



ELSEVIER

Computers & Geosciences ■ (■■■■) ■■■–■■■

COMPUTERS &
GEOSCIENCES

www.elsevier.com/locate/cageo

Scaling multiple-point statistics to different univariate proportions

Julián M. Ortiz^{a,*}, Steven Lyster^b, Clayton V. Deutsch^b

^aDepartment of Mining Engineering, University of Chile, Avenida Tupper 2069, Santiago 837 0451, Chile

^bCenter for Computational Geostatistics, Department of Civil and Environmental Engineering, University of Alberta, 220 Civil/Electrical Engineering Building, Edmonton, AB, Canada, T6G 2G7

Received 2 March 2006; received in revised form 13 June 2006; accepted 29 June 2006

Abstract

Multiple-point statistics are used in geostatistical simulation to improve forecasting of responses that are highly dependent on the reproduction of complex features of the phenomenon. Often, complex features cannot be captured by conventional two-point simulation methods, based on the variogram. Inference of multiple-point statistics requires a training image that depicts the geological features of the geological setting being modelled. The proportions of facies in the training image may not match the target statistics of the final model. This is a problem because taking multiple point statistics from a training image also takes the univariate proportions, that is, the multiple point statistics contain all lower order statistics. There is a need to scale multiple-point statistics to different target univariate proportions. In other cases, locally varying facies proportions must be honoured, but a single training image is available. The multiple-point statistics from the training image are scaled to the appropriate target univariate proportions of facies. An iterative scaling approach based on the expression for scaling multiple-point statistics in a purely random case is proposed. The implementation is illustrated through an example where it is shown that the proposed method lies between two extreme cases for a Boolean simulation, namely, the change in size of the objects and the change in their number of occurrences. A second example is presented to illustrate the potential use of this scaling procedure for nonstationary multiple-point geostatistical simulation. © 2006 Published by Elsevier Ltd.

Keywords: Facies simulation; Stochastic modelling; Training images; Geostatistics

1. Introduction

Multiple-point statistics can be used for improved geostatistical modelling of variables distributed in space. The relationships between several points at a time are estimated from training data and imposed during the simulation process in order to achieve a

numerical model that correctly represents the spatial features that conventional simulation cannot capture.

Fig. 1 shows a familiar reference image where the relationships between the different geological units or facies cannot be correctly captured by conventional simulation methods (Deutsch, 1992). Exhaustive images such as the one depicted here can be used as an analogue to the phenomenon that is being modelled; multiple-point statistics can be

*Corresponding author. Tel.: +56 2978 4585; fax: +56 2 672 3504.

E-mail address: jortiz@ing.uchile.cl (J.M. Ortiz).



Fig. 1. An exhaustive image showing intricate relationships of four facies in multiple-point patterns that cannot be easily captured by conventional simulation techniques.

extracted by scanning the image and computing the frequency with which facies arranged in specific patterns occur.

The training information requires abundant data located over a regular grid of points, in order to have enough replicates of each particular multiple-point event. This is often solved by utilizing a training image (Guardiano and Srivastava, 1993; Deutsch, 1992) or by using available pseudo-regularly spaced production data, as is the case in mining applications (Ortiz, 2003; Ortiz and Deutsch, 2004; Ortiz and Emery, 2005).

Multiple-point geostatistical simulation can be performed using any of the available methods. The single normal equation simulation proposed by Strebelle and Journel (2000) estimates the conditional distribution at every location given a multiple-point configuration by calculating the frequency with which the indicator at the location being simulated is one given that the multiple-point event occurs in the training image (see also Strebelle, 2002). Alternatively, simulated annealing (Deutsch, 1992) can be used to match the multiple-point frequencies extracted from a training image into a simulated numerical model. Again, the multiple-point statistics are read from the training image as frequencies of particular events occurring. Other methods such as neural networks also rely on the use of a training image to extract and reproduce the multiple-point statistics (Caers and Journel, 1998).

