European Journal of Soil Science

European Journal of Soil Science, April 2013, **64**, 183–191 doi: 10.1111/ejss.12033

Mapping complex soil patterns with multiple-point geostatistics

E. Meerschman, M. Van Meirvenne, E. Van De Vijver, P. De Smedt, M. M. Islam & T. Saey
Department of Soil Management, Ghent University, Coupure 653, B-9000 Gent, Belgium

Summary

The commonly used variogram function is incapable of modelling complex spatial patterns associated with repetitive, connected or curvilinear features, because it is a two-point statistic. Because this was strongly limiting to petroleum- and hydrogeologists, they developed multiple-point geostatistics (MPG), an approach that replaces the variogram by a training image (TI). However, soil scientists also face complex spatial patterns and MPG might be of use to them as well. Therefore, this paper aims to introduce MPG to soil science and demonstrate its potential with a case study of polygonal subsoil patterns caused by Weichselian periglacial frost cracks in Belgium. A high-resolution proximal soil sensing survey provided a reference image from which a continuous (655 sensor data) and a categorical (100 point observations) dataset were extracted. As a continuous TI, we used the geophysical data of another part of the field, and as categorical TI we used a classified photograph of an ice-wedge network in Alaska. The resulting MPG maps reconstructed the polygonal patterns very well and corresponded closely to the reference image. Consequently, we identify MPG as a promising technique to map complex soil patterns and suggest that it should be added to the pedometrician's toolbox.

Introduction

Multiple-point geostatistics (MPG) is a recently developed toolbox of simulation algorithms replacing the traditional variogram function as a model of the spatial structure by a training image (TI). The experimental variogram $\gamma(\mathbf{h})$ is calculated as

$$\gamma(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} \{z(\mathbf{x}_\alpha + \mathbf{h}) - z(\mathbf{x}_\alpha)\}^2, \quad (1)$$

with $N(\mathbf{h})$ the number of data pairs $\{z(\mathbf{x}_\alpha), z(\mathbf{x}_\alpha + \mathbf{h})\}$ separated by the vector $\mathbf{h}$. A variogram is a two-point statistic: it relates the same variable Z at two locations. However, a two-point statistic is not capable of modelling complex spatial patterns such as repetitive, connected or curvilinear patterns. Reproducing such complex patterns requires us to consider more than two locations at a time, in other words multiple-point statistics (Goovaerts, 1997). Guardiano & Srivastava (1993) were the first to suggest the derivation of multiple-point statistics from a TI instead of computing them from observations as their number is mostly too limited.

A TI is a conceptual image of the expected spatial structure reflecting the spatial dependence between multiple points. It can be constructed from expert knowledge, a hand drawing, a model output or an existing map or photograph that is assumed to be an analogue of the phenomenon under study. Alternatively, a TI can be constructed from more densely sampled zones (Goovaerts, 1997). Prerequisites for a good TI are that it represents both the geometrical features and the global statistics (Hu & Chugunova, 2008). A TI does not need to carry any local information of the studied phenomenon; it only needs to reflect a prior structural concept (Strebelle *et al*., 2003).

After the basic idea of MPG had been proposed in the early nineties, the first MPG algorithm, *snesim*, was developed by Strebelle (2002). During the last decade, MPG has become an active research topic and several MPG algorithms have been developed (see later) (Hu & Chugunova, 2008). Most MPG applications can be found in petroleum- and hydro-geological studies (Strebelle *et al*., 2003; Zhang *et al*., 2006a; Ronayne *et al*., 2008; Huysmans & Dassargues, 2009; Comunian *et al*., 2011; Le Coz *et al*., 2011). However, complex spatial patterns that are hard to model with traditional two-point geostatistics also occur in soil science. Soil-forming factors, including the human factor, induce complex patterns in soils that can be observed at scales ranging from landscape to microscopic. Examples are dune patterns, palaeochannels, limestone pavement, desiccation cracks, (relict) patterned ground, land-use patterns, sedimentary rock layers and soil pores. As such, local estimation of these

Correspondence: E. Meerschman. E-mail: Eef.Meerschman@UGent.be
Received 8 May 2012; revised version accepted 23 January 2013

soil features or simulating them for probabilistic forecasting (for infiltration rates for example) is an inherent part of soil science.

This paper aims to introduce MPG to soil science. Therefore, we first summarize the MPG theory and then provide a case study based on real data to demonstrate the applicability of MPG in soil science. A part (30 m × 30 m) of a geophysical image of polygonal textural patterns was used as the exhaustively-known reference image. From this reference image we extracted a continuous and a categorical dataset and evaluated the continuous and categorical MPG reconstruction of the complex soil patterns.

Theory

Two-point geostatistics

The two-point geostatistical (TPG) toolbox consists of estimation and simulation techniques. A characteristic of TPG estimation, known as kriging, is that each estimate $Z^*(\mathbf{x})$ is obtained independently of any neighbouring estimate. Kriging provides the best local estimate for each $\mathbf{x}$ in the least-squares sense by minimizing the variance of the estimation error. However, kriging maps are smoothed in their spatial variation and do not represent the true spatial variation properly (Goovaerts, 1997). Consequently, the local accuracy of kriged maps comes at the cost of a poor reproduction of the spatial pattern.

