Geoderma 213 (2014) 571–577

Contents lists available at ScienceDirect

Geoderma

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

Using bivariate multiple-point statistics and proximal soil sensor data to map fossil ice-wedge polygons

E. Meerschman *[,a], M. Van Meirvenne [a], G. Mariethoz [b], M.M. Islam [a], P. De Smedt [a], E. Van De Vijver [a], T. Saey [a]

[a] Research Group Soil Spatial Inventory Techniques, Department of Soil Management, Faculty of Bioscience Engineering, Ghent University, Coupure 653, 9000 Gent, Belgium
[b] National Centre for Groundwater Research and Training, School of Civil and Environmental Engineering, University of New South Wales, UNSW 2052, NSW, Sydney, Australia

ARTICLE INFO

Article history:
Received 25 April 2012
Received in revised form 3 December 2012
Accepted 27 January 2013
Available online 16 March 2013

Keywords:
Multivariate simulation
Pedometrics
Proximal soil sensor data
Multiple-point statistics
Training image

ABSTRACT

Multiple-point statistics (MPS) is a collection of geostatistical simulation algorithms that uses a multiple-point training image (TI) as structural model instead of a two-point variogram. MPS allows to simulate more complex random fields, like phenomena characterized by spatial connectivity. A very recent development is multivariate MPS in which an ensemble of variables can be simulated simultaneously using a multivariate TI. We investigated if multivariate MPS can be used for the processing of proximal soil sensor data, i.e. interpolating the sensor data and predicting the target variable. We measured a field with fossil ice-wedge polygons in the subsoil with an electromagnetic induction sensor and used these measurements to predict the location of wedge material in the subsoil. We built a bivariate TI with a categorical image of a random polygonal network as primary variable and a continuous image of the corresponding sensor values as secondary variable. Then, we performed a bivariate reconstruction with the recently developed Direct Sampling software. The resulting E-types provided an interpolated sensor data map and a probability map predicting the location of wedge material in the subsoil. This procedure was compared to the more traditional approach of interpolating the sensor data with ordinary kriging and performing a fuzzy *k*-means classification. Comparing the resulting maps with an aerial photograph revealing the location of the ice-wedges through polygonal crop marks, showed that MPS reconstructed the polygonal patterns much better. The local accuracy of the MPS maps was proven by an independent quantitative validation based on nine extra measurement lines and 94 bore hole samples. As a first application in soil science, our case study showed that multivariate MPS can be used for the processing of proximal soil sensor data. The flexibility of the technique opens perspectives for other new applications and therefore multivariate MPS can become a valuable part of the pedometrical toolbox.

1. Introduction

The key function in traditional geostatistics is the variogram, which is used as a model of the spatial structure. However, this two-point statistic is often not able to characterize complex random fields, such as phenomena showing spatial connectivity. To map complex random fields, multiple-point statistics (MPS) need to be considered (Guardiano and Srivastava, 1993). The fundamental idea of MPS is to replace the two-point variogram by a multiple-point training image (TI). A TI is a conceptual image of the expected spatial structure and is often built based on prior knowledge. A very recent development in MPS is multivariate MPS, in which an ensemble of variables can be simulated simultaneously using a multivariate TI (Mariethoz et al., 2010).

MPS was developed in petroleum geology and hydrogeology (Strebelle, 2002) and to date most of its applications can be found in these fields (e.g. Comunian et al., 2011; Huysmans and Dassargues, 2009; Le Coz et al., 2011; Ronayne et al., 2008; Strebelle et al., 2003; Zhang et al., 2006). Complex patterns, that are hard to model with traditional two-point geostatistics, also appear in soil science: a.o. dune patterns, paleochannels, limestone pavement, desiccation cracks, (relict) patterned ground, land-use patterns, sedimentary rock layers and soil pores. However, the use of MPS in the processing of soil data is still an open research question. In this paper, we investigated whether multivariate MPS can be used for the processing of proximal soil sensor data.

Proximal soil sensing is an increasingly used data source for soil inventory (McBratney et al., 2000). In a mobile setup, these sensors allow to rapidly collect indirect observations of the subsoil in a non-destructive way (Adamchuk et al., 2004). Processing proximal soil sensor data typically includes two steps: first the sensor data need to be interpolated to a regular grid and then this map can be used as a proxy to predict the target variable (de Gruijter et al., 2010).

Even though proximal soil sensor data are considered as high-resolution data, interpolating the data to a regular grid remains a crucial processing step. Sensor sampling is typically done with a sensor attached to a vehicle taking measurements at fixed intervals while driving

* Corresponding author. Tel.: +32 9 2646042; fax: +32 9 2646247.
E-mail address: eef.meerschman@ugent.be (E. Meerschman).

0016-7061/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.geoderma.2013.01.016

along parallel lines. With the instruments available today, the within-line sampling density is mostly no longer a limiting factor. The between-line distance, on the other hand, largely affects the costs of a field survey. Generally, this distance should be chosen based on the expected scale of the soil features one wants to map. Apart from interpolating the sensor data between measurement lines, spatial interpolation is required to complete the data set when some areas are inaccessible for the sensor survey.

To date, ordinary kriging (OK) is an often used method to interpolate sensor data because of its declustering ability. OK is a traditional geostatistical estimation technique based on a two-point variogram (Goovaerts, 1997). In our experience, OK is a successful method to interpolate sensor data. However, when the sensor data reflect subsoil phenomena that have a complex spatial pattern or are highly spatially connected, the two-point variogram is no longer sufficient. In practice, problems arise when the between-line distance is larger compared to the scale of the investigated soil features. Hence, it is worth investigating whether MPS can serve as a more suited interpolation technique for these situations.

