

Stochastic modeling in geology: Determining the sufficient number of models

Noémi Jakab*

Department of Geology and Paleontology, University of Szeged, Szeged, Hungary

Received: January 9, 2017; accepted: March 24, 2017

Finding the optimal number of realizations to represent the model uncertainty when applying stochastic approaches is still a relevant question in geostatistics. The essence of the method is to visualize the realizations in a suitably constructed attribute space. To construct this space, the static connectivity metrics of the realizations were used. Within this framework, the creation of new realizations can be regarded as a sampling process, in which each new stochastic image is the equivalent of a new sampling point in the attribute space. The sampling process begins with the first few realizations appearing in a dispersed manner in random parts of the attribute space. The addition of more realizations causes the gradual emergence of higher point densities, which in the end, results in a point structure where most of the points are located in areas of high point densities with areas of low point densities surrounding them. High point densities represent typical realizations showing very similar connectivity characteristics, whereas low point densities correspond to atypical realizations with stronger deviations from the bulk. In this sense, reaching the optimal number of realizations is the equivalent of reaching a state in the sampling process where high- and low point densities are present at the same time, yet high point densities do not dominate the overall structure of the attribute space, as they also reflect the redundancy of the information content. This desired structure is strongly analogous to the complete spatial randomness of spatial point processes, where the points are neither dispersed nor aggregated in space. Based on this analogy, the normalized version of Ripley's K -function and the L -function for the spatial inhomogeneous Poisson point process was applied to find the optimal number of realizations. The method is illustrated on a computed tomography slice and on the real-life data of the Tisza-2 reservoir.

*Corresponding address: Department of Geology and Paleontology, University of Szeged, Egyetem u. 2, H-6722 Szeged, Hungary

E-mail: j.noemi@geo.u-szeged.hu

This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium for non-commercial purposes, provided the original author and source are credited.

Keywords: geostatistics, uncertainty assessment, sequential Gaussian simulation, connectivity metrics, inhomogeneous Poisson process

Introduction

Giving estimates of values between sample data locations is a common task in the field of geology. Practitioners may face it in petroleum geology when working on reservoir modeling problems when characterizing ore reserves or modeling hydro-geologic features. In dealing with this sort of problem, it is important to keep in mind that estimates are always uncertain, regardless of the chosen estimation method. The sources of uncertainty may be the randomness of the studied process, our limited understanding and measurement error among others (Caers 2011).

This means that it is essential to characterize the inherent uncertainties in addition to producing estimates. Stochastic approaches enable us to do so by generating several equally probable realizations (Goovaerts 2001). In this way, the model uncertainty can be described and quantified by characterizing the differences between the individual stochastic images.

Possible ways for characterizing the model uncertainty

To characterize the model uncertainty, the probability distribution of the studied attribute must be described along the nodes of the simulation grid based on the realizations. Several descriptive parameters can be used to highlight areas of low, medium, and high uncertainties, such as the conditional variance linked to the E-type estimates, the confidence interval (Goovaerts 2001; Mucsi et al. 2013), and statistical entropy (Journel and Deutsch 1993; Geiger and Újhelyi 2012). Based on these parameters, we can obtain a rough idea of the regional uncertainties, but only if the number of the produced realizations is sufficient to provide a reliable outline.

However, in practice, the number of realizations is often limited by the available computational capacity, since in the case of a high grid resolution, simulating even one realization can be very time-consuming. Thus it is reasonable to ask how reliable an idea about the uncertainty can be that was constructed based on 5 or 10 stochastic realizations, how much new information can be gained by producing an additional 50 or 100 realizations, or from what number of realizations onward does the information content of any newly generated realization become redundant. The most important question, summarizing all previous ones, however, is: how many realizations are enough? Goovaerts (1999) sought the answers to these questions using the convergence of the variances of the stochastic realizations. Following his train of thought, Geiger et al. (2010) used variance decomposition and the convergence of the within-group and between-group variances to address the same question.

The generated stochastic images are often post-processed, as for example in the case of ore-body evaluations, where the net present value is calculated. In petroleum geology applications, a dynamic flow simulation may be performed; however, only a fraction of the realizations can be processed in this way, due to the time required for the flow calculations. The selection of these representative realizations can be performed subjectively, or by ranking them based on any parameter correlated with the studied flow property. In this manner, stochastic images corresponding to the P10, P50, and P90 cutoff values can be selected and passed on to the flow simulator (Ballin et al. 1992). Another possible way to select representative realizations was introduced by Scheidt and Caers (2009), who suggest representing the realizations in a Euclidean attribute space. The essence of the method is that distances between realizations in this attribute space are correlated with the differences between their flow parameters. As a result, the realizations become separable using any suitably selected multivariate method, such as principle component analysis or cluster analysis. Then, for example, the cluster centers can be selected as representative realizations for post-processing.

The aim of this study is to introduce a novel approach in harmony with the previously mentioned methods to address the question of the optimal number of stochastic realizations.