The TPG simulation algorithms, on the other hand, aim at a more realistic reproduction of the spatial pattern. They generate multiple realizations that can be true to the conditioning data and reproduce the sample histogram and variogram. Most TPG simulation algorithms are sequential simulation algorithms, such as sequential Gaussian simulation and sequential indicator simulation.

The key to sequential simulation algorithms is building a conditional probability distribution function (cpdf) for each successively visited location to be simulated $\mathbf{x}$:

$$f(\mathbf{x}; z|(n)) = \text{Prob}\{Z(\mathbf{x}) = z|(n)\}, \quad (2)$$

where $|(n)$ expresses the conditioning to local information (the conditioning data and the previously simulated grid nodes; Goovaerts, 1997). From this cpdf a simulated value for $Z(\mathbf{x})$ is drawn and the algorithm proceeds to the next $\mathbf{x}$. The TPG simulation algorithms build these cpdfs by considering the n neighbouring data one by one: for each location a kriging system is solved by using two-point variogram values.

Multiple-point geostatistics

To date, the MPG toolbox is restricted to simulation algorithms. Comparable to TPG realizations in which the sample histogram (a one-point statistic) and variogram (a two-point statistic) are reproduced, MPG simulation algorithms aim to generate realizations that honour the conditioning data and the multiple-point statistics of the TI. In other words, MPG simulation algorithms anchor the multiple-point patterns of the TI to the conditioning data (Caers & Zhang, 2004). A few alternative approaches to MPG are based on pattern classification (Zhang *et al*., 2006b; Arpat & Caers, 2007), but most MPG algorithms are sequential simulation algorithms (Hu & Chugunova, 2008).

In contrast to TPG, the cpdfs $f(\mathbf{x}; z|(n))$ in Equation (2) are built by considering the n neighbouring data jointly. For each successively visited $\mathbf{x}$, a data event $\mathbf{d}_n(\mathbf{x})$ of size n centred at location $\mathbf{x}$ is defined. This data event consists of the n neighbouring data values $z(\mathbf{x}+\mathbf{h}_\alpha)$ and the neighbouring data geometry, defined by the n vectors $\mathbf{h}_\alpha$ $(\alpha = 1, \ldots, n)$. Then, the TI is scanned for replicates of $\mathbf{d}_n(\mathbf{x})$. The TI scan is based on the principle that the conditional probability $f(\mathbf{x}; z|(n))$ corresponds to the ratio of the number of replicates with their central node equal to z, and the total number of replicates found (Strebelle, 2002).

Guardiano & Srivastava (1993) proposed rescanning the entire TI for each $\mathbf{x}$ to estimate $f(\mathbf{x}; z|(n))$. Unfortunately, this strategy was not practical to implement because it required a long computing time. The *snesim* code offered a practical solution: prior to sequential simulation, *snesim* scans the entire TI and stores all possible TI replicates in a search tree (Strebelle, 2002). For each $\mathbf{x}$, $f(\mathbf{x}; z|(n))$ is calculated from the tree structure. To date, *snesim* is still a popular MPG algorithm that is implemented in the freely distributed SGeMS software (Remy *et al*., 2009). An alternative is the *impala* code from the commercial Isatis software (Geovariances, 2011) that stores the catalogue of possible TI replicates in lists instead of tree structures (Straubhaar *et al*., 2011). Because all possible TI replicates are stored beforehand in a catalogue, memory use restricts both *snesim* and *impala* to the simulation of categorical variables. The first MPG sequential simulation technique that allows simulation of different variable types, including categorical, continuous and multivariate variables, is the direct sampling (DS) code (Mariethoz *et al*., 2010). The DS re-scans the TI for each $\mathbf{x}$ during sequential simulation, as was proposed first by Guardiano & Srivastava (1993), but it samples the TI directly without explicitly modelling $f(\mathbf{x}; z|(n))$. As yet, DS has not been implemented in a software package, but the code is available for academic purposes. Because DS is used for the case study presented in this paper, we discuss its implementation in more detail.

Figure 1 shows a flowchart of the DS algorithm. If there are conditioning data available, these are first assigned to their closest grid nodes in the simulation grid. Conditioning data are generally point observations that can be either categorical or continuous, but they can also be transect or (quasi) exhaustive samples. When no conditioning data are available, DS will generate unconditional simulations. Then, a path is defined through the remaining grid nodes with unknown values $\mathbf{x}$. This path is usually random, but the user has the option to define a unilateral path.

For each sequentially visited $\mathbf{x}$, DS finds the n closest neighbours (including conditioning data and previously simulated grid nodes) in order to define the data event $\mathbf{d}_n(\mathbf{x})$. The user defines the maximum number of neighbours (n_{max}) and the maximum

