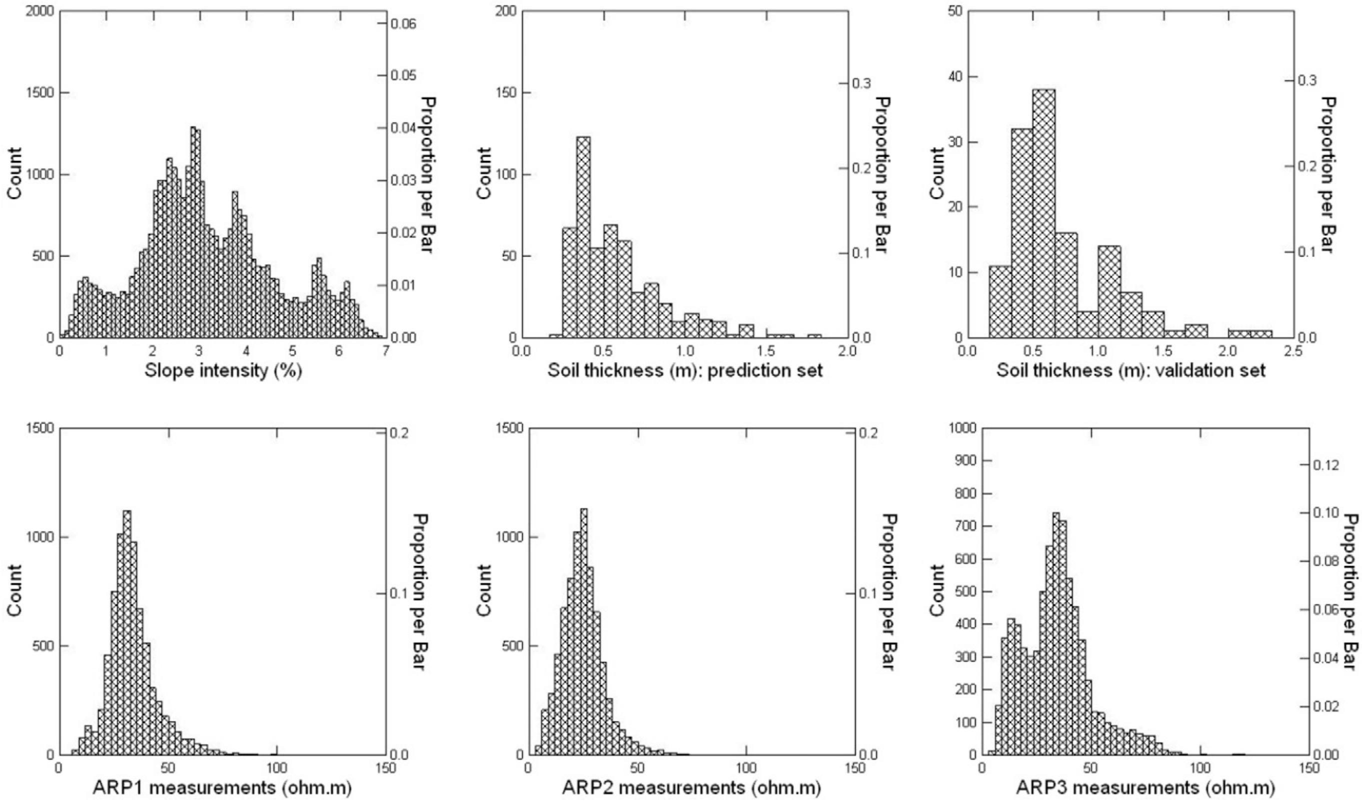

Fig. 2. Histograms of the target variable (soil thickness) and the secondary attributes (slope intensity and ARP measurements).

2.3.2. Accounting for secondary attributes

Let us consider the problem of estimating the value of a continuous soil variable z at an unsampled location $\mathbf{x}$ using data related to this variable, supplemented by values of secondary attributes that are exhaustively sampled. Several studies (e.g., Renard and Nai-Hsien, 1988; Chilès, 1991; Bourennane and King, 2003) have shown that kriging with multiple external drifts is more appropriate than cokriging and necessarily also more appropriate than ordinary kriging in such configuration sampling: the target variable is known in few locations, whereas the secondary attributes are exhaustively sampled over the study area. In this study, the decomposition described in the previous subsection was performed for the secondary attributes. The next step is to look at the correlation between the spatial components of secondary attributes and the target variable. The main idea consists of replacing the raw measurements of the secondary attributes in the kriging equations by the values of their spatial components that are the most strongly correlated with the target variable.

2.3.3. Principal component analysis

Principal component analysis (PCA) of a set of p images generally aims to summarize and hopefully improve the interpretation of the available information by using a few new images that are orthogonal

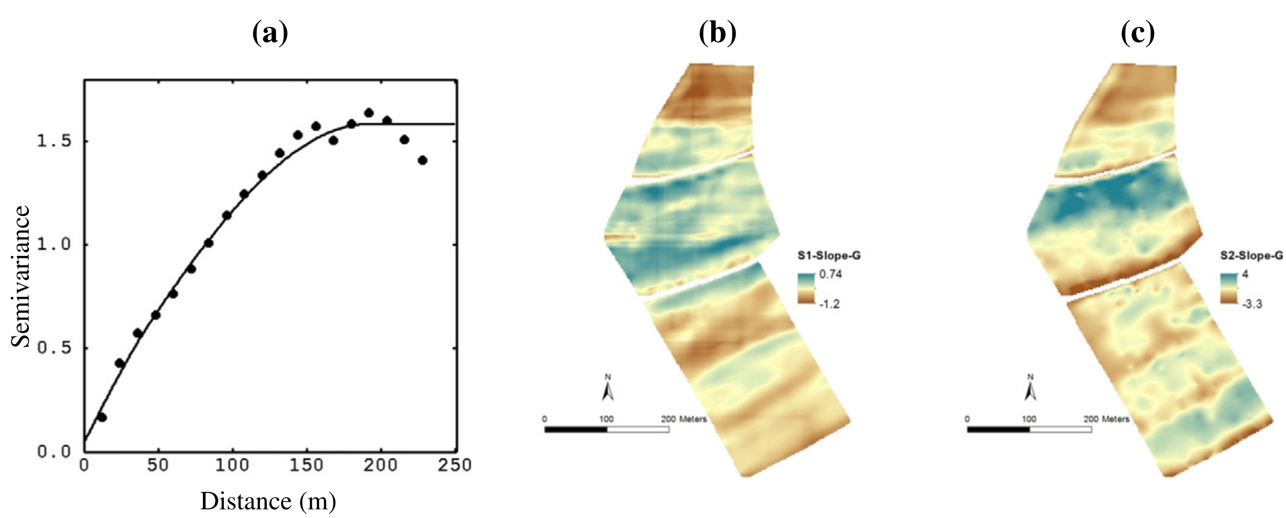

Fig. 3. Omnidirectional experimental (dots) and the theoretical (curves) variograms of the normal score transform values of slope intensity (a); maps of the decomposition of the original normal score transform values of slope intensity into two structures: (b) local: small-scale (S1-Slope-G); (c) regional: large-scale (S2-Slope-G).