If the sensor variable differs from the target variable (i.e. the soil variable of interest), a model is needed to predict the target variable from the sensor variable, which then serves as an ancillary or secondary variable (de Gruijter et al., 2010). For example, if the sensed variable is electrical resistivity and the variable of interest is porosity, the modeling of the relationship between these two attributes is critical. Depending on the specific situation and the type of target variable, a variety of pedometrical techniques can be used for this aim, ranging from numerical classification to CLORPT and hybrid techniques (McBratney et al., 2000). For instance, fuzzy *k*-means is an often used predictive classification technique to delineate zones with homogeneous soil properties based on proximal soil sensor data (Cockx et al., 2006, 2007; Islam et al., 2011; Vitharana et al., 2008b). Examples of CLORPT techniques are predicting the depth to contrasting soil layers from proximal soil sensor data with inverse modeling techniques (De Smedt et al., 2011; Saey et al., 2008, 2009) or predicting the soil clay content based on neural network approaches (Cockx et al., 2009). Vitharana et al. (2008a) used regression kriging to predict the depth to clay substratum and Triantafilis et al. (2001) compared different hybrid techniques to predict soil salinity from proximal soil sensor data.

Multivariate MPS is promising for both the interpolation of sensor data and the prediction of the target variable. This technique is mainly developed for situations where one variable is (partially) known and the other is to be simulated (the collocated simulation paradigm). Using a bivariate TI is especially interesting when the relationship between the variables is known through training data but cannot simply be expressed as a mathematical relationship (Mariethoz et al., 2010; Meerschman et al., 2013). To investigate the use of multivariate MPS, we applied it to a case study aiming to predict the location of fossil ice-wedge polygons in the subsoil based on electromagnetic induction (EMI) data.

Fossil ice-wedges polygons are a clear example of spatially connected subsoil features. They are remnants of thermal contraction cracks that were formed during glacial periods (Kolstrup, 1986). At the end of the glaciation these soil cracks were filled up and covered with wind and water transported sediments (French, 2007). Hence, today fossil ice-wedge polygons can be recognized as polygonal networks in the subsoil that are filled with soil material (wedge material) that is younger than the surrounding material (host material). Mapping these cryogenic features is of interest since they cause abrupt changes in the subsoil composition, possibly inducing preferential flow paths for e.g. agro-contaminants or nutrients. Furthermore, the morphology of this polygonal network is important for paleoclimatological reconstructions (Plug and Werner, 2002, 2008). It has recently been shown that EMI sensors are an effective aid in the mapping of fossil ice-wedge polygons, especially when the textural contrast between the wedge material and the host material is sufficiently strong (Cockx et al., 2006; Meerschman et al., 2011).

In this paper, we used an EMI sensor to measure a field with fossil ice-wedge polygons in the subsoil. Then, we applied bivariate MPS to interpolate the proximal soil sensor data to a regular grid and to simultaneously derive a map estimating the location of the fossil ice-wedge polygons in the subsoil. To set a comprehensive framework for the evaluation of the new method's prediction performance, we compared it with the often applied procedure of interpolating the sensor data with OK and then performing a fuzzy *k*-means classification to derive the possibility of finding wedge material in the subsoil.

2. Material and methods

2.1. Study area and data collection

Fig. 1a shows an oblique aerial photograph of an agricultural field in Belgium (central coordinates: 51°01′16″ N, 3°29′41″ E). The photograph was taken on 4 August 1996 when sugar beets were cultivated on the field. At that moment polygonal crop marks revealed the presence of an underlying network of fossil ice-wedges. Besides this

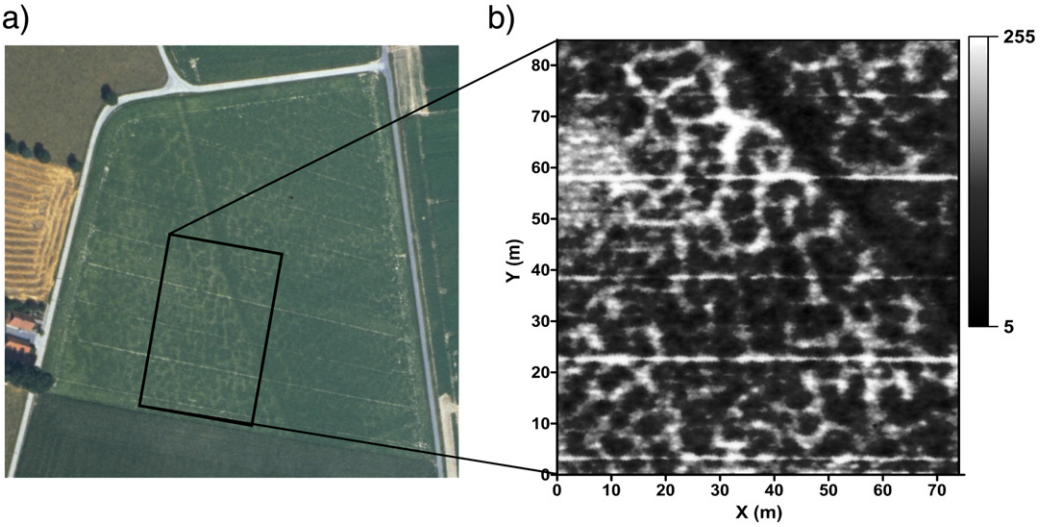

Fig. 1. (a) Aerial photograph taken on 4 August 1996 showing polygonal crop marks and a former field track (north-southeast oriented) with delineation of the study area (large rectangle) and (b) same aerial photograph after georectification, clipping, and color stretching. Coordinates are according to the Belgian metric Lambert-72 projection.
a, J. Bourgeois, Department of Archaeology and Ancient History of Europe, Ghent University, Belgium, photo: J. Semey.

polygonal pattern, the aerial photograph also shows a former field track from north to southeast and east–west stripes due to operation lines from farming vehicles. The photograph was georectified and color stretched to enhance the contrasts, after which a test area of 0.63 ha was selected and clipped (Fig. 1b) (Meerschman et al., 2011).