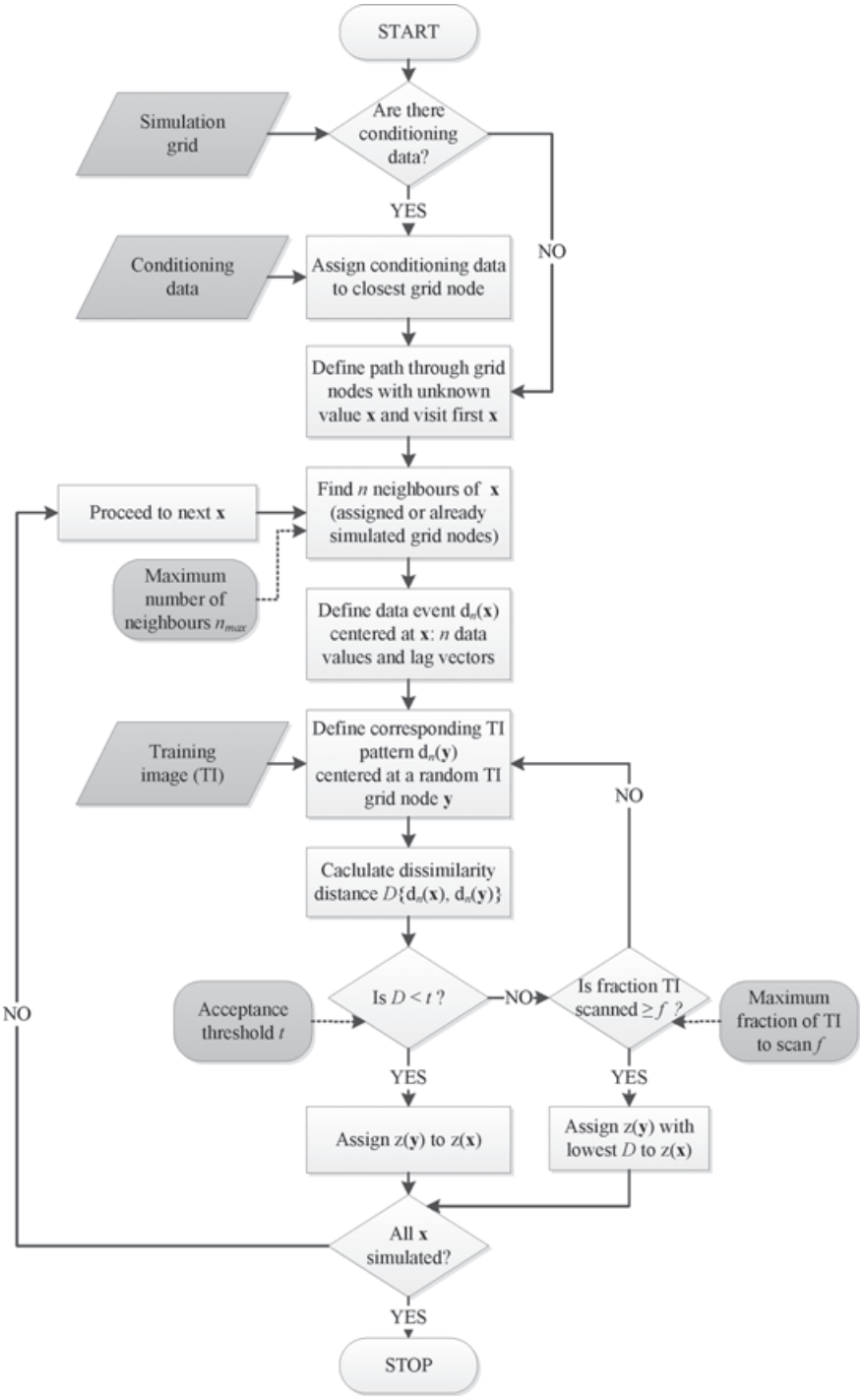

Figure 1 Flowchart of the multiple-point geostatistics algorithm used for this case study: the Direct Sampling algorithm.

search area (a rectangle). It is interesting to define a search area that equals the size of the simulation grid. When n_{max} is the only limiting factor, $\mathbf{d}_n(\mathbf{x})$ gradually decreases when the number of already simulated grid nodes increases, thus capturing patterns at different scales (Mariethoz *et al*., 2010).

Next, a TI scan is performed: for a random location $\mathbf{y}$ in the TI grid, the TI pattern $\mathbf{d}_n(\mathbf{y})$ is defined that has the same data geometry as $\mathbf{d}_n(\mathbf{x})$. The dissimilarity between the data values of $\mathbf{d}_n(\mathbf{x})$ and $\mathbf{d}_n(\mathbf{y})$ is quantified by a distance measure $D\{\mathbf{d}_n(\mathbf{x}), \mathbf{d}_n(\mathbf{y})\}$. As soon as D is smaller than a user-defined acceptance threshold t, the value at the central node of this TI pattern $z(\mathbf{y})$ is assigned to $z(\mathbf{x})$ in the simulation grid. If D is larger than t, the TI scan continues. For each variable type, one has only to select the appropriate dissimilarity distance D, making DS a flexible technique. The default distance type for categorical variables is the fraction of non-matching nodes between $\mathbf{d}_n(\mathbf{y})$ and

$\mathbf{d}_n(\mathbf{x})$. For continuous variables, it is the sum of the absolute value of the differences between the corresponding data values in $\mathbf{d}_n(\mathbf{y})$ and $\mathbf{d}_n(\mathbf{x})$. The latter is normalized, so both dissimilarity distances range between 0 and 1 (Mariethoz *et al*., 2010).