There is an implicit assumption of stationarity required to export the multiple-point statistics from a training image to the simulated model. A stationarity assumption stricter than the usual second order stationarity is required, since the use

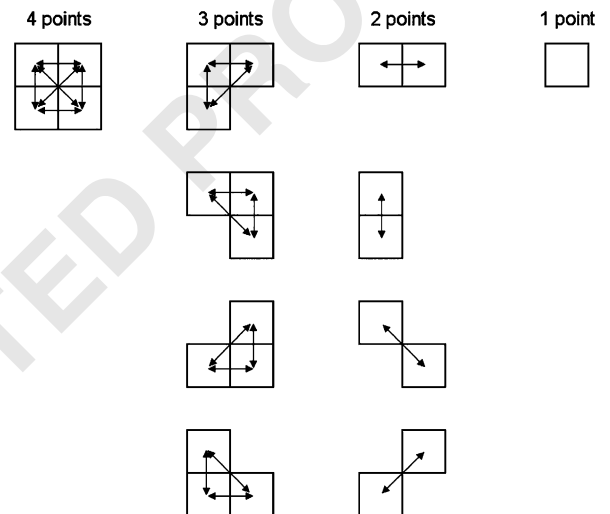


Fig. 2. A four-points pattern and all lower-order configurations that are implicitly matched by honoring the four-point statistics: three-points statistics in four configurations; two-points statistics for the corresponding lag distances separating the centers of the nodes (these represent indicator variogram values); and the one-point statistics that corresponds to the histogram. Nodes are represented by their surrounding squares for illustration.

of configurations of several points also locks all lower order statistics. For instance, if four-point configurations are matched during the multiple-point simulation process, then all three-point and two-point statistics whose configuration is included in the four-point configuration originally used are implicitly matched (Fig. 2). Most importantly, the histogram (univariate proportions of facies or rock types) is also locked when higher-order statistics are honoured. Univariate proportions are of first order importance in resource assessment.

This paper addresses the issue of scaling multiple-point statistics, which may be required in two distinct cases: (1) a training image with different proportions than the data, and (2) locally varying proportions.

First, consistency problems may arise when the training image does not have the same one-point distribution as that inferred from the available site-specific observations. Some kind of correction is required to ensure unbiased proportions in the output realization. Strebelle and Journel (2000) proposed the use of a servo-system that corrects the output mean from any bias caused by the mean of the training image or by the truncation of the multiple-point data events if inference is deemed unreliable. Ortiz and Deutsch (2004) standardized the overall mean when updating the proportions in an indicator based approach to incorporate multiple-point statistics to a continuous variable simulation.

These problems occur frequently, since training images are not readily available with the exact same proportions of facies that are being modelled. Direct use of the multiple-point statistics extracted from the training image will distort the histogram of the simulated realizations generating a bias in the proportions. A basic requirement for the simulated model to be accepted as a plausible representation of the true phenomenon is that it honours a given histogram. Scaling of multiple-point statistics is then required to allow the use of the training image in a consistent manner with the data and preserving its character, hence generating acceptable numerical models.

Another relevant use of a scaling procedure for multiple-point statistics is to impose non-stationary features, but preserving the character provided by the training image. One could devise the use of a single training image to model a field with locally varying proportions of the facies.

Scaling multiple-point statistics to a given set of univariate proportions of the facies could be done by generating a new image with an object-based model, if the conceptual model can be reasonably represented by a Boolean algorithm. Alternatively, dilution/erosion methods could be used to modify the original training image to adjust the proportions to the target ones (Serra, 1982). Local modification of the proportions for application of any multiple-point simulation method would require running the dilution/erosion algorithm locally, which could be inconvenient.

We present a methodology for scaling multiple-point statistics to generate consistent results from simulation methods that account for this information and to consider non-stationary features during the modeling process. The methodology is general and could be adapted to be used with continuous variables, as long as the data are coded as indicators by defining classes through a set of thresholds. The problem is presented in the case of a categorical variable, which is where simulation accounting for multiple-point statistics has seen a faster development.

2. Problem setting

Consider a training image that depicts the spatial arrangement of K categories or facies. The global proportions with which these categories are present in the training image are denoted: $p_k, k = 1, \dots, K$. The general appearance of the training image, deemed appropriate to model a given geological setting, hence the modeler decides the training image is to be used for inference of the multiple-point statistics that a given simulation algorithm will impose to a set of realizations.