A fossil ice-wedge was partly exposed by excavating a part of the field to a depth of 0.9 m (Meerschman et al., 2011). The wedges were formed in the Tertiary material which was covered by wind-blown loess at the end of the last Weichselian Glacial stage. This loess layer has an approximate thickness of 0.6 m. The wedges were between 0.3 and 1.2 m wide and their infillings contained more sand than the host material. Because of this difference in soil texture, we surveyed the field with an EM38-DD electromagnetic induction sensor measuring the apparent electrical conductivity (*ECa*) of an underlying soil volume in mS m^{-1} (Corwin and Lesch, 2005). We used a mobile setup with a between-line measurement distance of approximately 3 m and a within-line distance of 0.4 m (Fig. 2a). More details about the first processing steps, including temperature correction, drift correction, noise removal and spatial trend removal, can be found in Meerschman et al. (2011). The spatial trend was caused by larger *ECa* values at the location of the former field track, and was subtracted from the sensor data to highlight the polygon boundaries. Fig. 2a shows the final data set, i.e. the residuals $\Delta ECa = ECa -$ spatial trend, referred to hereafter as 'sensor data'.

To validate the interpolated sensor data maps we used nine extra measurement lines with a between-line distance of approximately 9 m and a within-line distance of 0.4 m (Fig. 2b). The lines were positioned in the middle of two conditioning data lines. They were measured on the same day as the conditioning data and were processed in a similar way. To validate the prediction of wedge material in the subsoil, we took 94 subsoil samples (0.6 m–0.8 m): half of them were sampled according to a regular grid and half of them were simply randomly sampled. We analyzed their textural fractions: 43 samples were classified as wedge material and 51 as host material (Fig. 2b) (Meerschman et al., 2011).

2.2. Traditional two-point geostatistics and predictive classification

First, the sensor data were interpolated to a regular grid (cell size 0.1 m × 0.1 m) with OK using a spherical variogram model ($C_0 = 0$, $C_1 = 1.7$, $a = 4.3$ m) (Fig. 3) (Goovaerts, 1997). The model was fit to the experimental variogram considering only data pairs in the direction of the driving lines. This directional variogram was more stable than the omnidirectional one which showed a jump at lag distances around 3 m, corresponding to the between-line distance. This strategy could be applied since we assumed that the anisotropy shown by the experimental variograms was caused by the sampling configuration, whereas the spatial process being studied was assumed isotropic. We defined an elliptical search window with the longest radius perpendicular to the driving direction to ensure that neighbors from different measurement lines were selected.

Then, we performed a fuzzy *k*-means classification of the interpolated sensor data. Since fuzzy-set theory allows dealing with uncertainty especially due to imprecise boundaries between categories (McBratney and Odeh, 1997), this technique was appropriate to classify the soil into two classes: one area with host material and one with wedge material in the subsoil (from 0.6 to 0.8 m depth). Although theoretically required for predictive classification (de Gruijter and McBratney, 1988), we did not add an extragrade class here since this would impede comparison with the MPG probability map. We used the FuzME software (Minasny and McBratney, 2002) and set the fuzziness exponent φ at 2.1 following the scheme proposed by McBratney and Moore (1985). Parameter φ controls the degree of fuzziness of the classification and has a value between 1 (hard classification) and ∞. The resulting fuzzy membership map for the wedge material class was interpreted as the possibility to find wedge material in the subsoil.

2.3. Multiple-point geostatistics

MPS requires the construction of a TI. In the case of bivariate MPS, a bivariate TI is needed. For this case study the TI needed to consist of a categorical image of the target variable (TI1), i.e. an indicator for the presence of wedge material in the subsoil, and a continuous image of the ancillary variable (TI2), i.e. the sensor data. Both TI1 and TI2 needed to represent the expected spatial structure of the corresponding variable and the bivariate image needed to represent the expected relationship between both variables. We built this bivariate TI based on our physical knowledge of the crack formation and the sensor measurements on the one hand, and the information we gathered during the field work on the other hand, i.e. the excavation and the prediction sensor data (Fig. 2a).

TI1 was built from a binary image of a polygonal network of desiccation cracks in a Mexico silt loam, that we selected from literature (Baer et al., 2009). We resized the image to an image of 700 pixels high and 700 pixels wide (bicubic interpolation), each pixel representing an area of 0.01 m^2. Then, we dilated the wedges considering the width of the

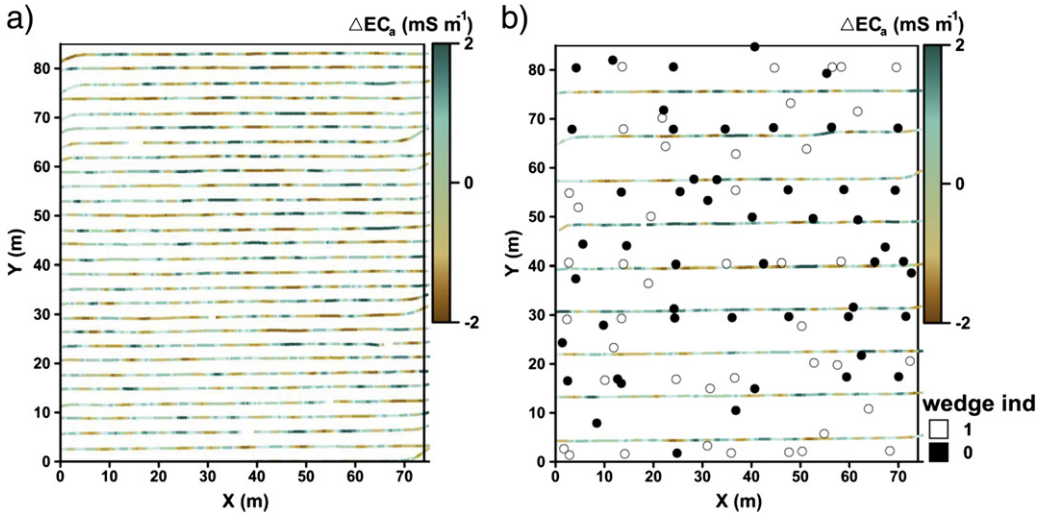

Fig. 2. Overview of the data collection with (a) the conditioning data: 28 measurement lines (in-line distance 0.4 m) of proximal soil sensor data (ΔECa in mS m^{-1}) and (b) the validation data: 9 measurement lines (in-line distance 0.4 m) of proximal soil sensor data (ΔECa in mS m^{-1}) and 94 classified bore hole samples indicating the presence (wedge indicator 1) or absence (wedge indicator 0) of wedge material in the subsoil (0.6–0.8 m).