Sequential Gaussian simulation (SGS)

SGS is a type of conditional stochastic simulation (Srivastava 1990), meaning it produces realizations honoring the input data values at their locations. It generates several equally probable alternative realizations of the studied spatial phenomenon, utilizing its modeled spatial continuity. The Gaussian approach assumes parametric distributions both in the case of the input data and the simulated values. The implementation of the SGS consists of the following steps (Gómez-Hernández and Srivastava 1990; Isaaks 1990; Deutsch and Journel 1998; Goovaerts 2001; Geiger 2006):

- (1) Normal score transform of the input data.
- (2) Define a simulation grid over the study area.
- (3) Define a random path through the grid nodes.
- (4) Construct the conditional distribution of the studied variable at the first grid node given the (n) conditioning data, then draw a random value from this distribution and assign it to the grid node.
- (5) Add this randomly assigned value to the conditioning dataset and use it when building the conditional distribution for the next grid node on the path according to step (4).
- (6) Repeat step (5) until a simulated value is generated for all grid nodes.
- (7) Back transform the simulated values.
- (8) Choosing different random paths and repeat steps (4)–(7) until the required number of realizations is produced.

Problem statement

Simulation techniques are usually considered as tools capable of creating a large number of equally probable alternatives which may contain an unmanageable amount of information if we disregard the problem of the required computational capacity. This impression is mostly inflicted by the multiplicity of random paths which can be chosen on the grid, and by the random selection used to assign the values to the grid nodes.

It is therefore crucial to consider how the effective difference, based on which two realizations may be deemed different at the decision-making level, can be defined between the several, pairwise different stochastic images.

Methods

Connectivity attributes

To define these differences, we must turn to the connectivity metrics of the simulated stochastic images. Connectivity metrics are widespread descriptive attributes used from the field of reservoir modeling (Deutsch 1998) through landscape ecology (McGarigal et al. 2012). They are used to characterize the heterogeneity of surfaces. Unlike spatial continuity, they are not included among the input parameters of the model-building phase; thus, they are able to highlight differences between realizations. A detailed overview of the several existing connectivity attributes can be found in the work of Renard and Allard (2013).

In this study, global and static metrics of connectivity were used. Based on the work of Renard and Allard (2013), static metrics are statistical measures describing the connectivity and thereby the flow characteristics of a field without considering a specific physical process. This means they characterize units within the studied field that show similar flow properties.

The global nature of these metrics, in turn, reflects that they provide information about the entirety of the model, and not just about a highlighted sub-unit, such as a selected well and its neighborhood. Such global metrics of connectivity are often referred to as geo-body or geo-object connectivity.

There are several software implementations available for performing the connectivity analysis by Deutsch (1998), Pardo-Igúzquiza and Dowd (2003), and McGarigal et al. (2012). These all perform the connectivity analysis following the two steps (Fig. 1):

- (1) Establish a binary indicator of net cells, separating cells that are considered net and non-net considering their flow properties.
- (2) Scan through the field aggregating corner- or edge-wise connected blocks of net cells, labeling each different connected unit.

Following the outlining of the geo-objects, their connectivity attributes can be calculated depending on the question at hand. The work by Deutsch (1998) only

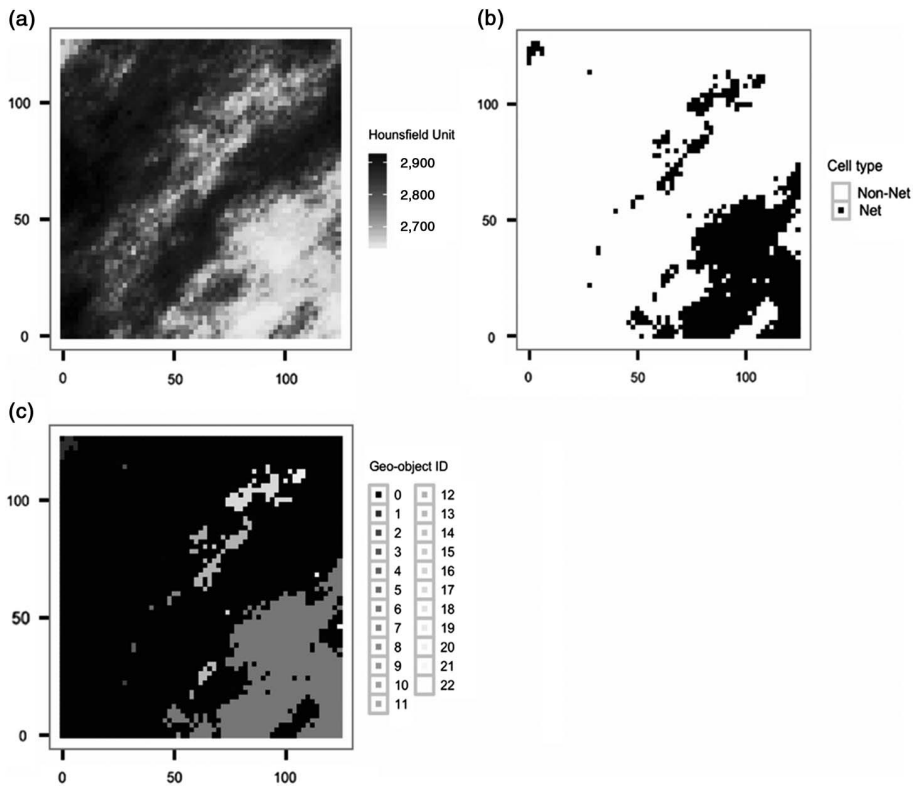


Fig. 1 Steps of the connectivity analysis: (a) a stochastic image, (b) its indicator transform, and (c) the identified geo-objects

utilizes the sizes of the different geo-objects, whereas in the paper by Pardo-Igúzquiza and Dowd (2003), various other metrics can be found, such as their average, minimal, and maximal areas and extents in X , Y , and Z dimensions. However, the most detailed characterization of the geo-objects can be found in the landscape ecology-themed work of McGarigal et al. (2012). In addition to the basic metrics, this also provides tools to characterize the geometric complexity of the geo-objects using metrics like the length of edges on the inner and outer boundaries of the geo-objects, the number of internal cells and the fractal dimension.