The user can determine the maximum fraction of the TI that is scanned for each $\mathbf{x}$ by setting parameter f to range between 0 and 1. If this maximum fraction has been scanned and still no TI pattern with $D < t$ has been found, DS assigns the central node of the TI pattern with the smallest D to $z(\mathbf{x})$. After all $\mathbf{x}$ values have been simulated, it is possible to resimulate the locations (except for the conditioning data locations) for which no TI pattern with $D < t$ had been found, with an entirely informed neighbourhood (not shown in the flowchart). Whether or not this post-processing is performed should be decided by the user, but it has been demonstrated that this additional step is beneficial because it reduces the level of noise (the isolated pixels that were wrongly simulated; Mariethoz *et al*., 2010; Meerschman *et al*., 2013).

Dataset of polygonal textural patterns

The geophysical dataset that we used was discussed in detail by Meerschman *et al*. (2011); therefore we limit ourselves here to a brief description. We surveyed a 0.6-ha field in the sandy silt region of Flanders, Belgium. This field was selected because an aerial photograph showed polygonal crop marks indicating textural patterns in the soil (Figure 2a). These textural patterns were caused by thermal contraction cracking in periglacial areas during the Weichselian glacial period and are known as fossil ice-wedge polygons or pseudomorphs (Kolstrup, 1986). At the end of the glaciation the soil cracks were filled up and covered with sediments (French, 2007). Mapping these cryogenic features is of interest because the abrupt changes in the subsoil composition affect soil management. Further, the morphology of their polygonal network is important for palaeo reconstructions (Plug & Werner, 2002). Note that a former field track crossed the field from north to southeast (Figure 2a).

To observe the polygonal textural patterns directly, we excavated a part of the field (approximately 6 m by 6 m) to a depth of 90 cm and uncovered an ice-wedge pseudomorph with a diameter of about 6 m. The host material had a sandy clay texture, whereas the younger wedge infilling contained a larger percentage of sand.

The field was surveyed with a mobile electromagnetic induction soil sensor (EM38-DD, Geonics, Mississauga, Canada) that measures the apparent electrical conductivity (EC_a) of an underlying soil volume down to approximately 1.5 m (Cockx *et al*., 2006). The inter-line distance was 0.75 m and the within-line distance between sensor response registrations was 0.15 m. Because the EC_a data were generally larger at the former field track, we subtracted a moving spatial average (radius = 3 m) from each measurement to highlight the polygon boundaries. The resulting residuals were expressed as $\Delta EC_a = EC_a$ - trend. Negative ΔEC_a values corresponded to the sand-rich wedge infilling and positive ΔEC_a values to the host material. The ΔEC_a data were interpolated by ordinary kriging (OK) to a regular grid with a resolution of 0.1 m by 0.1 m (Figure 2b). Despite the connectivity of the low values, estimation with OK produced good results because the sampling density was sufficient to reveal the complex spatial patterns.

Results

Continuous data

Continuous reference image and conditioning data. The lower left part of the ΔEC_a map (30 m × 30 m) (Figure 2b) was used as the continuous reference image (Figure 3, top right). Ten measurement lines within this area, having an inter-line distance of 3 m and a within-line distance of 0.4 m, were used as conditioning input data (655 data points) (Figure 3, top left).

Variogram modelling and mapping with traditional two-point geostatistics. Before reconstructing the image with MPG, we first applied a standard TPG approach of fitting a variogram model to the experimental variogram and used this model to generate an estimation and simulation map. The experimental variogram of the continuous data is given in Figure 3 (centre left). It shows no nugget effect and an almost linear increase to a sill that displays a hole effect, indicating a fairly regular repetition in the

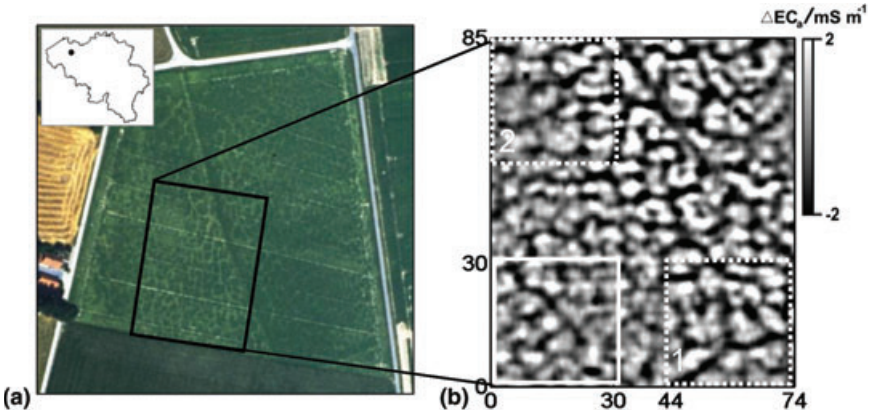

Figure 2 (a) Aerial photograph showing polygonal crop marks and a former field track (north-southeast oriented) with delineation of the surveyed area (black rectangle) (© J. Bourgeois, Department of Archaeology, Ghent University, Belgium: photo, J. Semey) and (b) resulting proximal soil sensor image (ΔEC_a/mS m^{-1}) (Meerschman *et al*., 2011): the left bottom part of this image was used as continuous reference image (white rectangle) and the right bottom and left upper part as continuous training images (white dashed rectangles).