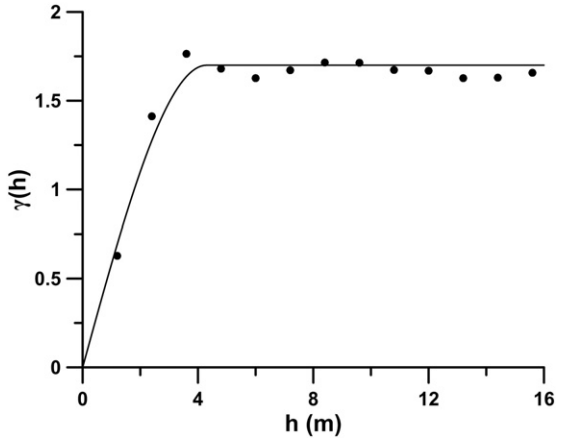

Fig. 3. Spherical variogram model with $C_0 = 0$, $C_1 = 1.7$ and $a = 4.3$ m used to interpolate the proximal soil sensor data with traditional two-point geostatistics (ordinary kriging).

excavated polygon (see 2.1). Fig. 4a shows the resulting image that was used as TI1. TI2 was obtained by a forward modeling procedure predicting the corresponding sensor data starting from TI1. We spatially filtered TI1 with a kernel (11×11 pixels) representing the depth–response curve of the EM38DD soil sensor (McNeill, 1980). This filtered image was histogram transformed targeting the histogram of the sensor data (Fig. 2a). The continuous TI (TI2) is shown in Fig. 4b. The image processing steps were performed in Matlab (Mathworks, R2011a).

We used the Direct Sampling (DS) code to generate bivariate multiple-point simulations (Mariethoz et al., 2010). As most of the MPS algorithms, DS is a sequential simulation technique. This means that the non-observed locations $\mathbf{x}$ of the simulation grid are visited according to a predefined (random or regular) path and that for each $\mathbf{x}$ a simulated value is drawn from a cumulative distribution function $F(z, \mathbf{x}, \mathbf{d}_n(\mathbf{x})) = Prob\{Z(\mathbf{x}) \leq z | \mathbf{d}_n(\mathbf{x})\}$ conditioned to a data event $\mathbf{d}_n(\mathbf{x})$ of size n centered at $\mathbf{x}$. This data event comprises the values of the n known neighboring grid nodes $\mathbf{x}_i$ ($i = 1,\ldots,n$), i.e. the conditioning data and the already simulated grid nodes, and their relative positions. Hence, MPS considers the n neighboring locations jointly, instead of pairwise.

Most MPS algorithms build $F(z, \mathbf{x}, \mathbf{d}_n(\mathbf{x}))$ by scanning the TI beforehand for replicates of all possible $\mathbf{d}_n(\mathbf{x})$'s (based on a predefined template) and storing the TI probabilities in a catalog (Straubhaar et al., 2011; Strebelle, 2002). The conditional probability is then calculated as the ratio of the number of replicates with their central node equal to z and the total number of replicates found. Therefore, these algorithms are restricted to categorical variables.

Conversely, DS can handle continuous and multivariate cases. This algorithm skips the prior scanning step and directly samples the TI during simulation. As soon as a TI pattern is found that matches $\mathbf{d}_n(\mathbf{x})$ exactly or as soon as the distance between the TI pattern and $\mathbf{d}_n(\mathbf{x})$ is lower than a user-defined threshold, the value at the central node of the TI pattern is directly pasted to $\mathbf{x}$. Different dissimilarity distances can be selected, depending on the application and the type of variable (Mariethoz et al., 2010). Hereafter, we briefly discuss the DS parameters that were particularly interesting for this case study. More details about DS and its implementation can be found in Mariethoz et al. (2010) and Meerschman et al. (in press). For practical guidelines about how to set the DS input parameters, the reader is referred to Meerschman et al. (2013).

DS generates multivariate simulations by defining a path through all the non-observed grid nodes x for each of the m variables. This means that when one variable is simulated at one location, the other variable at the same location can be simulated later in the path. For each x a multivariate data event $\mathbf{d}_n(\mathbf{x})$ is defined that contains the neighboring data for the m variables. Based on a weighted average of the m selected dissimilarity distances, the multivariate TI pattern is chosen that is most similar to the multivariate $\mathbf{d}_n(\mathbf{x})$ and the value at the central node of this TI pattern is pasted in the simulation grid at location x. Both the type of dissimilarity distance and the weight given to each distance w_m are user-defined (Mariethoz et al., 2010). In this paper we used the fraction of non-matching nodes for the categorical variable and the mean absolute error for the continuous variable. The continuous variable was given a weight three times larger than the categorical variable.

If conditioning data are given for all or some of the m variables, they will be honored by assigning them to the closest grid node prior to sequential simulation. Although local accuracy is assured this way, it is important that the assigned grid nodes are embedded in the spatial pattern and do not appear as noisy pixels. Therefore, DS allows to give the conditioning data grid nodes a higher weight when calculating the distance between $\mathbf{d}_n(\mathbf{x})$ and the TI pattern. In this study the weight given to the conditioning data was set five times larger than the weight given to the already simulated grid nodes.