The multiple-point statistics to be considered are defined by a spatial arrangement of nodes and by the combination of facies values in these nodes. If we consider the case where an N -points statistic is considered, then K^N possible combinations are available. Each of these combinations occurs in the training image with a given frequency. In fact, as soon as N or K become relatively large, many of the K^N combinations will not occur in the training image.

Each one of the possible combinations is identified with an index that completely defines the facies values within the N points, however the ordering to identify the points in the pattern must be defined prior to the calculation of the index of each multiple-point event (Fig. 3). The index of each multiple-point configuration is calculated as

$$j = 1 + \sum_{n=1}^N (i_n - 1)K^{n-1}$$

where i_n is the code of the n^{th} node of the pattern that identifies its facies. The facies are denoted by consecutive integers starting with code 1.

The frequency of each multiple-point event in the training image is denoted: $f_j, j = 1, \dots, K^N$. Knowing the indexing and the frequencies with which

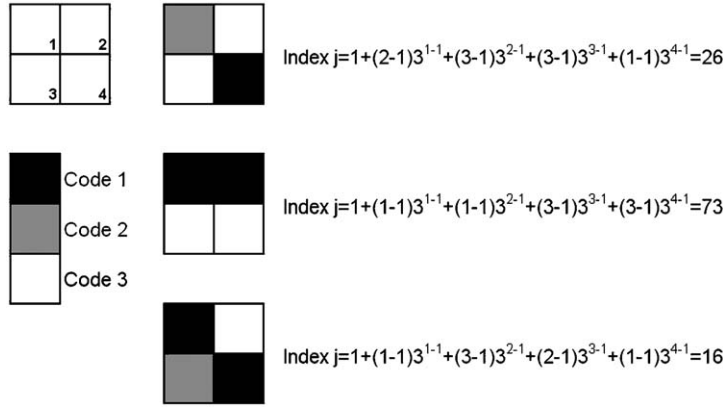


Fig. 3. A four-point configuration, the order for considering the nodes and codes of the facies. Three examples of calculation of the multiple-points indexes are illustrated.

each multiple-point configuration occurs, permits calculation of the facies proportions $p_k, k = 1, \dots, K$. Denoting by $p_{kj}, k = 1, \dots, K; j = 1, \dots, K^N$, the proportion of facies k in the multiple-point arrangement identified with the index j , the proportions of the facies in the training image can be retrieved from the multiple-point statistics as follows. Consider the multiple-point index of interest is j . We can calculate the value of the n^{th} node, by taking:

$$i_N = \text{int} \left(\frac{j-1}{K^{N-1}} \right),$$

where int represents the integer part of the division. Subsequently, the indexes of the $n-1$ remaining nodes of the pattern can be calculated recursively using the residual of the division (fractional part):

$$i_n = \text{int} \left(\frac{\text{frac}(j-1/K^{N-1})}{K^{n-1}} \right), \quad n = 1, \dots, N-1.$$

Knowing the facies values of the N nodes of each multiple-point index and their frequencies, the univariate proportions can be easily calculated.

Now, the simulated model must honor a set of statistics inferred from a set of samples. For instance, the available data may show that the global declustered proportions of the facies in the domain are: $p_k^{\text{target}}, k = 1, \dots, K$, with p_k not necessarily equal to p_k^{target} for some $k = 1, \dots, K$.

The goal of this paper is to propose a methodology to calculate the corrected frequencies of multiple-point events that will honor the target proportions, preserving the “character” of the training image, that is, keeping the features that

make it distinct. These corrected frequencies are denoted: $f_j^*, j = 1, \dots, K^N$.

It should be mentioned that dimensionality becomes quickly a problem when dealing with a large number of categories and a pattern with many points. The indexing proposed above does not preclude the use of the scaling approach proposed. The methodology could be applied in a different context, for instance, when using the search tree considered by Strebel (2002).