Table 2
The parameters of the experimental model and the fitted models for the secondary attributes.

Variable	Lag (m)	Number of lags	Nugget	Model	Range (m)	Sill
Slope-G	12	20	0.05	Spherical	45	0.08
				Spherical	192	1.45
PC1-ARP	20	13	0.13	Spherical	34	0.35
				Spherical	99	2.07

linear combinations of the original images, referred to as PCs (e.g., Oliver et al., 2000; Van Meirvenne and Goovaerts, 2002). For this reason, a PCA was used in our study to improve and summarize the interpretation of ARP measurements for the three depths (ARP1, ARP2 and ARP3) that contain a large proportion of redundant information. Therefore, the first step consisted of performing a PCA on the resistivity data resulting from the three depths of investigation. The PCA result was then mapped to summarize the resistivity signal for both the horizontal and vertical dimensions of soil. The second step involved a decomposition of each image, principal components in our case, into the low-frequency component or regional variability, the high-frequency component or local variability, and noise component (nugget effect).

2.3.4. Mapping procedures

The mapping procedures of soil thickness included the ordinary kriging and kriging with external drift. In the ordinary kriging, the predictor of $z(\mathbf{x}_0)$ for an unsampled location $\mathbf{x}_0$ is

$$z_{OK}{}^{*}(\mathbf{x}_0) = \sum_{\alpha=1}^{n} w_\alpha z(\mathbf{x}_\alpha), \tag{6}$$

where w_α are weights associated with the n sampling points. The weights sum to unity, a condition that ensures a zero error in expectation.

Kriging with external drift (KED) is a particular formulation of universal kriging (e.g., Goovaerts, 1997; Wackernagel, 1998). It allows the use of secondary information to account for the spatial variation of the local mean of the primary variable. The secondary variables are chosen for their strong correlation with the variable of interest and should be available at every location of the primary variable and every estimation point. KED consists of incorporating supplementary universality conditions about one or several external drift variables measured exhaustively in the spatial domain into the kriging system. For a thorough presentation of those methods, the reader should refer to books or papers on the subject, such as those by Goovaerts (1997), Wackernagel (1998), Bourennane and King (2003) and Bourennane et al. (2006).

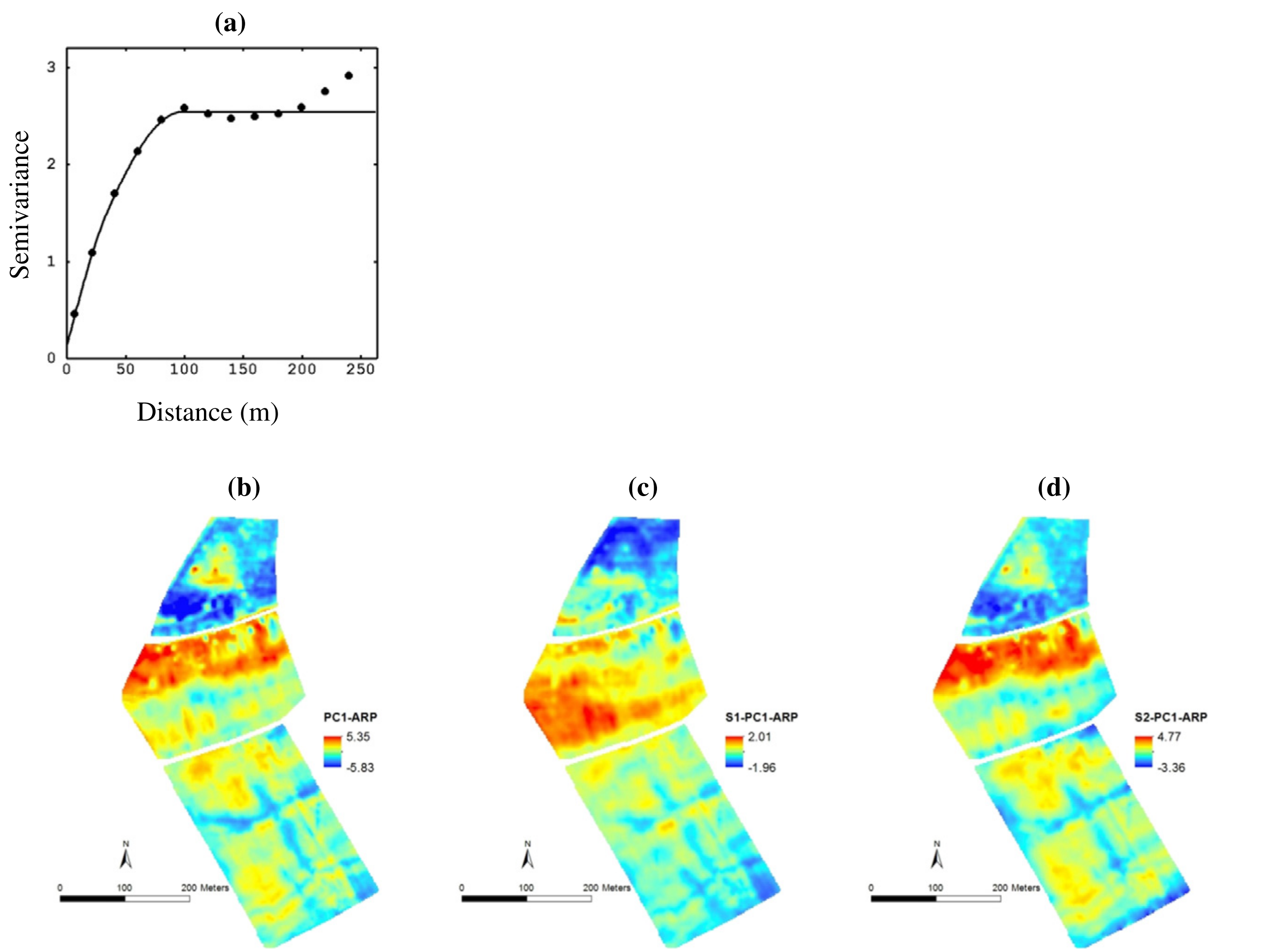

Fig. 4. Omnidirectional experimental (dots) and the theoretical (curves) variograms of the first principal component (a); maps of the decomposition of the original first principal component: PC1-ARP (b) into two structures (c) local: small scale (S1-PC1-ARP); (d) regional: large scale (S2-PC1-ARP).