Sampling the field of uncertainty

Returning to the work of Scheidt and Caers (2009), which suggests representing the realizations in an n -dimensional attribute space, let us imagine there is a way to limit the location of the realizations in this space. This would mean that the realizations are

not spread in the infinite attribute space in an unbounded manner, but rather that they are confined to a precisely defined finite subspace. Such finite n -dimensional space can be created if the defining axes are selected to represent attributes, which can take their values only over closed intervals. As a result, the realizations can be enclosed in an n -dimensional box, its size defined by the theoretically possible values of the realizations on the attribute axes.

The resulting subspace can be regarded as an uncertainty field, which is sampled with the simulation process to obtain an understanding of the spread of the realizations in it. This means that each individual realization can be regarded as a sample in this space.

Following the introduced logic, the question asked at the beginning of this study can be rephrased in this way: How many realizations are required for the sampling of the uncertainty field to be considered sufficient? And the other aspect of the same question: After what number of realizations can it be stated that the uncertainty field is oversampled, meaning part of the information content of the realizations is redundant? To answer these questions, the spread of the points representing the realizations within the field of uncertainty must be studied.

One important characteristic of the simulation approach is that it only ensures reproduction of the input statistics, such as the estimated value and covariance, in the average sense, if a reasonably high number of realizations are generated. Thus, comparing the model statistics with those of the individual realizations, different levels of discrepancies will be found (Deutsch and Journel 1998). This is the phenomenon referred to in geostatistics as ergodic fluctuation. It also causes the variability of other derivative attributes, such as the connectivity parameters, which is thus observable in the field of uncertainty.

Consequently, if a high enough number of realizations is generated, the point cloud representing them in the uncertainty field will have different point densities. High point densities represent the bulk of the realizations, in the form of groups of typical stochastic images having similar connectivity attribute combinations with each other. Conversely, areas of low point densities suggest connectivity attribute combinations that are more uncommon, representing realizations characterized by higher ergodic fluctuations, deviating strongly from the bulk. In addition, there are also areas in the uncertainty field that are devoid of realizations, which may represent one of the three possible cases. First, these may represent the theoretically possible attribute combinations that are not depicted by any of the realizations. These may be extremities and their neighborhoods, such as realizations with all net or all non-net cells, which are impossible in practice. Second, realizations that are probable in practice but are not obtainable using the given model of spatial continuity and the selected simulation settings, which also belong to these areas. And lastly, these are also the places of realizations that are theoretically possible and with the chosen simulation settings obtainable, but not yet included in the generated n realizations.

This type of representation of the stochastic images can also facilitate assessment whether the number of realizations is sufficient to represent the model

uncertainty. If we imagine the simulation as a process where the points representing the realizations are placed in the uncertainty field one by one, the following can be seen: The first few points are randomly dispersed in the uncertainty field, without any structure suggesting where groups of typical realizations may form later. The addition of more realizations causes the gradual emergence of higher point densities, with areas of low point densities still present. After creating a few hundred realizations, this process results in a point structure where most of the points are located in areas of high point densities with areas of low point densities surrounding them. At this stage, any new realizations generated are most likely to appear in the areas of high point densities.

In this sense, reaching the optimal number of realizations is the equivalent of reaching a state in the sampling process where high- and low point densities are present at the same time, yet high point densities do not dominate the overall structure of the attribute space. The reason for this is that besides outlining groups of typical realizations, high point densities also reflect the redundancy of the information content, as they suggest similar attribute combinations.

Ripley's K-function

To find the number of realizations from which the sampling stops being sufficient and becomes redundant, it may help to consider the uncertainty field and the points in it representing the realizations as a spatial point process. A spatial point process is a stochastic process each of whose realizations consists of a finite or countably infinite set of points in the plane (Gelfand et al. 2010).

In the study of spatial point processes, the Poisson process is one of the most commonly used reference models, due to the fact that it embodies complete spatial randomness (Illian et al. 2008). Since in practice complete spatial randomness implies that the points are neither dispersed nor aggregated in space, it may be a suitable benchmark for the optimal number of realizations.

The Poisson point process can be characterized by its first moment, the intensity (λ), giving the mean number of points per unit area. For homogeneous Poisson processes the intensity is constant, whereas for the inhomogeneous case the intensity varies in space (Illian et al. 2008). Due to the uncertainty field having high- and low point densities, the inhomogeneous Poisson process is the most appropriate model for its characterization.

To check whether a point process is dominantly dispersed, aggregated, or spatially random, Ripley's K -function can be applied. The K -function is the second moment of the point process. It is the expected number of points N within a distance r of another point normalized by the intensity (Ripley 1977),

$$K(r) = \frac{1}{n} \sum_{i=1}^n \frac{N_{p_i}(r)}{\lambda}$$

where p_i is the i th point. The value of $K(r)$ for an arbitrary Poisson point process is $r^2\pi$. Consequently, to decide whether an empirical point process is aggregated, dispersed, or random, its K -function should be compared with the K -function of the theoretical Poisson process. Empirical $K(r)$ values below $r^2\pi$ suggest the dispersion of the point process, whereas higher empirical $K(r)$ values imply aggregation. Instead of the $K(r)$ function, often its normalized version, the L -function (Besag 1977) is used because it is linear with an expected value of r .

$$L(r) = \sqrt{\frac{K(r)}{\pi}}$$

While the $K(r)$ and $L(r)$ functions were originally defined for the homogeneous Poisson point processes, Baddeley et al. (2000) modified them for the inhomogeneous case.