3. A scaling approach for multiple-point statistics

To understand the concept of scaling multiple-point statistics, the following example illustrates possible outcomes from an increase in the proportion of a facies in a binary case and where the scaled models preserve the character of the original training image. A field of 1000 by 1000 pixels is populated with 10 by 10 pixels squares, where the center of the squares are randomly located in the field, that is they are generated through a Poisson process (Fig. 4). There are enough squares to cover 20% of the domain, leaving the remaining 80% as background. Scaling this training image so that the proportion of squares goes up to 40% may generate two equally valid outcomes: we can have more 10 by 10 pixels squares or have a smaller number of larger squares, say 20 by 20 pixels squares. These situations are of course extremes and since we do not know the exact multiple-point statistics of the variable and we borrow this information from a training image that does not exactly match the proportions, a scaling procedure that lays somewhere in between the two extremes presented above

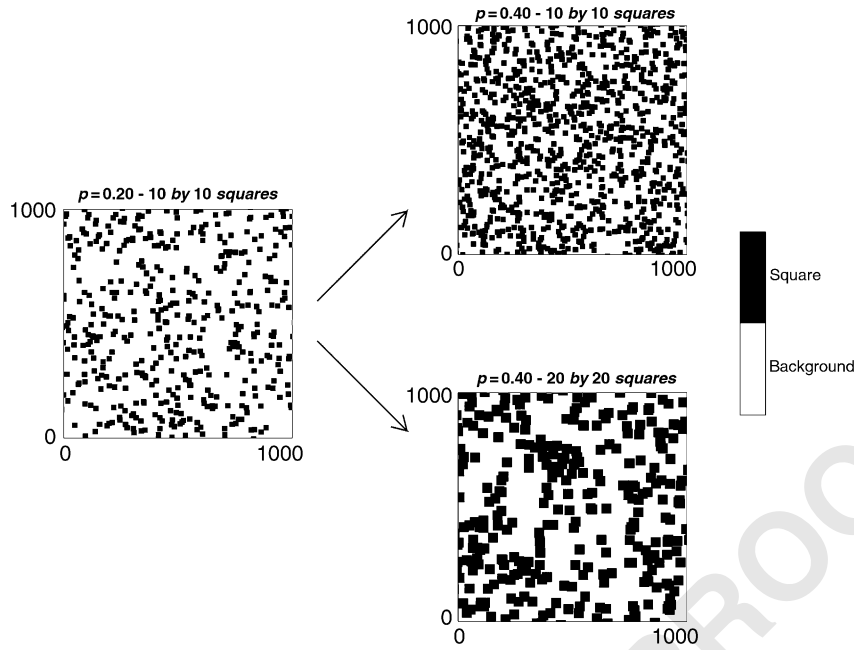


Fig. 4. Two possible upscaled models from the training image on the left. The top right map shows more squares of the same dimension as in the reference image; the bottom right map shows larger squares. Both maps have a proportion of pixels belonging to squares of 0.40.

can be used as a modelling decision to handle the problem.

Dimensionality is always a problem when dealing with multiple-point statistics since the combinatorial becomes very large as the number of points in the multiple-point configuration N or the number of facies K increase. For example, 10 facies can be arranged on a 9 points pattern in 1 billion possible combinations. Of course, this example is extreme, but pattern size often increases in an exponential fashion: 4, 9, 16, 25 points in 2D, or 8, 27, 64, 125 points in 3D. This problem may be partially solved by proceeding pairwise, that is, separate the most relevant facies and group all other facies together. Then, freezing the most important facies already simulated, one can simulate another relevant facies against all remaining facies on a reduced domain, and so on, until all facies have been individually taken into account. There are many implementation considerations, but the focus here is on scaling the multiple-point statistics to be representative.

The proposed approach is based on how multiple-point statistics change when the facies codes of the points in the pattern are randomly distributed. In this case, the frequency with which a multiple-point event occurs can be calculated as the product of the probabilities of occurrence (proportions) of each facies values in its nodes. Since the facies values are

considered uncorrelated, it is straightforward to scale the frequency of occurrence of multiple-point statistics. The frequency of a multiple-point configuration can be calculated by

$$f_j = \prod_{k=1}^K p_k^{Np_{k,j}}.$$

Since $p_{k,j}$ is the proportion of facies k in class j , $N \times p_{k,j}$ is the number of occurrences of facies k in class j . For example (Fig. 5), considering a case where two facies are available and the probability of facies 1 prevailing at a given location is 0.25, then the probability of having a four-points configuration where facies 1 prevails in two nodes and facies 2 prevails in the remaining nodes would be calculated as:

$$f_{13} = 0.25^{4 \times 0.5} \times 0.75^{4 \times 0.5} = 0.03515625.$$

Scaling the random case is quite simple; multiplying the original frequency by a series product of the desired frequency over the old frequency results in a simple equation:

$$f_j^* = f_j \prod_{k=1}^K \left(\frac{p_k^{\text{target}}}{p_k} \right)^{Np_{k,j}}.$$