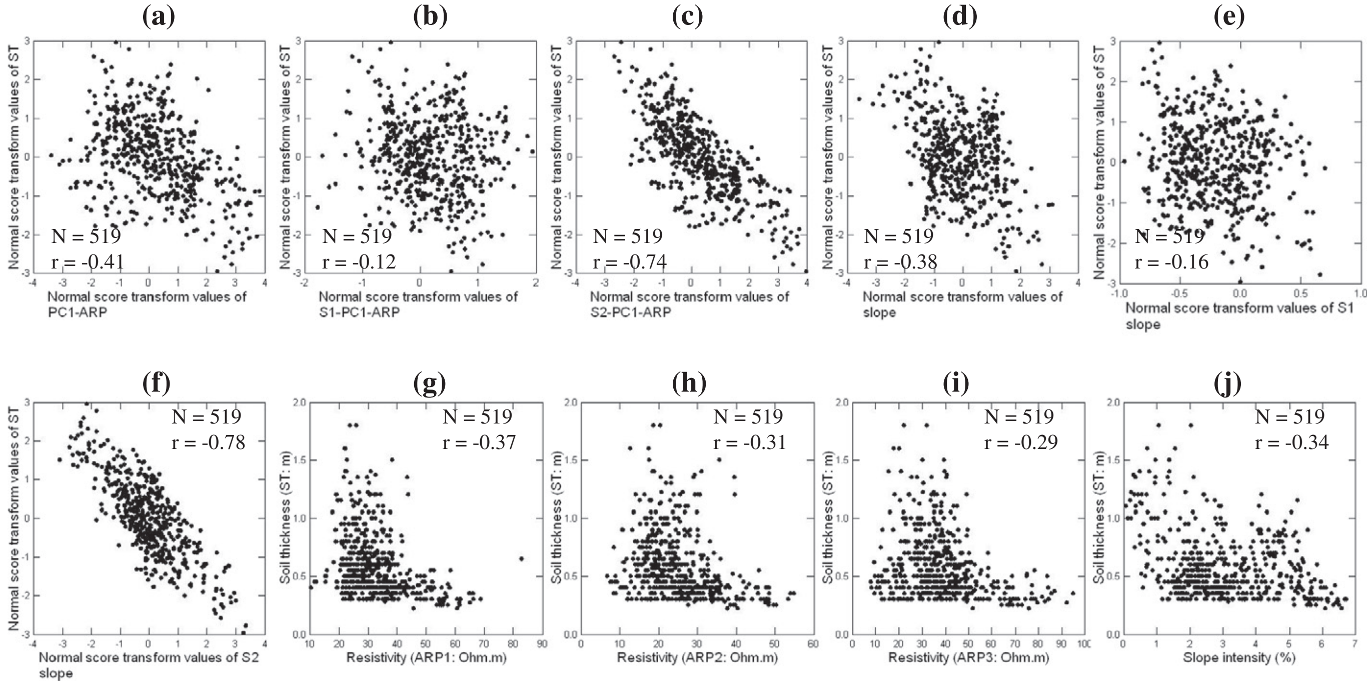

Fig. 5. Scattergrams between soil thickness (ST) and: PC1-ARP (a) and its decomposition by kriging into two structures (b) S1-PC1-ARP (small-scale); (c) S2-PC1-ARP (large-scale); slope (d) and its decomposition by kriging into two structures (e) S1-slope (small-scale); (f) S2-slope (large-scale); (g to j) original auxiliary data: ARP and slope intensity data vs ST original data.