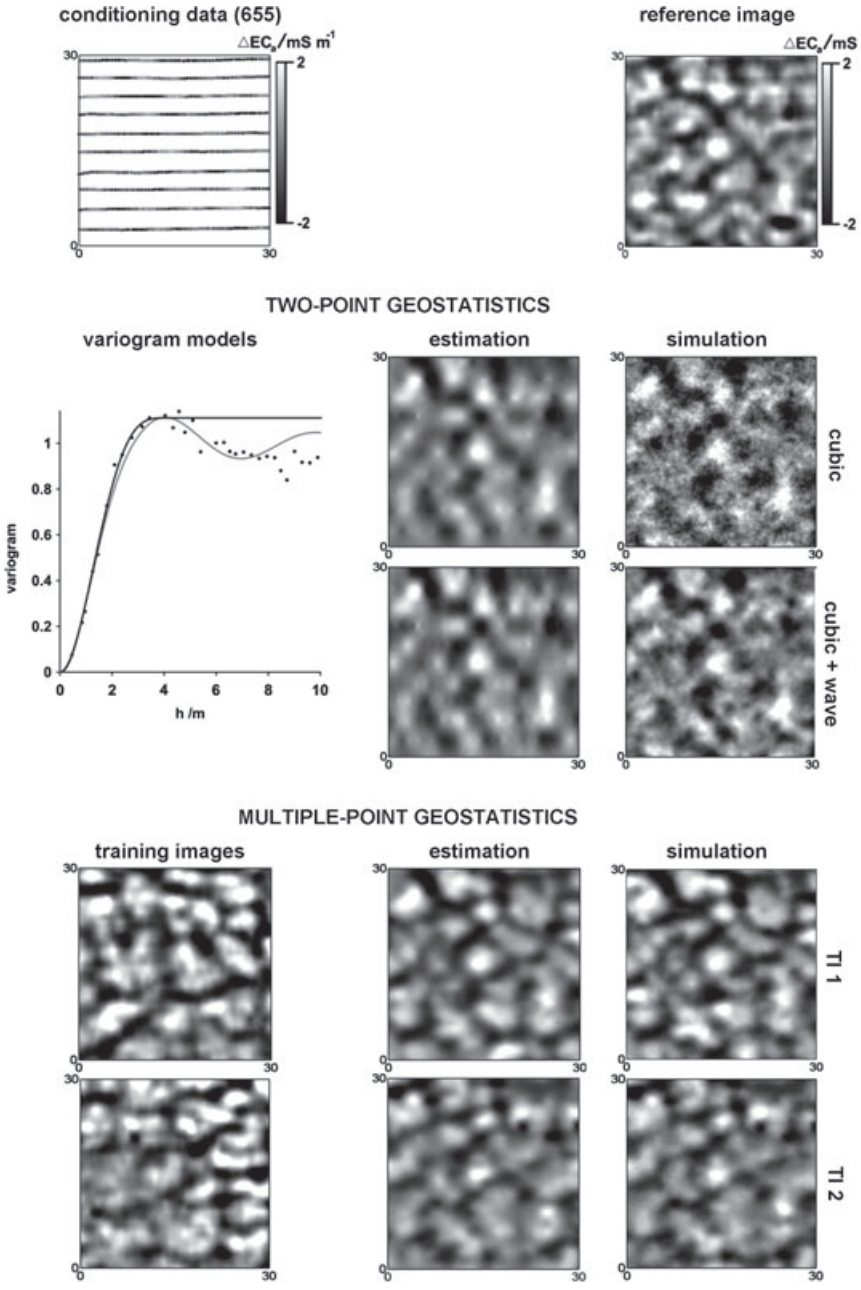

Figure 3 Reconstruction of the continuous reference image with two-point and multiple-point geostatistics starting from 655 conditioning data. The two-point estimation (ordinary kriging) and simulation (first Gaussian simulation realization) maps were based on a non-periodic (cubic) and a periodic (cubic + cardinal sine) variogram model. The two multiple-point estimation (E-type of 100 direct sampling (DS) realisations) and simulation (first DS realization) maps were based on two different continuous training images.

process (Webster & Oliver, 2007). To evaluate the contribution of the hole effect, we fitted both a non-periodic and a periodic variogram model. The non-periodic variogram model was a cubic function:

$$\gamma(h) = \begin{cases} c\left[7\left(\frac{h}{a}\right)^2 - 8.75\left(\frac{h}{a}\right)^3 + 3.5\left(\frac{h}{a}\right)^5 - 0.75\left(\frac{h}{a}\right)^7\right] & \text{for } h \le a, \\ c & \text{for } h > a, \end{cases} \quad (3)$$

with $c = 1.11$ m and $a = 4.1$ m. Similar to a Gaussian function, a cubic function is a bounded model with reverse curvature near the origin (Webster & Oliver, 2007). We incorporated periodicity in the variogram model by fitting a combination of a cubic and cardinal sine variogram:

$$\gamma(h) = \begin{cases} c_1\left[7\left(\frac{h}{a_1}\right)^2 - 8.75\left(\frac{h}{a_1}\right)^3 + 3.5\left(\frac{h}{a_1}\right)^5 - 0.75\left(\frac{h}{a_1}\right)^7\right] \\ \quad + c_2\left[1 - \frac{\sin\left(\frac{h}{a_2}\right)}{\left(\frac{h}{a_2}\right)}\right] & \text{for } h \le a_1, \\ c_1 + c_2\left[1 - \frac{\sin\left(\frac{h}{a_2}\right)}{\left(\frac{h}{a_2}\right)}\right] & \text{for } h > a_1, \end{cases} \quad (4)$$

with $c_1 = 0.48$, $a_1 = 3.2\,\text{m}$, $c_2 = 0.52$ and $a_2 = 0.9\,\text{m}$. The cardinal sine function is a simple periodic function that is valid in one, two and three dimensions (Webster & Oliver, 2007).