Demonstrating the method on the example of a computed tomography (CT) slice

To demonstrate the approach, a CT slice of a core-size sedimentary structure was used. The image and the corresponding dataset consisted of 16,000 Hounsfield Unit values measured on a 125×128 regular grid. From this dataset, 100 data locations and the corresponding measurements were randomly chosen as the input data (Figs 2 and 3).

After the normal score transform of the input data, the spatial continuity was modeled using two empirical directional variograms and one omnidirectional empirical variogram. The variogram modeling was performed according to the Matheronian method (Journel and Huijbregts 1978), by averaging the squared differences of values at the ends of h vectors of different lengths:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(u_i) - z(u_i + h)]^2 \quad (1)$$

where $z(u_i)$ and $z(u_i + h)$ are the measured Hounsfield Unit values at the ends of vector h , and $N(h)$ is the number of data points separated by h distance. From the permissible models, an exponential model with the main continuity direction of 45° , an anisotropy ratio of 0.5, and a range of 52.23 units was fitted. The goodness of the fitted model was assessed based on the visual inspection of the curves and the variogram surface (Fig. 4).

The SGS was used to render 1,024 realizations on a 2×2 grid resolution. The realizations were then back-transformed and connectivity analysis was performed on them using the software package by McGarigal et al. (2012). For the connectivity analysis, the value of 2,700 Hounsfield Units was set as the indicator threshold, as values below this correspond to the highest porosities on the original CT slice

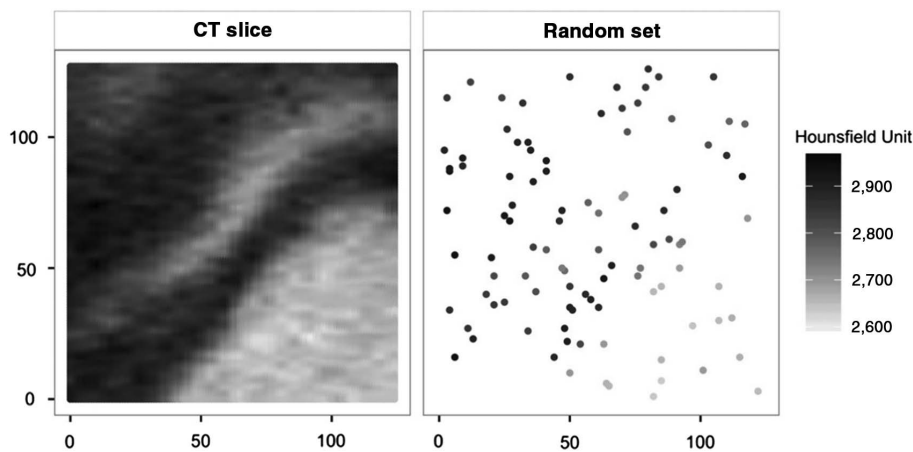


Fig. 2
The CT slice used as the exhaustive dataset and the random sample

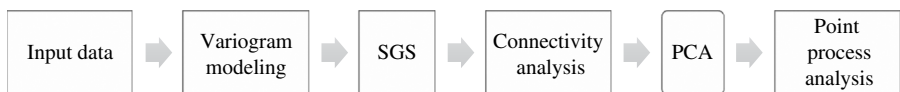


Fig. 3
Steps of the workflow

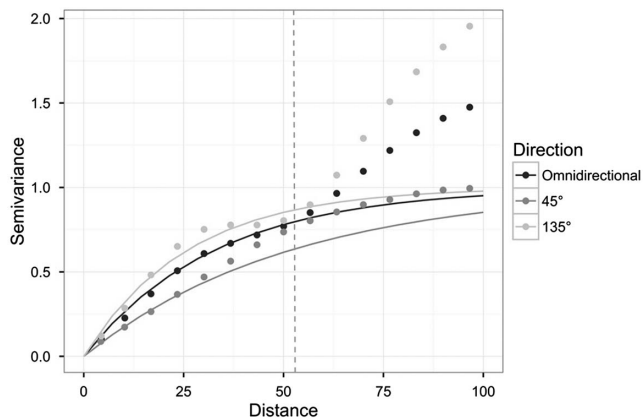


Fig. 4
Fitted variogram models. The range is indicated by the vertical dashed line ($R^2 = 0.96$, $SS_{Err} = 0.15$)

according to Györy et al. (2012). As net cells tend to concentrate into one large geo-object on all of the generated stochastic images, only the connectivity parameters of this largest geo-object were used. The parameters were the following:

- (1) Area
- (2) Area of inner net cells (net cells not in contact with non-net cells)
- (3) Area of boundary net cells inside the geo-object (net cells in contact with non-net cells on the inside of the geo-object)
- (4) Area of boundary net cells on the boundary of the geo-object (net cells in contact with non-net cells on the edge of the geo-object)
- (5) Perimeter–area ratio
- (6) Shape index (ratio of the perimeter to the square root of area adjusted for a square standard)
- (7) Fractal dimension
- (8) Ratio of inner net cells to the area of the geo-object.

As the next step, a principal component analysis was performed on the data to facilitate visualization of the uncertainty field in two dimensions. The principal component axes were calculated for the sequence of 8, 16, 32, 64, 128, 256, 512, and 1,024 realizations to follow the sampling process of the uncertainty field. The retained two principal component axes were able to preserve ~99% of the variance of the original attribute space. The stochastic images were then placed into the uncertainty fields defined by the corresponding two principal component axes. The resulting point clouds were then regarded as two-dimensional spatial point processes.