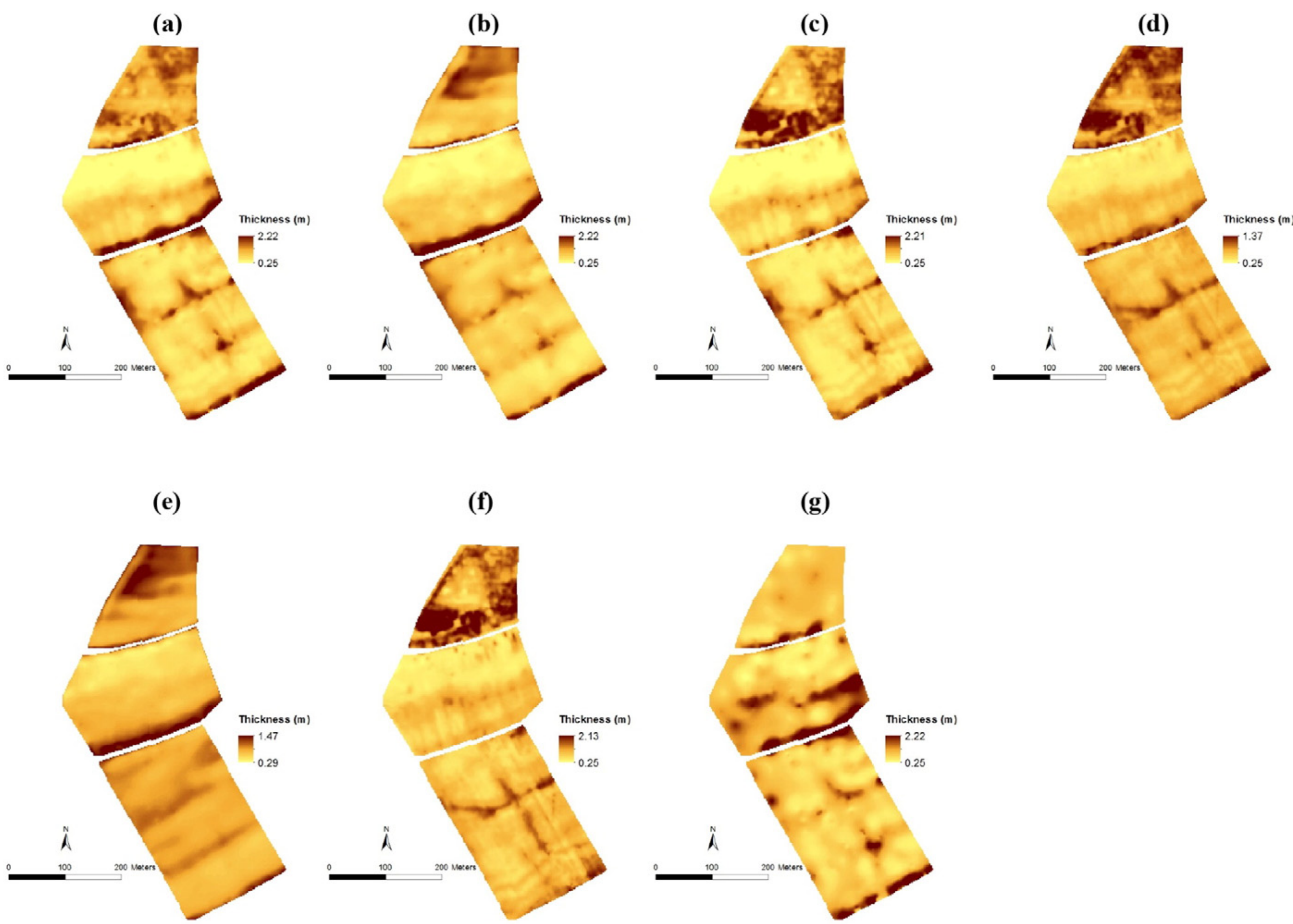

Fig. 6. Soil thickness predicted by kriging using: (a) large-scales of slope intensity and PC1-ARP as external drifts; (b) large-scale of slope as an external drift; (c) large-scale of PC1-ARP as an external drift; (d) slope intensity and PC1-ARP as external drifts; (e) slope as an external drift; (f) PC1-ARP as an external drift; (g) only punctual measurements of soil thickness.

Table 3
Mean prediction error (ME), root mean square error (RMSE) and relative improvement (RI) of measured versus predicted ST at each of the validation sites: statistics based on 131 samples.

	ME (m)	RMSE (m)	RI (%)
ST using both S2-slope and S2-PC1 as external drifts	−0.03	0.17	0
ST using S2-slope as an external drift	−0.07	0.25	−47
ST using S2-PC1 as an external drift	0.03	0.22	−29
ST using both slope and PC1 as external drifts	0.07	0.29	−70
ST using slope as an external drift	0.07	0.30	−76
ST using PC1 as an external drift	0.11	0.34	−100
ST using ordinary kriging	−0.05	0.29	−70

2.4. Validation procedure

The performances of mapping procedures were assessed using several criteria. The accuracy of the mapping procedures was assessed by computing (1) scatter plots of measured versus predicted values at each validation site, (2) the mean prediction error (ME) and (3) the root mean square error of prediction (RMSE), which are defined as follows:

$$\mathrm{ME} = \frac{1}{n}\sum_{i=1}^{n}[Z(\mathbf{x}_i) - Z^*(\mathbf{x}_i)] \tag{7}$$

$$\mathrm{RMSE} = \left\{\frac{1}{n}\sum_{i=1}^{n}[Z(\mathbf{x}_i) - Z^*(\mathbf{x}_i)]^2\right\}^{0.5} \tag{8}$$

The ME criterion was used to check the conditional bias property, which consists of assuming that prediction errors cancel out, leading to an unbiased estimator over the entire range of values. The RMSE criterion is a measure of accuracy of the various prediction methods. The RMSE value should be as small as possible for accurate prediction (e.g., Hyndman and Koehler, 2006). Finally, the relative improvement (RI) of accuracy was calculated by

$$\mathrm{RI} = \frac{(\mathrm{RMSE_R} - \mathrm{RMSE_E})}{\mathrm{RMSE_R}} \times 100 \tag{9}$$

where $\mathrm{RMSE_R}$ and $\mathrm{RMSE_E}$ are the root mean square errors for the reference and evaluated methods, respectively. Thus, if RI is positive, the accuracy of the evaluated method is superior to that of the reference and inferior otherwise (Zhang et al., 1992).

3. Results and discussion

3.1. Descriptive statistics

The descriptive statistics, as well as the values of the Kolmogorov-Smirnov statistic for normality test, for raw values of both the target variable and the secondary attributes are summarized in Table 1. Since the soil thickness (the target variable), as well as the slope intensity and ARP measurements (the secondary attributes), was found to be moderately to highly skewed (Fig. 2 and Table 1), normal score transforms were carried out beforehand for all data using the normal score transform (Goovaerts, 1997; Deutsch and Journal, 1998; Chilès and Delfiner, 1999):

$$y(\mathbf{x}_\alpha) = G^{-1}\left[\hat{F}(z(\mathbf{x}_\alpha))\right] \qquad \alpha = 1, \ldots, n \tag{10}$$

where $G^{-1}(.)$ is the inverse Gaussian cumulative distribution function (cdf) of the random function (RF) Y ($\mathbf{x}$), and $\hat{F}$ is the sample cumulative distribution of z.