We ran 50 bivariate simulations with the constructed bivariate TI (Fig. 4) and the sensor data as continuous conditioning data (Fig. 2a). The resulting E-type for the continuous variable served as an interpolated

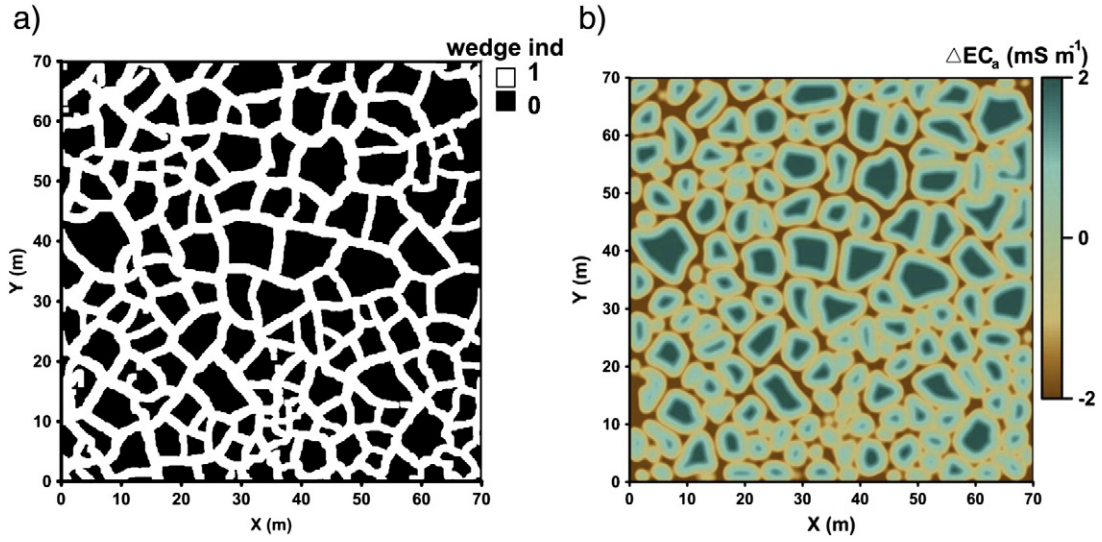

Fig. 4. Bivariate TI used to interpolate the proximal soil sensor data and predict the target variable with multiple-point geostatistics with (a) the categorical image representing the spatial pattern of the wedge indicator (TI1) and (b) the continuous image representing the spatial pattern of the sensor data (TI2).

sensor data map and the E-type for the categorical (binary) variable served as a probability map for the presence of wedge material in the subsoil.

2.4. Validation

The interpolated sensor data maps were validated by comparing the measured sensor values in the independent measurement lines (Fig. 2b) with the estimated values at the closest grid node. For both the map interpolated with two-point geostatistics and the one interpolated with multiple-point geostatistics, we made a scatterplot and calculated five validation indices: the mean estimation error (MEE), the root mean square estimation error (RMSEE), the mean absolute estimation error (MAEE), the Pearson's correlation coefficient (r) and the Spearman's rank correlation coefficient (r_R).

Based on the 94 classified bore hole samples (Fig. 2b), we validated the two maps predicting the presence of wedge material in the subsoil by calculating their receiver-operating characteristic (ROC) curve (Pontius and Schneider, 2001). This method was chosen since a ROC curve evaluates the two-class prediction performance of the maps independent of the chosen decision threshold. This is important to compare the fuzzy membership value map more objectively with the probability map derived with MPS. The effect of the degree of fuzziness, as is defined by φ, will not influence the comparison. The ROC space is defined by the 1-specificity (false positive rate) and the sensitivity (true positive rate) as x- and y-axes respectively, considering a continuous range of decision thresholds. The top left corner is the optimal location of the ROC space since there both the specificity and the sensitivity are 1. The area under the ROC curve (AUC) measures the two-class prediction performance. An AUC of 0.5 indicates a classification performance no better than chance. The closer the AUC is to 1, the better is the classification potential of the maps (Cockx et al., 2007).

3. Results and discussion

Fig. 5 shows the maps generated with traditional two-point geostatistics (left) and multiple-point geostatistics (right). When we compare these maps with the georectified aerial photograph of the polygonal crop marks (Fig. 1b), it is clear that both maps delineate the major ice-wedges very well, especially considering the between-line