Results

The point processes representing the selected number of stochastic images can be seen in Fig. 5. The sampling process of the PCA-derived uncertainty field described on the previous pages can be followed through the sequence of plots. Figure 5a and b shows the initial dispersion of the point process followed by the emergence of local high point densities as can be seen in Fig. 5c–e. The gradually increasing aggregation of points reflecting the redundancy of the information content behind the stochastic images can be seen in Fig. 5f–h.

The corresponding inhomogeneous L -functions can be seen in Fig. 6. The continuous lines with a gradient of 1 represent the L -functions of the theoretical inhomogeneous Poisson point process, whereas the dashed lines represent the L -functions of empirical inhomogeneous Poisson point processes calculated using two different edge corrections. Figure 6a–e shows the dispersion of the point processes very clearly: the dashed lines run below the continuous line, suggesting lower empirical point densities compared with the theoretical inhomogeneous Poisson point process. Figure 6f–h suggests increasing empirical point densities, initially only at

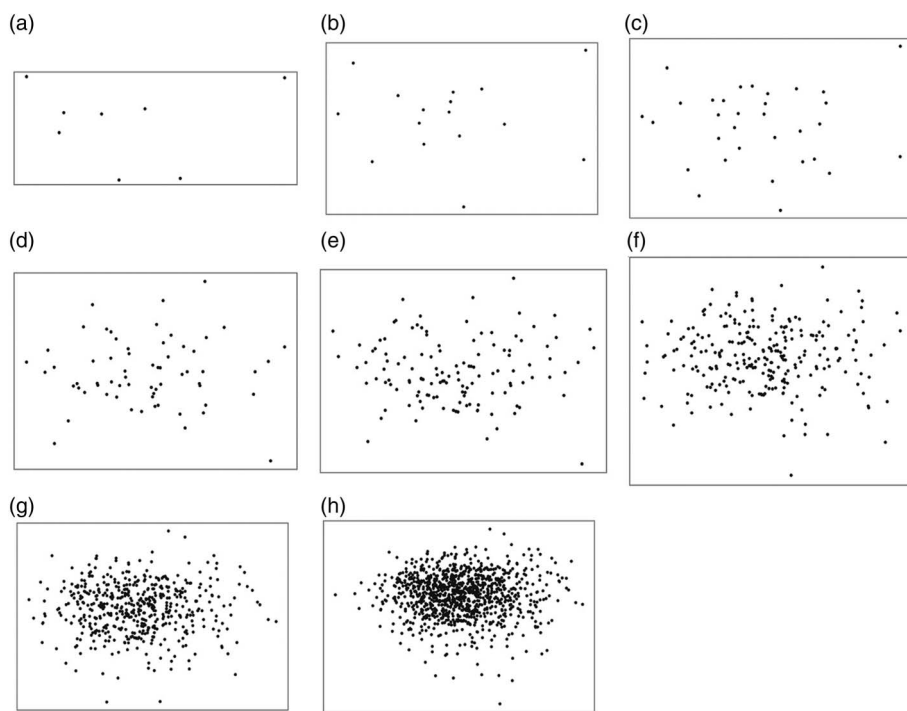


Fig. 5
Point processes representing the following number of realizations: (a) 8, (b) 16, (c) 32, (d) 64, (e) 128, (f) 256, (g) 512, and (h) 1,024

longer distances, then also at smaller distances, depending on the edge correction used. According to Fig. 6, the optimal number of realizations is between 128 (Fig. 6e) and 256 (Fig. 6f).

To find the optimum state, realization numbers between 128 and 256 were checked using smaller increments. The point processes representing these were also generated using the two principal component axes corresponding to the studied number of stochastic images. Based on the L -functions, the optimum state is reached at about 151 realizations (Fig. 7), where the structure of the empirical point process most closely approximates the structure of the empirical inhomogeneous Poisson process (Fig. 7b).

Since the hypothesis of having a Poisson process cannot be validated solely relying on the second moment, a model approximating the intensity of the inhomogeneous Poisson point process was fitted to the empirical data. After comparing the fitted model to the empirical data, the hypothesis of having a Poisson process was accepted (two-sample Kolmogorov–Smirnov test: x coordinate $D = 0.047965$, p value = 0.878; y coordinate $D = 0.04014$, p value = 0.9681).