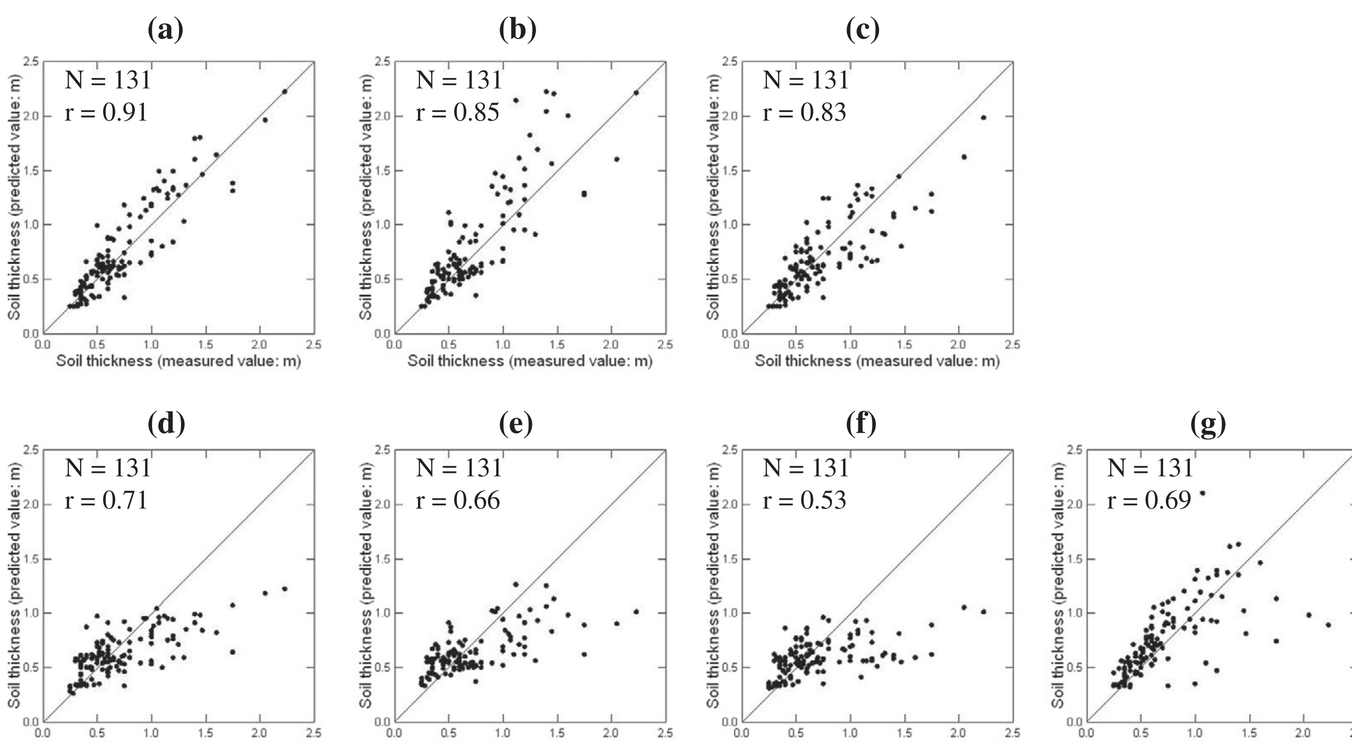

Fig. 7. Measured values of soil thickness versus predicted values by: (a) kriging using both ARP large-scale of PC1 and large-scale of slope intensity as external drifts; (b) kriging using large-scale of slope as an external drift; (c) kriging using ARP large-scale of PC1 as an external drift; (d) kriging using both PC1 of ARP and slope intensity as external drifts; (e) kriging using slope intensity as an external drift; (f) kriging using PC1 of ARP as an external drift; (g) ordinary kriging from only punctual measurement of soil thickness.

Indeed, methods based on second-order moments of distributions, such as PCA, are sensitive to skewed data (e.g., Goovaerts, 1997; Chilès and Delfiner, 1999). Obviously, back-transformation (normal score transformed to raw) into the original unit of the target variable is performed finally using

$$z^{(l)}\left(\mathbf{x}_j'\right) = \hat{F}^{-1}\left[\left(G\left(y^{(l)}\left(\mathbf{x}_j'\right)\right)\right] \qquad j = 1, \ldots, N \tag{11}$$

3.2. Kriging analysis of the secondary attributes

3.2.1. Kriging spatial components of slope intensity

To account for the variability of the slope intensity at both short and long distances, a nested variogram was used to fit the experimental variogram of the slope. The variability of slope intensity at short distances is ascribed to two types of linear anthropogenic landforms encountered over the study area: lynchets and undulations (Chartin et al., 2011). Thus, the experimental variogram of the normal score transform values of slope intensity (Slope-G) was calculated and modeled (Fig. 3a) with a nested variogram composed of a nugget and two spherical models. Table 2 summarizes the parameters of the experimental model and the fitted models. Thus, based on the model presented in Fig. 3a, geostatistical filtering has been used to decompose Slope-G into a long wavelength structure (regional component) corresponding to the global trend (Fig. 3c: S2-Slope-G) and a short wavelength structure (local component) corresponding to residual anomalies (Fig. 3b: S1-Slope-G).

3.2.2. ARP data: principal component analysis and kriging spatial components

The apparent resistivity, provided by Automatic Resistivity Profiling, contains a large proportion of redundant information. In addition, the resistivity values (ARP1, ARP2 and ARP3) are intrinsically correlated. Indeed, the codispersion coefficients (results not shown), which describe the correlation between the resistivity values measured at the three different depths as a function of the spatial scale, are constant. In other words, the codispersion coefficients are quasi-equal to the correlation coefficient values. Such a result allows us to conclude that the resistivity variability with depth is scale independent

Thus, prior to the decomposition of the raw signal into several spatial components, a classical method of multivariate analysis, namely, principal component analysis (PCA), was performed on the normal score transform values of resistivity. The first PC (PC1) accounted for 84.58% of the total variance, the second for 12.57%, and the third for 2.85%. In the sequel, we focused on PC1, as it accounts for most of the total variance measured.

The experimental variogram of PC1 was calculated and modeled (Fig. 4a) with the basic models that consisted of a nugget component and two spherical models. The parameters of the experimental model and the fitted models are summarized in Table 2. Thus, using these models and the kriging equation system, PC1 was mapped (Fig. 4b) on a 2 m regular grid over the entire study area. This map summarizes the relationships among the resistivity values measured at the three depths of investigation, and it accounts for >84% of the total variance measured. The positive score values correspond to the larger resistivity values and the negative values to smaller ones.

Geostatistical filtering was used to decompose PC1 into a long wavelength structure (regional component) corresponding to the global trend (Fig. 4d: S2-PC1-ARP) and a short wavelength structure (local component) corresponding to residual anomalies (Fig. 4c: S1-PC1-ARP).

3.3. Mapping soil thickness (ST)

The correlations between the original auxiliary data (ARP and slope intensity data) and soil thickness (ST original data) are weak and nonlinear (Fig. 5g to j). Conversely, the correlation between the first principal component resulting from the ARP measurements' PCA (PC1-ARP) and ST (Fig. 5a) is moderately negative and can be assumed to be linear. Anyway, in all cases, these linear relationships are statistically significant. The closeness between PC1-ARP and ST significantly increased when PC1 was filtered to isolate the regional component (Fig. 5c). It appears also that ST variation is not reflected in the local component of PC1-ARP (Fig. 5b).