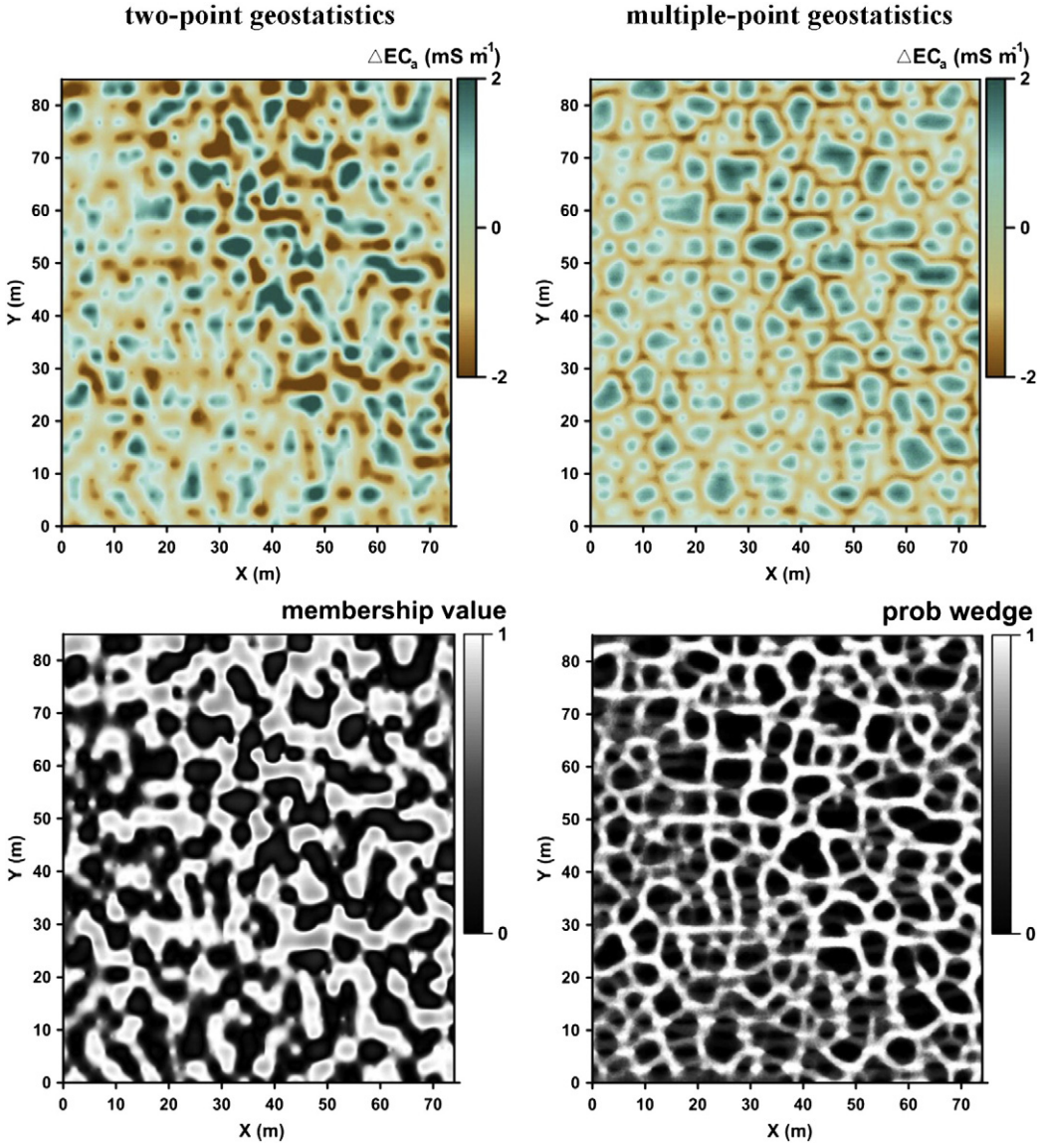

Fig. 5. Map of the sensor data interpolated with ordinary kriging and derived fuzzy membership value map indicating the possibility to find wedge material in the subsoil (left) and map of the sensor data and probability map to find wedge material in the subsoil simultaneously generated with MPS (E-type of 50 simulations) (right).

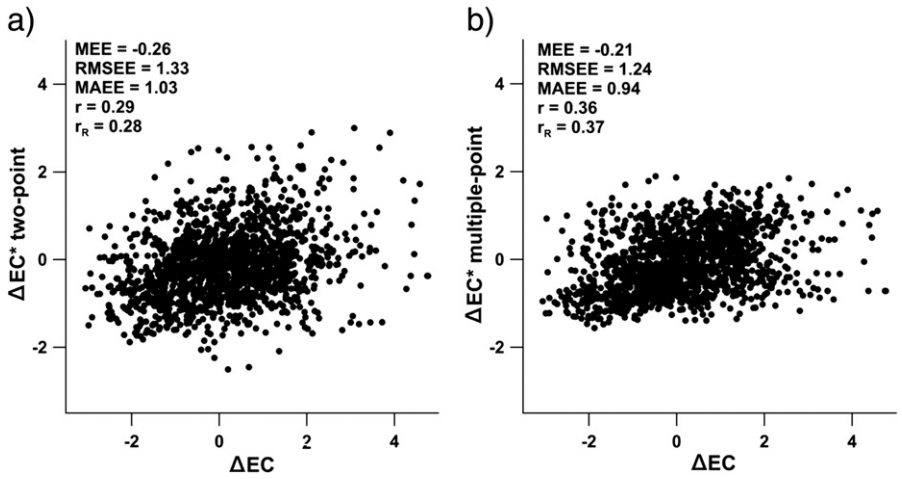

Fig. 6. Validation results for both interpolated sensor data maps (Fig. 5 — top) using the independent validation data of 9 measurement lines (Fig. 2b): scatterplots and validation indices (MEE = mean estimation error, RMSEE, root mean square estimation error, MAEE = mean absolute estimation error, r = Pearson's correlation coefficient, r_R = Spearman's rank correlation coefficient) for the map interpolated with two-point geostatistics (left) and the map interpolated with multiple-point geostatistics (right).

distance of the input data which was large in relation to the scale of the soil features (Fig. 2a).

However, the polygonal pattern was much better reconstructed in the MPS maps. The maps based on two-point geostatistics showed more smoothed polygons and a lack of connectivity for the smaller polygons. This better pattern reconstruction of the MPS maps is due to the use of a TI as a structural model which explicitly implies a multiple-point pattern. Since the used TI strongly influences the resulting map, selecting an appropriate TI is crucial. Nevertheless, one should realize that using a variogram as a structural model also has consequences for the higher order statistics. By only implying a two-point statistical model, the higher order statistics remain uncontrolled and are thus blindly accepted (Journel and Zhang, 2006).

In addition to reconstructing the patterns correctly, the prediction maps also need to be locally accurate. To quantify this local accuracy, we validated the maps as described in Section 2.4. Fig. 6 shows the validation results for the interpolated sensor data maps (Fig. 5 — top). Although the validation scatterplots show a smoothing effect for both maps (the slope shown by the data in both plots is less than one), they predicted the sensor data reasonably well. The scatterplot cloud was more elongated for the MPS E-type map, the correlation coefficients were closer to 1, and the validation indices closer to 0. This shows that the enhanced pattern reconstruction obtained with MPS does not come at the cost of local accuracy.

Fig. 7 shows the ROC curves for the maps predicting the presence of wedge material in the subsoil (Fig. 5 — bottom) using the 94 classified bore hole samples (Fig. 2b). The fuzzy membership value map had an AUC of 0.73 and the probability map created with MPS an AUC of 0.84. This means that the probability of ranking a randomly chosen location with wedge material higher than a randomly chosen location with host material, is higher for the MPS map than for the fuzzy membership value map. Hence, the MPS map was better able to locate the fossil ice-wedge polygons in the subsoil.