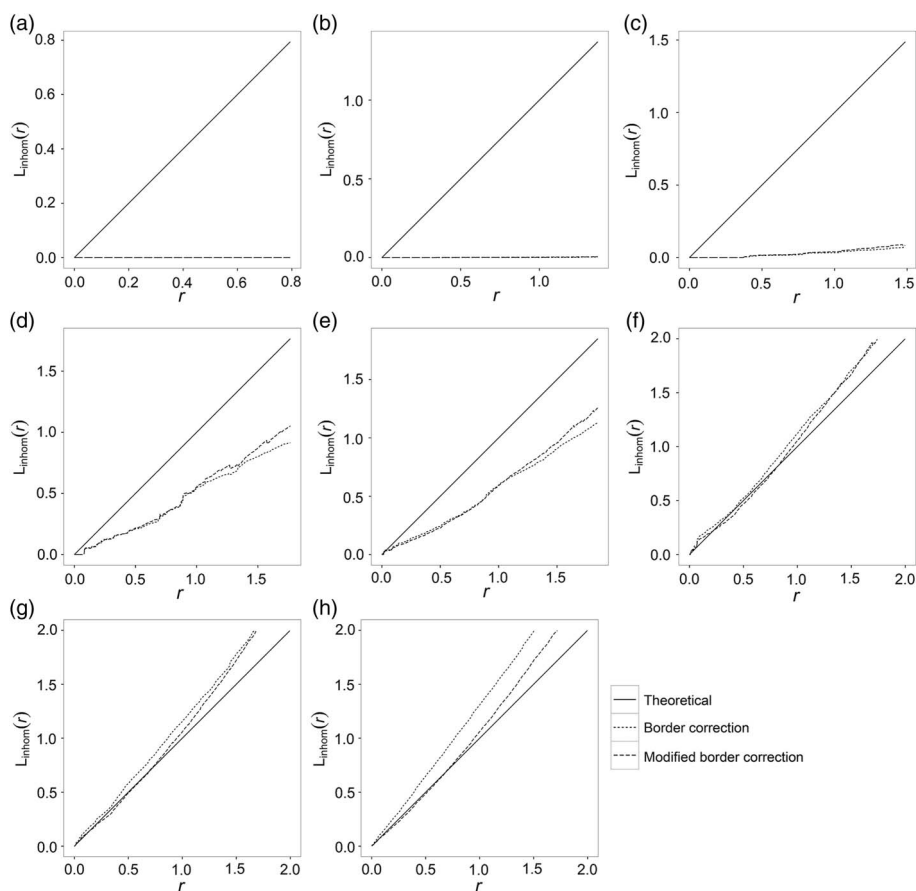


Fig. 6
Inhomogeneous L -functions representing the point processes for (a) 8, (b) 16, (c) 32, (d) 64, (e) 128, (f) 256, (g) 512, and (h) 1,024 realizations

Case study: Tisza-2 reservoir

A real-life application of the method is presented using the example of the Tisza-2 reservoir. The rock body is situated at the top of the Upper Pannonian sequence of the Algyő Field. The best reservoir qualities occur in the southeastern part of the area, where very fine, fine, and medium-grained sandstone can be found in a thickness of 10–15 m. The characteristic well log shapes of the area suggest three typical fluvial sub-environments: point bars, crevasse splays, and channels. The detailed study regarding this rock body can be found in Geiger (2006), where a SGS was used to produce 100 realizations of the well-averaged porosity values

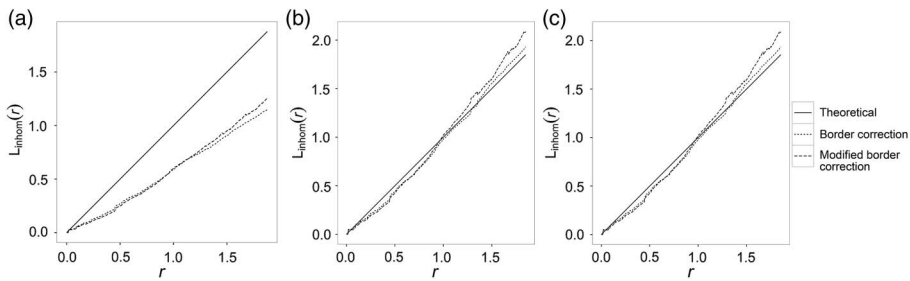


Fig. 7
L-functions for (a) 150, (b) 151, and (c) 152 realizations

(Fig. 8). Since the prospective parts of the reservoir are only found in the southeastern part of the area, only this part of the realizations was used in the further steps.

A 25% porosity value was selected as the threshold between net and non-net cells for the connectivity analysis; values higher than 25% were used to outline the geo-objects. Analyzing the point processes with the inhomogeneous L -function (Fig. 9) showed that for this particular dataset approximately 64 stochastic images are sufficient to obtain a reliable understanding of the field of uncertainty. Creating additional realizations is more likely to increase redundancy of the information content than to provide valuable new information about the possible arrangement of net cells over the area.

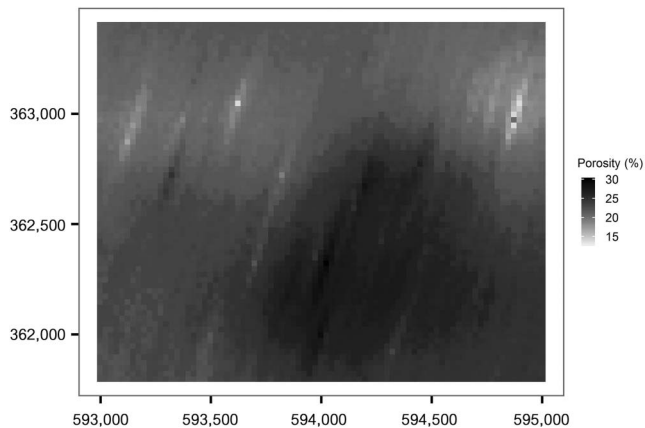


Fig. 8
Southeastern part of the E-type map for 100 realizations of the SGS for the Tisza-2 reservoir (Geiger 2006)