Similar results (Fig. 5d to f) were observed when looking at relationships between soil thickness (ST) and the spatial components of the slope intensity.

Therefore, in the next step and as a first approach to mapping ST, raw values of the secondary attributes (ARP measurements and slope intensity) are replaced in the kriging equations when mapping ST over the entire study area (Fig. 6a) by the values of spatial components of the secondary attributes (S2-Slope and S2-PC1-ARP) that are the most strongly correlated with ST.

To examine the need to use two external drifts in predicting the target variable, two other maps of the ST variable were predicted. The first (Fig. 6b) was obtained by kriging using S2-Slope as an external drift and the second (Fig. 6c) using S2-PC1-ARP as an external drift.

Although the better correlations (Fig. 5) between the large-range components of the secondary variables and the target variable suggest that using these will improve upon using the raw secondary variables, it was necessary to demonstrate that improvements in predictions result from using the large-range components of the secondary variables rather than the raw secondary variables. Accordingly, ST was also predicted by kriging (Fig. 6d to f) using the raw secondary variables: PC1-ARP and slope intensity as external drifts; slope intensity as an external drift; and PC1-ARP as an external drift.

Finally, ST was also mapped (Fig. 6g) by ordinary kriging on a 2 m regular grid over the entire study area based on only the 519 punctual measurements of ST.

3.4. Validation results

Thus, each of the 131 individuals of the validation set was assigned a value of ST from each predicted map, and then, the criteria listed above were computed. Table 3 summarizes the different values of these criteria. The lowest bias (ME values: Table 3) is achieved by kriging accounting for the large scale components of the two external drifts (S2-slope and S2-PC1-ARP) and by kriging using S2-PC1-ARP as an external drift.

The RMSE values (Table 3) allow the conclusion that accounting for the large scale components of two exhaustive secondary data in the kriging of the ST variable is the most accurate prediction method. The relative improvement (RI) of the accuracy is at least equal to 29% (Table 3) between the approach accounting for both large scale components of secondary attributes in the spatial estimates of ST and the other approaches of estimates considered in this study. The results summarized in Fig. 7 and Table 3 indicate that improvements in predictions result from using the large-range components of the secondary variables rather than the raw secondary variables themselves. In addition, they stress the fact that the use of raw exhaustive variables in order to improve spatial estimates of a target variable can lead to inaccurate estimates compared to estimates based only on the measurement of the target variable (ST by ordinary kriging in this study).

4. Summary and conclusions

The focus of this paper has been on the spatial prediction of a target variable by capitalizing on the better correlation between the target variable and two exhaustive auxiliary attributes (slope intensity and electrical resistivity measurements: ARP) at specific spatial scales. The values of ARP and slope intensity were replaced in the kriging equations by the values of the spatial components that are the most strongly correlated with the soil thickness (target variable). For this purpose, the kriging spatial component technique was applied to separate the nuggets and small- and large-scale structures for both ARP and slope intensity. Then, the spatial structures in the ARP and slope intensity data, which are poorly correlated with soil thickness, were filtered out prior to integrating both ARP and slope intensity data into the soil thickness estimation. The filtering of nuggets and small scales of the secondary attributes greatly increases the weak correlations between soil thickness and auxiliary attributes (ARP measurements and slope intensity). Our findings also show that improvements in predictions result from using the large-scale components of the secondary variables rather than the raw secondary variables. The results also warn against the systematic use of a raw auxiliary variable to map a sparse target variable. In fact, the use of a noisy auxiliary variable, exhibiting weak-to-moderate correlation with the target variable, can lead to worse predictions compared to a prediction using only the measurements of the target variable.

In regard to our study, the correlation between electrical resistivity and soil thickness was improved by performing the kriging spatial component technique on the principal component achieved using the three measurement depths of electrical resistivity as variables in a principal component analysis. The results suggest that the correlation between soil thickness and the first principal component (PC1), resulting from a principal component analysis carried out on electrical resistivity measurements of the three prospecting depths, is improved from −0.41 to −0.74 when the nugget and the small structure of the resistivity data have been filtered out. A similar enhancement (−0.38 to −0.78) of the relationship between soil thickness and slope intensity was obtained when the nugget and small structure of the slope data have been filtered out. In terms of prediction, kriging with the large scale components of two external drifts was compared with ordinary kriging and kriging with either the large scale of slope or the large scale of the electrical resistivity's PC1 as an external drift to predict the soil thickness. To demonstrate that improvements in predictions result from using the large-scale components of the secondary variables rather than the raw secondary variables, soil thickness was also predicted by kriging using the raw secondary variables: PC1-ARP and slope intensity as external drifts; slope intensity as an external drift; and PC1-ARP as an external drift. The results indicated reasonably low bias of prediction by all of the approaches. The root mean square error values have shown that kriging accounting for two large scales of exhaustive variables is the most accurate.

One of the significant results of our study is that removing nuggets and small scale variability allows improvement of the correlation between electrical resistivity and soil thickness on the one hand and between slope and soil thickness on the other hand. This leads to an improvement in the spatial interpolation of the target variable. Another major result consists in the fact that our findings empirically demonstrate that the use of a raw auxiliary variable to map a sparse target variable can lead to worse predictions compared to a prediction using only the measurements of the target variable. This study also confirms the robustness and the generic aspect of noise filtering by kriging analysis in order to extract target features in redundant exhaustive information provided currently by various sensors. Furthermore, we can state that the originality of this work lies in the use of raw data of the apparent resistivity. In fact, strong assumptions underlie the inversion models. Thus, we have chosen to focus on a geostatistical technique that uses raw data and filters the spatial structures of the raw data that have low correlation with the target variable before integrating the filtered exhaustive data into the estimate of the target variable. However, further analysis could be performed in order to compare, in terms of mapping performances of a target variable, the approach developed in this study to an approach in which ARP data are inverted beforehand. In addition, we should keep in mind that by only considering PC1 in our approach, we admit to neglecting 16% of the variability in ARP measurements.

Finally, this study suggests that multivariate predictions of soil properties should use the information that is best correlated with the variable of interest. In the presence of nested variogram models, which indicate the existence of several scales of spatial variation, we should investigate whether the correlation between the variable of interest and auxiliary information can be enhanced by filtering some spatial structures using kriging of spatial components. For example, the weak correlation for a given spatial scale can hide the real correlation between raw measurements. Filtered auxiliary information can then be incorporated using kriging with external drift since it is available at all estimation grid nodes.