This case study illustrates one potential application of multivariate MPS in soil science, but the flexibility of the method opens up a wide range of potential applications. The variables to be simulated can be categorical and/or continuous and for each variable conditioning data can be given as input data. Furthermore, the (multivariate) TI can be data driven, knowledge driven or a combination of both, like the TI used in this paper.

The use of a TI is the strength of MPS, as it allows to simulate complex spatial structures and multivariate relationships based on different types of prior information. At the same time this TI is also the bottleneck of MPS: constructing a TI, especially a bivariate TI, can require a large effort and there is a need for prior information that can be visualized. In this study, we started the TI construction from a simple binary image taken from literature. However, comparing different methods to construct soil TIs is a topic for further research.

4. Conclusions

As a first application in soil science, this case study has shown that bivariate MPS can be used for the processing of proximal soil sensor data. Based on a bivariate TI, we interpolated the proximal soil sensor data and simultaneously predicted a target variable, i.e. the location of fossil ice-wedge polygons in the subsoil. The use of the sensor data as ancillary variable guaranteed local accuracy, while the multiple-point structural model (TI) ensured pattern reconstruction.

Since soil scientists often face complex patterns that are hard to model with traditional two-point geostatistics, we believe that (multivariate) MPS can be a valuable part of the pedometrical toolbox. It is an innovative and flexible hybrid approach, which can be both data and knowledge driven.

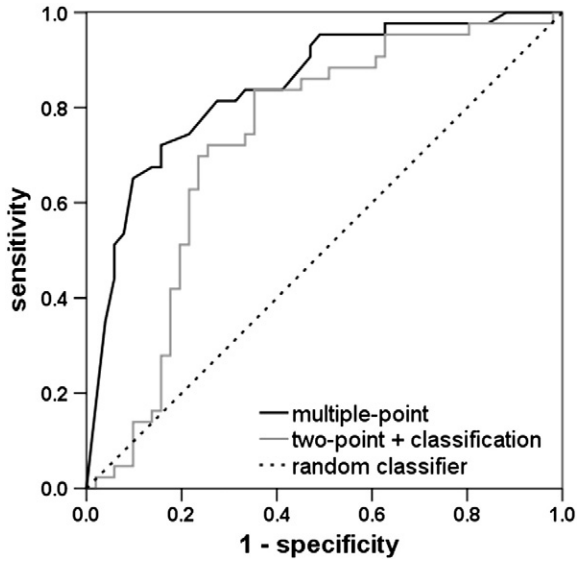

Fig. 7. Validation results for both maps predicting the presence of wedge material in the subsoil (Fig. 5 — bottom) using the 94 classified bore hole samples (Fig. 2b): receiver-operating characteristic (ROC) curves.

Acknowledgments

This research was supported by the Fund for Scientific Research-Flanders (FWO-Vlaanderen). The authors thank the farmer for granting access to his field and Valentijn Van Parys for his help with the field work.