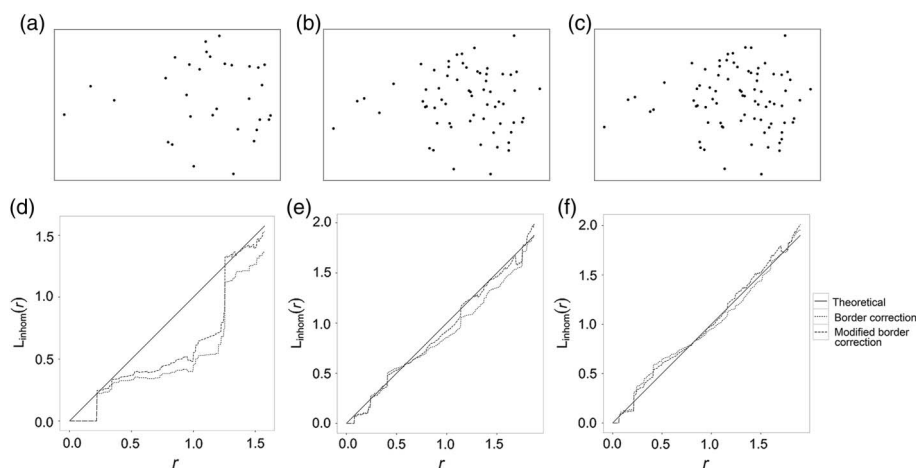


Fig. 9

(a–c) Point processes and (d–f) L -functions for 32, 64, and 70 realizations of the Tisza-2 reservoir

Discussion

For the studied CT slice, it can be stated that the number of realizations should not exceed 151 as any additional realizations will most likely enhance the aggregation of the point process (Fig. 6f–h), instead of providing new information. As can be seen in Fig. 5, the changes in the global structure of the point process become less prominent as more stochastic images are generated. This means that 1,024 realizations (Fig. 5h) do not provide proportionally more insight into the model uncertainty than 512 (Fig. 5g) or 256 realizations do (Fig. 5f). This proves that there is a limit, above which it is not sensible to increase the number of stochastic images.

At the same time, in the initial stages of sampling the uncertainty field, the structure of the point process changes very dynamically, as can be seen in Fig. 5. Thus, at this stage, the decision regarding which realizations may be deemed typical or atypical strongly depends on the number of realizations the decision is based on. This subjectivity cannot be bypassed until the structure of the point process, and thus the structure of the uncertainty field, stabilizes.

While the SGS for the CT measurements required 151 realizations, the SGS for the Tisza-2 well-averaged porosity values only required around 64. It is also worth noticing that, while in the case of the CT slice, the optimum state was reached with a very crisp change in the structure of the point process linked to the 151st realization, in the case of the Tisza-2 reservoir the transition was much smoother.

The reason for this is that the optimal number of realizations depends on the geometry of the studied phenomenon, the chosen grid resolution, the range of the modeled spatial continuity, the location of the input data as well as the number of

grid nodes used to construct the local probability distributions during the simulation.

If the input data and the simulation settings jointly overdetermine the location of simulated net cells on the grid, redundancy may set in after generating a relatively small number of realizations. This means the simulation algorithm is not able to generate any more significantly different new realizations, despite choosing new random paths as we saw in the case of the Tisza-2 reservoir. On the other hand, if the input data and simulation settings allow greater variability in the spatial distribution of simulated net cells, a larger number of realizations may be necessary to reach sufficient sampling of the uncertainty field; hence, more than twice as many realizations were needed in the case of the CT slice.

The advantage of the presented method is that it highlights the optimal number of realizations and facilitates the selection of representative realizations at the same time. This is due to the fact that it places the stochastic images into an attribute space where distances between the realizations reflect the differences between them (Fig. 10).

Since point processes can extend beyond two dimensions, the applicability of the method is also extendable to cases where two attributes are not sufficient to represent the variation of the original attribute space. Similarly, the applicability of the approach is not dependent on the choice of using connectivity attributes, since the uncertainty field may be constructed based on any suitable multidimensional attribute space that is able to reflect the differences between the individual stochastic images. The chosen

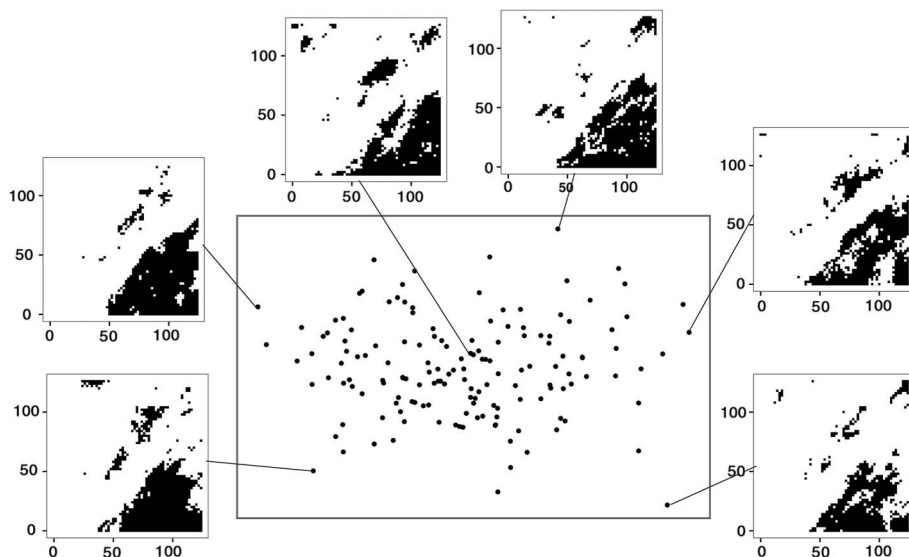


Fig. 10
Point process characterized by 151 stochastic images and position of a few highlighted realizations

dimension reduction technique, if required, however, may affect the structure of the resulting point process. Thus, the effect of using multidimensional scaling, for example, instead of principal component analysis, should be studied.