Acknowledgments

Financial support provided by the ANR (Agence Nationale de la Recherche) (ANR-08-VMCS-00) VMCS project LANDSOIL is gratefully acknowledged.

References

Afshar, F.A., Ayoubi, S., Besalatpour, A.A., Khademi, H., Castrignano, A., 2016. Integrating auxiliary data and geophysical techniques for the estimation of soil clay content using CHAID algorithm. J. Appl. Geophys. 126, 87–97.

Akumu, C.E., Woods, M., Johnson, J.A., Pitt, D.G., Uhlig, P., McMurray, S., 2016. GIS-fuzzy logic technique in modeling soil depth classes: using parts of the Clay Belt and Hornepayne region in Ontario, Canada as a case study. Geoderma 283, 78–87.

Alcaydé, G., Coubès, J., Macaire, J.-J., 1989. Feuille Loudoun (513), Carte géologique de la France (1:50000), Orléans, BRGM.

Allaire, S.E., Lange, S.F., Lafond, J.A., Pelletier, B., Cambouris, A.N., Dutilleul, P., 2012. Multiscale spatial variability of CO_2 emissions and correlations with physico-chemical soil properties. Geoderma 170, 251–260.

Bellemlih, S., 1999. Stocks particulaires holocènes et bilans de matières dans un bassin fluviatile en domaine sédimentaire - le bassin du Négron, Sud-ouest du Bassin Parisien, France. Ph.D. thesis. Université de Tours, France.

Besalatpour, A.A., Ayoubi, S., Hajabbasi, M.A., Mosaddeghi, M.R., Schulin, R., 2013. Estimating wet soil aggregate stability from easily available properties in a highly mountainous watershed. Catena 111, 72–79.

Besson, A., Cousin, I., Bourennane, H., Nicoullaud, B., Pasquier, C., Richard, G., Dorigny, A., King, D., 2010. The spatial and temporal organisation of the soil water content at the field scale as described by electrical resistivity measurements. Eur. J. Soil Sci. 61, 120–132.

Bocchi, S., Castrignano, A., Fornaro, F., Maggiore, T., 2000. Application of factorial kriging for mapping soil variation at field scale. Eur. J. Agron. 13, 295–308.

Bollinne, A., 1971. Les rideaux en Hesbaye gembloutoise - etude morphologique et sédimentologique. Bull. Soc. Géogr. Liège 7, 61–67.

Bourennane, H., King, D., 2003. Using multiple external drifts to estimate a soil variable. Geoderma 114, 1–18.

Bourennane, H., King, D., Chéry, P., Bruand, A., 1996. Improving the kriging of a soil variable using slope intensity as external drift. Eur. J. Soil Sci. 47, 473–483.

Bourennane, H., King, D., Le Parco, R., Isambert, M., Tabbagh, A., 1998. Three-dimensional analysis of soils and surface materials by electrical resistivity survey. Eur. J. Environ. Eng. Geophys. 3, 5–23.

Bourennane, H., Salvador-Blanes, S., Cornu, S., King, D., 2003. Scale of spatial dependence between chemical properties of topsoil and subsoil over a geologically contrasted area (Massif Central, France). Geoderma 112, 235–251.

Bourennane, H., Nicoullaud, B., Couturier, A., King, D., 2004. Exploring the spatial relationships between some soil properties and wheat yields in two soil types. Precis. Agric. 5, 521–536.

Bourennane, H., Dère, C., Lamy, I., Cornu, S., Baize, D., van Oort, F., King, D., 2006. Enhancing spatial estimates of metal pollutants in raw wastewater irrigated fields using a topsoil organic carbon map predicted from aerial photography. Sci. Total Environ. 361, 229–248.

Bourennane, H., Nicoullaud, B., Couturier, A., Pasquier, C., Mary, B., King, D., 2012. Geostatistical filtering for improved soil water content estimation from electrical resistivity data. Geoderma 183-184, 32–40.

Bourennane, H., Couturier, A., Pasquier, C., Chartin, C., Hinschberger, F., Macaire, J.J., Salvador-Blanes, S., 2014. Comparative performance of classification algorithms for the development of models of spatial distribution of landscape structures. Geoderma 119-220, 136–144.

Bourgault, G., 1994. Robustness of noise filtering by kriging analysis. Math. Geol. 6, 733–752.

Boutin, D., Froger, D., Rassineux, J., 1990. Feuille Loudun (1724–1624), Carte des sols du Département de la Vienne et de la région Centre au 1:50000, Chambre d'Agriculture de la Vienne - IGN - INRA.

Castrignano, A., Giugliarini, L., Risaliti, R., Martinelli, N., 2000. Study of spatial relationships among some soil physico-chemical properties of a field in central Italy using multivariate geostatistics. Geoderma 97, 39–60.

Chartin, C., Bourennane, H., Salvador-Blanes, S., Hinschberger, F., Macaire, J.-J., 2011. Classification and mapping of anthropogenic landforms on cultivated hillslopes using DEMs and soil thickness data — example from the SW Parisian Basin, France. Geomorphology 135, 8–20.