The estimation maps were created by ordinary kriging (OK) and the simulation maps by sequential Gaussian simulation (SGS) (Goovaerts, 1997) (Figure 3, centre right). The ΔEC_a data distribution followed a standard normal distribution, making a normal score transformation unnecessary. Both the variogram modelling and the two-point geostatistical mapping were performed with the Isatis software (Geovariances, Avon Cedex, France). The search area was set to the size of the study area ($30\,\text{m} \times 30\,\text{m}$) and the maximum number of neighbours to 50.

TI construction and mapping with multiple-point geostatistics. Whereas 655 observations are generally considered as sufficient to infer two-point statistics, this number is not enough to infer multiple-point statistics. Therefore, we applied an alternative strategy and used another part of the ΔEC_a map (Figure 2b) as a continuous TI (TI 1) (Figure 3, bottom left). This mimics a situation in which a field is partially sampled at a very high resolution to infer multiple-point statistics and partially at a lower resolution (larger inter-line distance) to minimize cost. To get an idea about how sensitive the MPG maps are to the TI, we repeated the reconstruction using another part of the field as TI (TI 2) (Figure 2b).

The estimation maps were obtained as the E-type (conditional mean) of 100 DS realizations and the simulation maps were obtained as the first DS realizations (Figure 3, bottom right). We used the default distance type for continuous data and selected the post-processing option. Parameter t was set to 0.02, f to 0.75, n_{max} to 50 and the maximum search area equal to the size of the study area ($30\,\text{m} \times 30\,\text{m}$).

Evaluation of the two-point and multiple-point maps. Both the TPG and MPG estimation maps corresponded reasonably well to the reference image (Figure 3). The mean absolute estimation error (MAEE) was 0.51 for the TPG map based on the cubic variogram, 0.52 for that based on the periodic variogram, 0.64 for the MPG map with TI 1 and 0.54 for the MPG map with TI 2. However, pattern reconstruction was better for the MPG maps: the connectivity of the smaller values was better reproduced.

The TPG estimation maps were very similar ($r = 0.99$) and thus rather insensitive to the hole effect of the variogram model, whereas the two MPG estimation maps differed more ($r = 0.77$). This demonstrates that changing the TI has larger consequences than changing the variogram: a TI has greater control over the spatial structure (Boisvert *et al*., 2007). To interpret this correctly, the fundamental distinction between two-point and multiple-point techniques should be understood. For MPG simulations, the user provides a prior multiple-point structural model, the TI. This model allows us to link the n neighbouring data jointly to $Z^*(\mathbf{x})$. For TPG simulations, the user provides only a prior one-point (the histogram) and two-point structural model (the variogram), linking the n neighbouring data pairwise to $Z^*(\mathbf{x})$. Beyond the histogram and variogram, TPG algorithms use their own intrinsic prior structural model that is beyond the control of the user. For SGS this model is multivariate Gaussian, a model that imposes maximum entropy for the high-order statistics (Journel & Zhang, 2006).

The different multiple-point structural models are to some extent visualized in the simulation maps (Figure 3). The TPG simulation maps show a spatial distribution of greater entropy: the extremes are more spatially fragmented. The spatial patterns in the MPG simulation maps can be considered as better structured and correspond more closely to those in the reference image. This is of course due to the lower entropy (spatially connected small values) prior multiple-point models we defined by means of the TIs. Note that the conditioning data strongly guide the pattern reconstruction. The differences between the TPG and MPG simulation maps would have been more profound for a smaller number of conditioning data.

Categorical data

Categorical reference image and conditioning data. To obtain a categorical reference image, the continuous one was classified by a k-means classification after running a contrast enhancement filter (Figure 4, top right). We chose $k = 3$ to have a strict categorical dataset and not a binary one. The three classes represent wedge, host and intermediate materials. From this classified reference map 100 data points were extracted according to a stratified random sampling scheme, mimicking a soil sampling campaign where the subsoil textural class was estimated by field texturing (Figure 4, top left).

TI construction and estimation with multiple-point geostatistics. The experimental set-up of the categorical case study required an alternative approach to construct the TI. We assumed that the only information that we had about the spatial structure came from the small excavated area and that there was no exhaustively measured neighbouring field to derive a TI from directly, as was done in the continuous case. Hence, we chose an existing photograph from the literature and rescaled it using the information that we gathered from the excavation because it is beneficial when the size of the TI patterns corresponds more or less with the true pattern size.