The scaled multiple-point frequencies honour the target univariate proportions:

$$\begin{aligned}
 f_j^* &= f_j \prod_{k=1}^K \left(\frac{p_k^{\text{target}}}{p_k} \right)^{Np_{k,j}} = f_j \frac{\prod_{k=1}^K (p_k^{\text{target}})^{Np_{k,j}}}{\prod_{k=1}^K (p_k)^{Np_{k,j}}} \\
 &= f_j \frac{f_j^*}{f_j} = f_j^*.
 \end{aligned}$$

Following up on the previous example, we can see

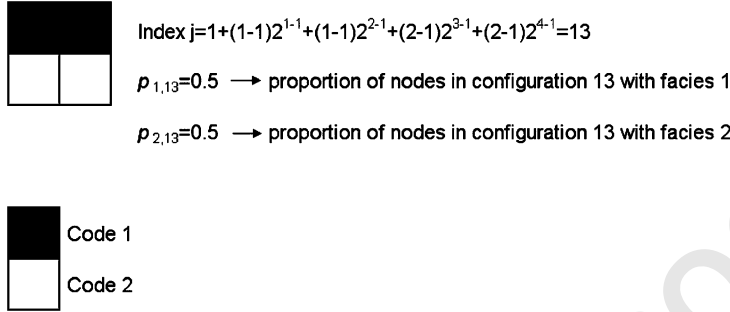


Fig. 5. Example of four-points configuration with two facies to calculate the probability of occurrence of an event in the random case.

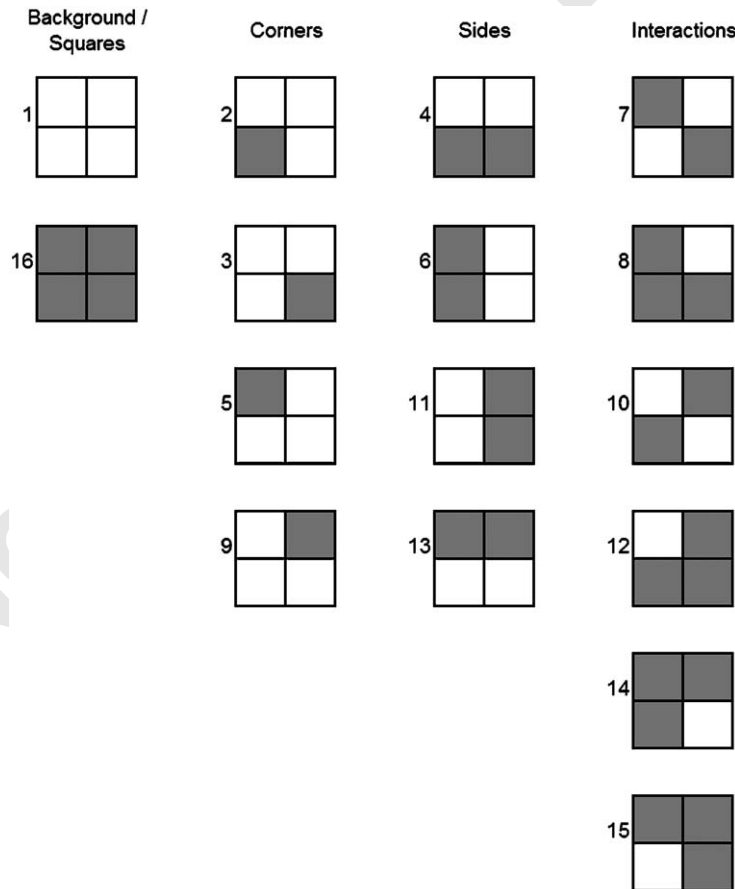


Fig. 6. Geometric interpretation of the different multiple-point configurations for the example of drawing squares randomly in the domain.