Acknowledgements

The author would like to thank her supervisor and colleagues for their invaluable insights and support. The author would also like to express her gratitude to Tamás Földes (Petrosoft) for providing the CT measurement data essential for this research.

References

- Baddeley, A.J., J. Møller, R. Waagepetersen 2000: Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. – *Statistica Neerlandica*, 54/3, pp. 329–350.
- Ballin, P.R., A.G. Journel, K. Aziz 1992: Prediction of uncertainty in reservoir performance forecast. – *Journal of Canadian Petroleum Technology*, 31/4, pp. 52–62.
- Besag, J. 1977: Comments on Ripley's paper. – *Journal of the Royal Statistical Society: Series B*, 39/2, pp. 193–195.
- Caers, J. 2011: *Modeling Uncertainty in the Earth Sciences*. – John Wiley & Sons, New York, NY, 246 p.
- Deutsch, C.V. 1998: Fortran programs for calculating connectivity of three-dimensional numerical models and for ranking multiple realizations. – *Computers & Geosciences*, 24/1, pp. 69–76.
- Deutsch, C.V., A.G. Journel 1998: *GSLIB: Geostatistical Software Library and User's Guide*. – Oxford University Press, New York, NY and Oxford, UK, 369 p.
- Geiger, J. 2006: Szekvenciális sztochasztikus szimuláció az övzátontestek kisléptékű heterogenitásának modellezésében (Sequential Gaussian simulation in the modelling of small-scale heterogeneity of point-bars). – *Földtani Közlemények*, 136/4, pp. 527–546. (in Hungarian)
- Geiger, J., J. Újhelyi 2012: Application of Bayes' theorem and entropy sets in the evaluation of uncertainty. – In: Geiger, J., E. Pál-Molnár, T. Malvić (Eds): *Theories and Applications in Geomathematics*. GeoLittera, Szeged, Hungary, pp. 15–37.
- Geiger, J., T. Malvić, J. Horváth, K. Novak-Zelenika 2010: Statistical characters of realizations derived from sequential indicator simulation. – *Proceedings of 14th Annual Conference of the International Association for Mathematical Geosciences, IAMG 2010, Budapest, Hungary*, 5 p.
- Gelfand, A.E., P.J. Diggle, M. Fuentes, P. Guttorp (Eds) 2010: *Handbook of Spatial Statistics*. – CRC Press, Boca Raton, FL, 619 p.
- Gómez-Hernández, J., R.M. Srivastava 1990: ISIM3D: An ANSI-C three-dimensional multiple indicator conditional simulation program. – *Computers & Geosciences*, 16/4, pp. 395–440.
- Goovaerts, P. 1999: Impact of the simulation algorithm, magnitude of ergodic fluctuations and number of realizations on the spaces of uncertainty of flow properties. – *Stochastic Environmental Research and Risk Assessment*, 13/3, pp. 161–182.
- Goovaerts, P. 2001: Geostatistical modeling of uncertainty in soil science. – *Geoderma*, 103/1–2, pp. 3–26.
- Györy, L., G. Kristóf, M. Balogh, J. Geiger, J. Horváth 2012: I-core numerical rock and pore model. – In: Geiger, J., E. Pál-Molnár, T. Malvić (Eds): *Theories and Applications in Geomathematics*. GeoLittera, Szeged, Hungary, pp. 49–70.
- Illian, J., A. Penttinen, D. Stoyan, H. Stoyan 2008: *Statistical Analysis and Modeling of Spatial Point Patterns*. – John Wiley & Sons, Chichester, England, 534 p.
- Isaaks, E.H. 1990: *The application of Monte Carlo methods to the analysis of spatially correlated data*. – PhD Thesis, Stanford University, Stanford, CA, 426 p.

- Journel, A.G., C.V. Deutsch 1993: Entropy and spatial disorder. – *Mathematical Geology*, 25/3, pp. 329–355.
- Journel, A.G., C.H. Huijbregts 1978: *Mining Geostatistics*. – Academic Press, London, UK, 600 p.
- McGarigal, K., S.A. Cushman, E. Ene 2012: FRAGSTATS v4: Spatial pattern analysis program for categorical and continuous maps. – <http://www.umass.edu/landeco/research/fragstats/fragstats.html>.
- Mucsi, L., J. Geiger, T. Malvić 2013: The advantages of using sequential stochastic simulations when mapping small-scale heterogeneities of the groundwater level. – *Journal of Environmental Geography*, 6/3–4, pp. 39–47.
- Pardo-Igúzquiza, E., P.A. Dowd 2003: CONNEC3D: A computer program for connectivity analysis of 3D random set models. – *Computers & Geosciences*, 29/6, pp. 775–785.
- Renard, P., D. Allard 2013: Connectivity metrics for subsurface flow and transport. – *Advances in Water Resources*, 51, pp. 168–196.
- Ripley, B.D. 1977: Modeling spatial patterns. – *Journal of the Royal Statistical Society: Series B*, 39, pp. 172–212.
- Scheidt, C., J. Caers 2009: Representing spatial uncertainty using distances and kernels. – *Mathematical Geosciences*, 41/4, pp. 397–419.
- Srivastava, R.M. 1990: An application of geostatistical methods for risk analysis in reservoir management. – Paper presented at the Society of Petroleum Engineers (SPE) 65th Annual Technical Conference and Exhibition, SPE Paper No. 20608, New Orleans, LA, pp. 825–834.