We selected a near-infrared aerial photograph of a present-day ice-wedge network in Alaska (Plug & Werner, 2002), assuming that a similar genetic process was at the basis of both ice-wedge patterns. The photograph was rescaled, equalling its average polygon size to that of the textural polygon that was observed in the Belgian field by excavation (Figure 5a). Then, we applied a pixel-wise (7×7 kernel) adaptive Wiener low-pass filter for noise removal (Lim, 1990) and classified the image into three classes with a k-means classification using MATLAB R2011a (Mathworks, Natick, MA, USA) (Figure 5b). We chose

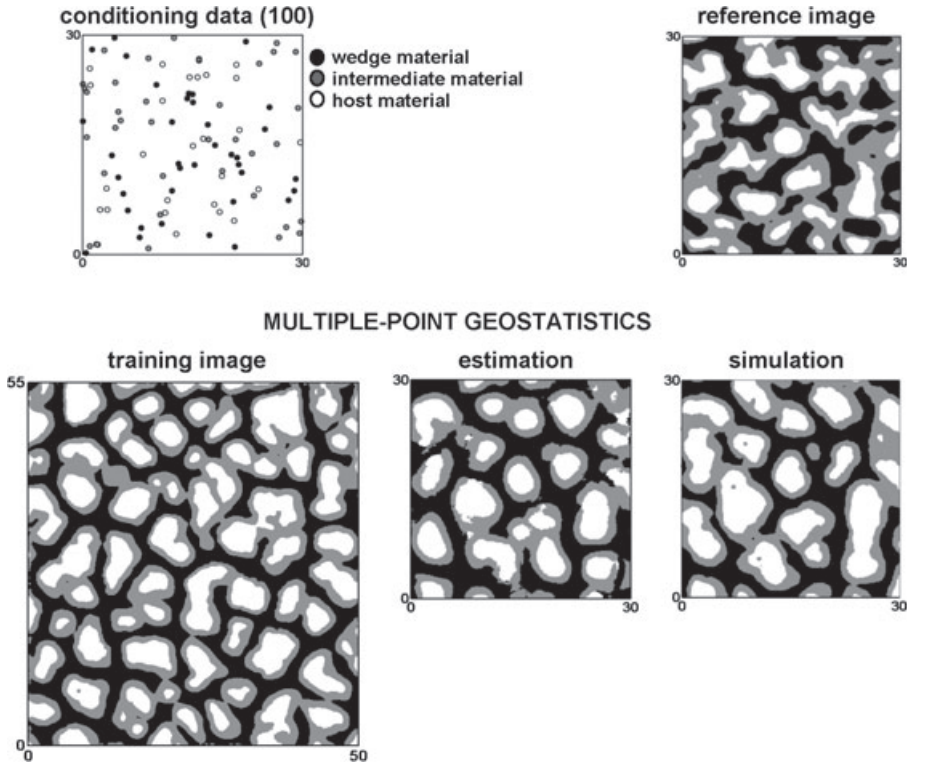

Figure 4 Reconstruction of the categorical reference image with multiple-point geostatistics starting from 100 conditioning data. The multiple-point estimation (conditional mode of 100 direct sampling (DS) realisations) and simulation (first DS realization) maps were based on a categorical training image.

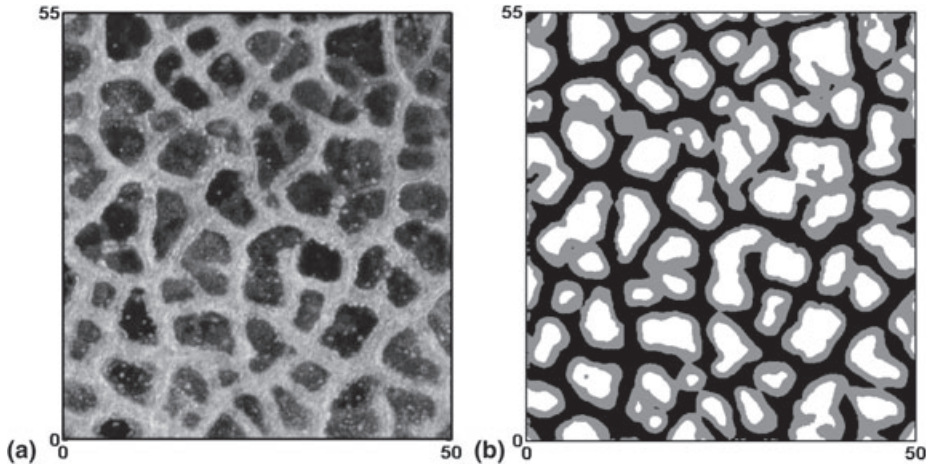

Figure 5 (a) The original, rescaled near-infrared aerial photograph of an ice-wedge network on the floor of a drained lake near Espenberg, northwest Alaska (Plug & Werner, 2002), used to construct (b) the categorical training image.

a photograph from Alaska, and not from Belgium, to illustrate better the strong concept of the TI as a database of representative patterns, independent of its origin.

The estimation map was the most probable category (conditional mode) of 100 DS realizations and the simulation map was the first DS realization. We used the default distance type and set t to 0.05, f to 0.5 and n_{max} to 50. As for the continuous case, the maximum search area was set to be equal to the size of the study area and after each simulation one post-processing step was performed for noise removal.

Evaluation of the multiple-point maps. Figure 4 compares the multiple-point categorical reconstruction with the categorical reference image. Most of the polygons were correctly identified. The categorical estimation map had a correct classification rate of 54%. The largest source of error between the estimation map and the reference image was due to the difference in spatial pattern between the chosen TI and the reference image. The TI contains polygons with smoother boundaries, and slightly over-estimates the connectivity of the wedge material and under-estimates the connectivity of the intermediate material.

Note that we did not make the comparison with TPG here because the number of observations was too small to reveal the spatial pattern, and the patterns themselves, especially the pattern of the intermediate material, were too complex to be modelled with a two-point variogram function.