Chilès, J.P., 1991. Application du krigeage avec dérive externe à l'implantation d'un réseau de mesures piézométriques. Sci. Terre Sér. Inf. 30, 131–147.
Chilès, J.P., Delfiner, P., 1999. Geostatistics: Modeling Spatial Uncertainty. Wiley, New York.
Chilès, J.P., Guillan, A., 1984. Variogrammes et krigeages pour la gravimétrie et le magnétisme. In: Royer, J.-J. (Ed.)Computer in Earth Sciences for Natural Resources Characterization Vol. 20. Sciences de la Terre, Nancy, France, pp. 455–468.
Coscia, I., Linde, N., Greenhalgh, S., Günther, T., Green, A., 2012. A filtering method to correct time-lapse 3D ERT data and improve imaging of natural aquifer dynamics. J. Appl. Geophys. 80, 12–24.
Davydenko, A.Y., Grayver, A.V., 2014. Principal component analysis for filtering and leveling of geophysical data. J. Appl. Geophys. 109, 266–280.
De Benedetto, D., Castrignano, A., Sollitto, D., Modugno, F., Buttafuoco, G., lo Papa, G., 2012. Integrating geophysical and geostatistical techniques to map the spatial variation of clay. Geoderma 171–172, 53–63.
De Fouquet, Ch., Malherbe, L., Ung, A., 2011. Geostatistical analysis of the temporal variability of ozone concentrations. Comparison between CHIMERE model and surface observations. Atmos. Environ. 45, 3434–3446.
Deutsch, C., Journal, A., 1998. GSLIB: Geostatistical Software Library and User's Guide. Oxford University Press, Oxford. UK.
Diek, S., Temme, A.J.A.M., Teuling, A.J., 2014. The effect of spatial soil variation on the hydrology of a semi-arid Rocky Mountains catchment. Geoderma 235-236, 113–126.
Dietrich, W.E., Reiss, R., Hsu, M.L., Montgomery, D.R., 1995. A process-based model for colluvial soil depth and shallow landsliding using digital elevation data. Hydrol. Process. 9, 383–400.
Dobermann, A., Goovaerts, P., George, T., 1995. Sources of soil variation in an acid Ultisol of the Philippines. Geoderma 68, 173–191.
Dobermann, A., Goovaerts, P., Neue, H.U., 1997. Scale-dependent correlations among soil properties in two tropical lowland rice fields. Soil Sci. Soc. Am. J. 61, 1483–1496.
FAO, 1998. World reference base for soil resources. Food and Agriculture Organization of the United Nations. World Soil Resources Report No 84. FAO, Rome.
Gessler, P.E., Moore, I.D., Mckenzie, N.J., Ryan, P.J., 1995. Soil–landscape modelling and spatial prediction of soil attributes. Int. J. Geogr. Inf. Syst. 9, 421–432.
Goovaerts, P., 1997. Geostatistics for Natural Resources Evaluation. Oxford University Press.
Goovaerts, P., Webster, R., 1994. Scale-dependent correlation between topsoil copper and cobalt concentrations in Scotland. Eur. J. Soil Sci. 45, 79–95.
Goovaerts, P., Jacquez, G.M., Marcus, A., 2005. Geostatistical and local cluster analysis of high resolution hyperspectral imagery for detection of anomalies. Remote Sens. Environ. 95, 351–367.
Goulard, M., Voltz, M., 1992. Linear coregionalization model: tool for estimation and choice of cross-variogram matrix. Math. Geol. 24, 269–286.
Herbst, M., Diekkrüger, B., Vereecken, H., 2006. Geostatistical co-regionalization of soil hydraulic properties in a micro-scale catchment using terrain attributes. Geoderma 132, 206–221.
Hyndman, R.J., Koehler, A.B., 2006. Another look at measures of forecast accuracy. Int. J. Forecast. 22, 679–688.
Kim, D., Zheng, Y., 2011. Scale-dependent predictability of DEM-based landform attributes for soil spatial variability in a coastal dune system. Geoderma 164, 181–194.
Lin, Y.P., 2002. Multivariate geostatistical methods to identify and map spatial variations of soil heavy metals. Environ. Geol. 42, 1–10.
Ma, Y.Z., Royer, J.J., 1988. Local geostatistical filtering: application to remote sensing. Sci. Terre Sér. Inf. 27, 17–36.
Macaire, J.-J., Bellemlih, S., Di Giovanni, C., De Luca, P., Visset, L., Bernard, J., 2002. Sediment yield and storage variations in the Negron river catchment (south western Parisian Basin, France) during the Holocene period. Earth Surf. Process. Landf. 27, 991–1009.
Matheron, G., 1982. Pour une analyse krigeante des données régionalisées. Internal Note N-732. Fontainebleau, Centre de Géostatistique.
Mehnatkesh, A.M., Ayoubi, S., Jalalian, A., Sahrawat, K.L., 2013. Relationships between soil depth and terrain attributes in a semi-arid hilly region in western Iran. J. Mt. Sci. 10, 163–172.
Milne, A.E., Webster, R., Ginsburg, D., Kindred, D., 2012. Spatial multivariate classification of an arable field into compact management zones based on past crop yields. Comput. Electron. Agric. 80, 17–30.
Minsley, B.J., Smith, B.D., Hammack, R., Sams, J.I., Veloski, G., 2012. Calibration and filtering strategies for frequency domain electromagnetic data. J. Appl. Geophys. 80, 56–66.
Oliver, M.A., Webster, R., Slocum, K., 2000. Filtering SPOT imagery by kriging analysis. Int. J. Remote Sens. 21, 735–752.
Renard, D., Nai-Hsien, M., 1988. Utilisation de dérives externes multiples. Sci. Terre Sér. Inf. 28, 281–301.
Salvador-Blanes, S., Cornu, S., Couturier, A., King, D., Macaire, J.-J., 2006. Morphological and geochemical properties of soil accumulated in hedge-induced terraces in the Massif Central, France. Soil Tillage Res. 85, 62–77.
Sandjivy, L., 1984. The factorial kriging analysis of regionalized data. Its application to geochemical prospecting. In: Verly, G., David, M., Journel, A.G., Marechal, A. (Eds.), Geostatistics for Natural Resources Characterization. Reidel, Dordrecht, pp. 559–571.
Tesfa, T.K., Tarboton, D.G., Chandler, D.G., McNamara, J.P., 2009. Modeling soil depth from topographic and land cover attributes. Water Resour. Res. 45, W10438. http://dx.doi.org/10.1029/2008WR007474.
Van Meirvenne, M., Goovaerts, P., 2002. Accounting for spatial dependence in the processing of multi-temporal SAR images using factorial kriging. Int. J. Remote Sens. 2, 371–387.
Wackernagel, H., 1998. Multivariate Geostatistics. Springer-Verlag.
Wahren, F.T., Julich, S., Nunes, J.P., Gonzalez-Pelayo, O., Hawtree, D., Feger, K.H., Keizer, J.J., 2016. Combining digital soil mapping and hydrological modeling in a data scarce watershed in north-central Portugal. Geoderma 264, 350–362.
Webster, R., Atteia, O., Dubois, J.P., 1994. Coregionalization of trace metals in the soil in the Swiss Jura. Eur. J. Soil Sci. 45, 205–218.
Zhang, R., Warrick, A.W., Myers, D.E., 1992. Improvement of the prediction of soil particle size fractions using spectral properties. Geoderma 52, 223–234.
Zhu, Q., Lin, H., 2011. Influences of soil, terrain, and crop growth on soil moisture variation from transect to farm scales. Geoderma 163, 45–54.