References

Adamchuk, V.I., Hummel, J.W., Morgan, M.T., Upadhyaya, S.K., 2004. On-the-go soil sensors for precision agriculture. Computers and Electronics in Agriculture 44, 71–91.
Baer, J.U., Kent, T.F., Anderson, S.H., 2009. Image analysis and fractal geometry to characterize soil desiccation cracks. Geoderma 154, 153–163.
Cockx, L., Ghysels, G., Van Meirvenne, M., Heyse, I., 2006. Prospecting ice-wedge pseudomorphs and their polygonal network using the electromagnetic induction sensor EM38DD. Permafrost and Periglacial Processes 17, 163–168.
Cockx, L., Van Meirvenne, M., De Vos, B., 2007. Using the EM38DD soil sensor to delineate clay lenses in a sandy forest soil. Soil Science Society of America Journal 71, 1314–1322.
Cockx, L., Van Meirvenne, M., Vitharana, U.W.A., Verbeke, L.P.C., Simpson, D., Saey, T., Van Coillie, F.A.B., 2009. Extracting topsoil information from EM38DD sensor data using a neural network approach. Soil Science Society of America Journal 73, 2051–2058.
Comunian, A., Renard, P., Straubhaar, J., Bayer, P., 2011. Three-dimensional high resolution fluvio-glacial aquifer analog — part 2: geostatistical modeling. Journal of Hydrology 405, 10–23.
Corwin, D.L., Lesch, S.M., 2005. Characterizing soil spatial variability with apparent soil electrical conductivity. I. Survey protocols. Computers and Electronics in Agriculture 46, 103–133.
de Gruijter, J.J., McBratney, A.B., 1988. A modified fuzzy *k*-means method for predictive classification. In: Bock, H.H. (Ed.), Classification and Related Methods of Data Analysis. Elsevier, North Holland, pp. 97–104.
de Gruijter, J.J., McBratney, A.B., Taylor, J., 2010. Sampling for high-resolution soil mapping Chapter 1 In: Rossel, V.R.A., McBratney, A.B., Minasny, B. (Eds.), Proximal Soil Sensing. Progress in Soil Science, 1. Springer Science+Business Media B.V., pp. 3–14.
De Smedt, P., Van Meirvenne, M., Meerschman, E., Saey, T., Bats, M., Court-Picon, M., De Reu, J., Zwertvaegher, A., Antrop, M., Bourgeois, J., De Maeyer, P., Finke, P.A., Verniers, J., Crombe, P., 2011. Reconstructing palaeochannel morphology with a mobile multicoil electromagnetic induction sensor. Geomorphology 130, 136–141.
French, H.M., 2007. The Periglacial Environment, Third ed. John Wiley & Sons, Chichester.
Goovaerts, P., 1997. Geostatistics for Natural Resource Evaluation. Oxford University Press, New York.
Guardiano, F., Srivastava, R.M., 1993. Multivariate geostatistics: beyond bivariate moments. In: Soares, A. (Ed.), Geostatistics-Troia, Vol. 1. Kluwer, Dordrecht, pp. 133–144.
Huysmans, M., Dassargues, A., 2009. Application of multiple-point geostatistics on modelling groundwater flow and transport in a cross-bedded aquifer (Belgium). Hydrogeology Journal 17, 1901–1911.
Islam, M.M., Saey, T., Meerschman, E., De Smedt, P., Meeuws, F., Van De Vijver, E., Van Meirvenne, M., 2011. Delineating water management zones in a paddy rice field using a floating soil sensing system. Agricultural Water Management 102, 8–12.
Journel, A., Zhang, T., 2006. The necessity of a multiple-point prior model. Mathematical Geology 38, 591–610.
Kolstrup, E., 1986. Reappraisal of the upper Weichselian periglacial environment form Danish frost wedge casts. Palaeogeography, Palaeoclimatology, Palaeoecology 56, 237–249.
Le Coz, M., Genthon, P., Adler, P.M., 2011. Multiple-point statistics for modeling facies heterogeneities in a porous medium: the Komadugu-Yobe Alluvium, Lake Chad Basin. Mathematical Geosciences 43, 861–878.
Mariethoz, G., Renard, P., Straubhaar, J., 2010. The Direct Sampling method to perform multiple-point geostatistical simulations. Water Resources Research 46 (W11536).
McBratney, A.B., Moore, A.W., 1985. Application of fuzzy sets to climatic classification. Agricultural and Forest Meteorology 35, 165–185.
McBratney, A.B., Odeh, I.O.A., 1997. Applications of fuzzy sets in soil science: fuzzy logic, fuzzy measurements and fuzzy decisions. Geoderma 77, 85–113.
McBratney, A.B., Odeh, I.O.A., Bishop, T.F.A., Dunbar, M.S., Shatar, T.M., 2000. An overview of pedometric techniques for use in soil survey. Geoderma 97, 293–327.
McNeill, J.D., 1980. Electromagnetic terrain conductivity measurement at low induction numbers. Technical Note TN-6. Geonics Limited, Missisauga, Ontario, Canada.
Meerschman, E., Van Meirvenne, M., De Smedt, P., Saey, T., Islam, M.M., Meeuws, F., Van De Vijver, E., Ghysels, G., 2011. Imaging a polygonal network of ice-wedge casts with an electromagnetic induction sensor. Soil Science Society of America Journal 75, 2095–2100.
Meerschman, E., Pirot, G., Mariethoz, G., Straubhaar, J., Van Meirvenne, M., Renard, P., 2013. A practical guide to performing multiple-point statistical simulations with the Direct Sampling algorithm. Computers & Geosciences 52, 307–324.
Meerschman, E., Van Meirvenne, M., Van De Vijver, E., De Smedt, P., Islam, M.M., Saey, T., in press. Mapping complex soil patterns with multiple-point geostatistics. Eur. J. Soil. Sci.
Minasny, B., McBratney, A.B., 2002. FuzMe version 3.0. Australian Centre for Precision Agriculture, The University of Sidney, Australia.
Plug, L.J., Werner, B.T., 2002. Nonlinear dynamics of ice-wedge networks and resulting sensitivity to severe cooling events. Nature (London, U.K.) 417, 929–933.
Plug, L.J., Werner, B.T., 2008. Modelling of ice-wedge networks. Permafrost and Periglacial Processes 19, 63–69.
Pontius Jr., R.G., Schneider, L.C., 2001. Land-cover change model validation by an ROC method for the Ipswich watershed, Massachusetts, USA. Agriculture, Ecosystems and Environment 85, 239–248.
Ronayne, M., Gorelick, S., Caers, J., 2008. Identifying discrete geologic structures that produce anomalous hydraulic response: an inverse modeling approach. Water Resources Research 44 (W08426).
Saey, T., Simpson, D., Vitharana, U.W.A., Vermeersch, H., Vermang, J., Van Meirvenne, M., 2008. Reconstructing the paleotopography beneath the loess cover with the aid of an electromagnetic induction sensor. Catena 74, 58–64.
Saey, T., Simpson, D., Vermeersch, H., Cockx, L., Van Meirvenne, M., 2009. Comparing the EM38DD and DUALEM-21S sensors for depth-to-clay mapping. Soil Science Society of America Journal 73, 7–12.
Straubhaar, J., Renard, P., Mariethoz, G., Froidevaux, R., Besson, O., 2011. An improved parallel multiple-point algorithm using a list approach. Mathematical Geosciences 43, 305–328.
Strebelle, S., 2002. Conditional simulation of complex geological structures using multiple-point statistics. Mathematical Geology 34, 1–21.
Strebelle, S., Payrazyan, K., Caers, J., 2003. Modeling of a deepwater turbidite reservoir conditional to seismic data using principal component analysis and multiple-point geostatistics. Society of Petroleum Engineers Journal 8, 227–235.
Triantafilis, J., Odeh, I.O.A., McBratney, A.B., 2001. Five geostatistical models to predict soil salinity from electromagnetic induction data across irrigated cotton. Soil Science Society of America Journal 65, 869–878.
Vitharana, U.W.A., Saey, T., Cockx, L., Simpson, D., Vermeersch, H., Van Meirvenne, M., 2008a. Upgrading a 1/20,000 soil map with an apparent electrical conductivity survey. Geoderma 148, 107–112.
Vitharana, U.W.A., Van Meirvenne, M., Simpson, D., Cockx, L., De Baerdemaeker, J., 2008b. Key soil and topographic properties to delineate potential management classes for precision agriculture in the European loess area. Geoderma 143, 206–215.
Zhang, T., Bombarde, S., Strebelle, S., Oatney, E., 2006. 3D porosity modeling of a carbonate reservoir using continuous multiple-point statistics simulation. Society of Petroleum Engineers Journal 11, 375–379.