Discussion and conclusions

Variogram-based geostatistics is mostly data driven (Heuvelink & Webster, 2001), whereas the construction of a multiple-point structural model usually requires extra information. Therefore, an often-mentioned constraint of MPG is that the construction of a TI requires a larger effort than variogram modelling.

Today there is, however, an increasing availability of soil covariates that help the soil scientist to model spatial variation accurately. This soil covariate information, like aerial photographs or proximal soil sensor images, might be used to construct TIs that can be more generally applied (because they do not need to contain any local information). As illustrated in the continuous case, additional high-density proximal soil sensor data were used. Using such high-resolution spatial datasets as a model of the spatial structure to interpolate less densely sampled data is an elegant approach.

Moreover, soil scientists often have some prior conceptual knowledge about the studied phenomenon. Although attempts have been made to translate this prior knowledge into a variogram function (Truong *et al*., 2012), it is easier to translate these concepts into an image (Journel & Zhang, 2006), as has been illustrated in the categorical case study.

A second perception is that TI construction is more subjective than variogram modelling. Indeed, constructing a TI forces the user to make decisions about the multiple-point statistics instead of just accepting them implicitly. However, we believe that the explicit visualization of multiple-point statistics in a TI should be seen as an advantage rather than inducing some degree of subjectivity. Multiple-point maps are very sensitive to the chosen TI, but it should be realised that this sensitivity is not stronger than the sensitivity to the combination of a variogram model and implicit high-order assumptions (Journel & Zhang, 2006).

Soil scientists frequently face periodic, connected or curvilinear patterns. We believe that MPG is a promising and accessible technique to model these complex spatial soil patterns. Consequently, MPG deserves to be added to the pedometrician's toolbox.

Acknowledgements

This work was funded by the Flemish Fund for Scientific Research (FWO-Vlaanderen).

References

Arpat, G.B. & Caers, J. 2007. Conditional simulation with patterns. *Mathematical Geology*, **39**, 177–203.

Boisvert, J.B., Pyrcz, M.J. & Deutsch, C.V. 2007. Multiple-point statistics for training image selection. *Natural Resources Research*, **16**, 313–321.

Caers, J. & Zhang, T. 2004. Multiple-point geostatistics: a quantitative vehicle for integrating geologic analogs into multiple reservoir models. *Integration of Outcrop & Modern Aanalog Data in Reservoir Models: AAPG Memoir*, **80**, 383–394.

Cockx, L., Ghysels, G., Van Meirvenne, M. & Heyse, I. 2006. Prospecting frost-wedge pseudomorphs and their polygonal network using the electromagnetic induction sensor EM38DD. *Permafrost & Periglacial Processes*, **17**, 163–168.

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.

French, H.M. 2007. The Periglacial Environment, 3rd edn. John Wiley and Sons, Chichester.

Goovaerts, P. 1997. Geostatistics for Natural Resources Evaluation. Oxford University Press, New York.

Guardiano, F.B. & Srivastava, R.M. 1993. Multivariate geostatistics: beyond bivariate moments. In: Geostatistics-Troia, Volume **1** (ed. A. Soares), pp. 133–144. Kluwer Academic Publishers, Dordrecht.

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

Hu, L.Y. & Chugunova, T. 2008. Multiple-point geostatistics for modeling subsurface heterogeneity: a comprehensive review. *Water Resources Research*, **44**, W11413.

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.

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 from 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.

Lim, J.S. 1990. Two-Dimensional Signal and Image Processing. Prentice Hall, Upper Saddle River, NJ.

Mariethoz, G., Renard, P. & Straubhaar, J. 2010. The direct sampling method to perform multiple-point geostatistical simulations. *Water Resources Research*, **46**, W11536.

Meerschman, E., Van Meirvenne, M., De Smedt, P., Saey, T., Islam, M.M., Meeuws, F., *et al*. 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.

Plug, L.J. & Werner, B.T. 2002. Nonlinear dynamics of ice-wedge networks and resulting sensitivity to severe cooling events. *Nature*, **417**, 929–933.

Remy, N., Boucher, A. & Wu, J. 2009. Applied Geostatistics with SGeMS: A User's Guide. Cambridge University Press, New York.

Ronayne, M.J., Gorelick, S.M. & Caers, J. 2008. Identifying discrete geologic structures that produce anomalous hydraulic response: an inverse modeling approach. *Water Resources Research*, **44**, W08426.

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 deep-water turbidite reservoir conditional to seismic data using principal component analysis and multiple-point geostatistics. *SPE Journal*, **8**, 227–235.

Truong, P.N., Heuvelink, G.B.M. & Gosling, J.P. 2012. Web-based tool for expert elicitation of the variogram. *Computers & Geosciences*, **51**, 390–399.

Webster, R. & Oliver, M.A. 2007. Geostatistics for Environmental Scientists, 2nd edn. John Wiley & Sons, Chichester.

Zhang, T., Bombarde, S., Strebelle, S. & Oatney, E. 2006a. 3D porosity modeling of a carbonate reservoir using continuous multiple-point statistics simulation. *SPE Journal*, **11**, 375–379.

Zhang, T., Switzer, P. & Journel, A. 2006b. Filter-based classification of training image patterns for spatial simulation. *Mathematical Geology*, **38**, 63–80.