prevailing at a location goes up to 0.4, leaving a probability of 0.6 for facies 2. Considering the same four-points configuration, we can calculate its new probability of occurrence:

$$f_{13}^{\text{new}} = 0.4^{4 \times 0.5} \times 0.6^{4 \times 0.5} = 0.0576.$$

It can be checked that the expression provided for scaling multiple-point statistics in the random case provides the same result calculated above:

$$\begin{aligned} f_{13}^* &= f_{13} \left(\frac{p_1^{\text{target}}}{p_1} \right)^{Np_{1,13}} \left(\frac{p_2^{\text{target}}}{p_2} \right)^{Np_{2,13}} \\ &= 0.03515625 \left(\frac{0.4}{0.25} \right)^{4 \times 0.5} \left(\frac{0.6}{0.75} \right)^{4 \times 0.5} = 0.0576. \end{aligned}$$

This expression is valid only in the case of a multiple-point statistics from an uncorrelated variable (pure nugget effect). As soon as the variable is spatially correlated, multiple-point proportions cannot be directly calculated. An iterative approach is proposed next to calculate the multiple-point

frequencies that honour the target proportions, from the initial multiple-point frequencies inferred from the training image.

The formula above overcorrects the multiple-point frequencies. Convergence can be achieved by iterating using the following modified expression:

$$f_j^* = f_j \prod_{k=1}^K \left(\frac{p_k^{\text{target}}}{p_k} \right)^{p_{k,j}}.$$

Multiple-point frequencies for all indexes $j = 1, \dots, K^N$ must be updated. The new global proportions $p_k, k = 1, \dots, K$ are recalculated for each iteration. This formula does not require a large number of iterations, and usually, the desired multiple-point frequencies can be obtained with a nearly perfect match of the target proportions, with less than 50 iterations, which takes only a few seconds.

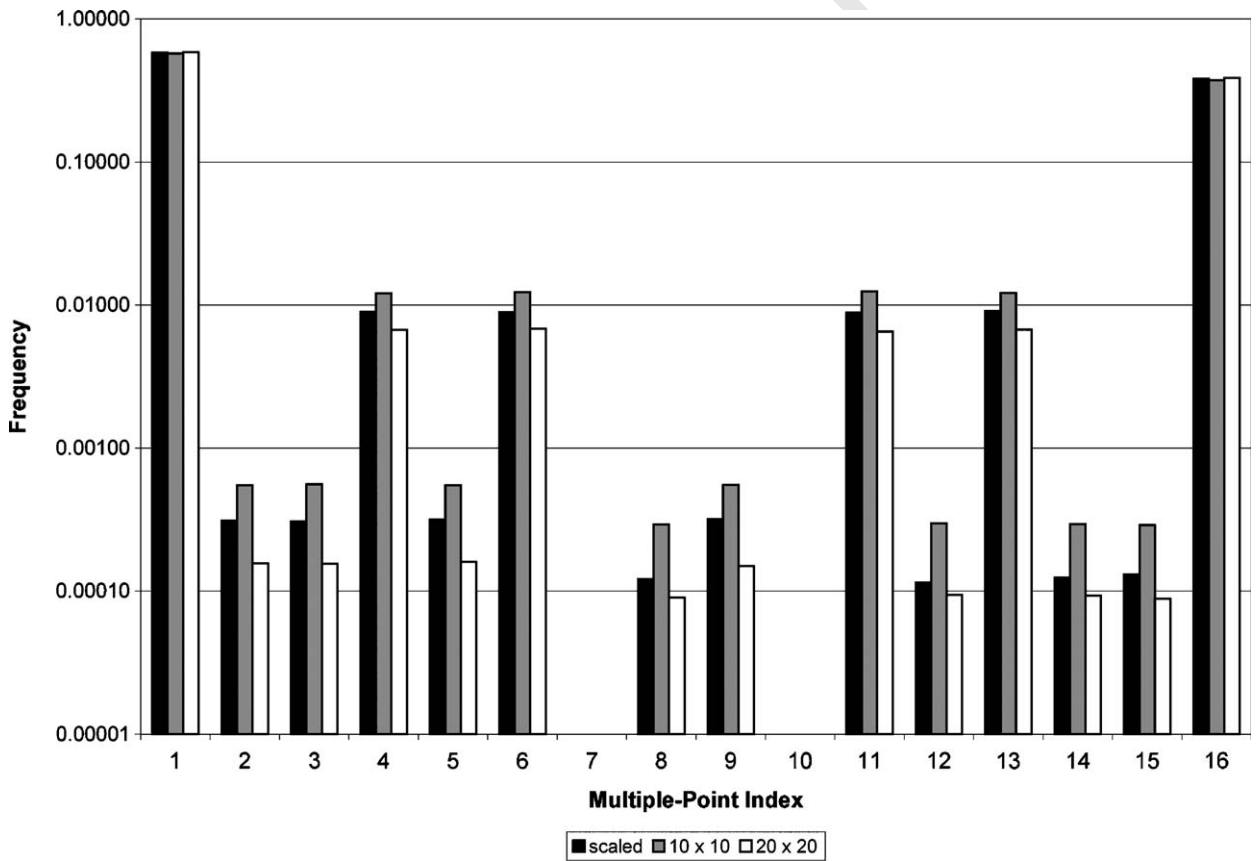


Fig. 7. Multiple-point histogram for the scaled statistics and the two extreme cases: increased number of 10 by 10 squares and increased size of the squares to 20 by 20. A logarithmic scale has been used for the frequency axis to better show that the scaled statistics lie in between the frequencies for the extreme cases.

4. Examples

First, to illustrate the use of the scaling approach, the small example presented in Fig. 4 is expanded. The multiple-point histogram for a two by two points pattern is calculated. The frequency of occurrence of all 16 multiple-point events is computed and a histogram of the frequency with which each indexed event occurs is plotted as a summary. The 16 possible events can be easily interpreted in terms of the geometry of the objects (squares), as shown in Fig. 6.

The reference image (map on the left hand side of Fig. 4) is used to calculate the multiple-point frequencies in a 2 by 2 pixels pattern and these are scaled to reach a global proportion of squares of 0.40. These scaled statistics are then compared with the two extreme cases presented on the right hand side of Fig. 4.

Fig. 7 shows the three multiple-point histograms superimposed. It can be seen that the scaling using the random approach provides statistics that lie in between the two extreme cases. We can expect that these scaled up statistics represent a case where

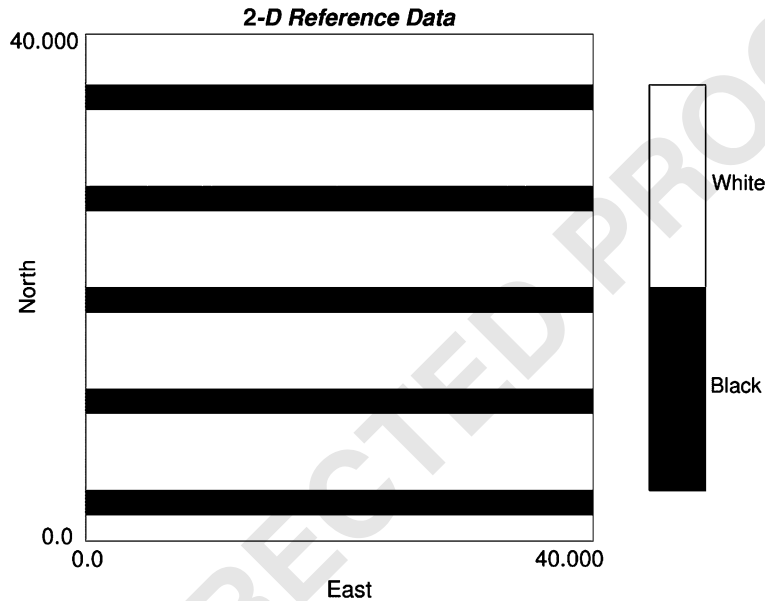


Fig. 8. Reference training image. Black stripes are two nodes wide and separated by 6 nodes in the north direction.

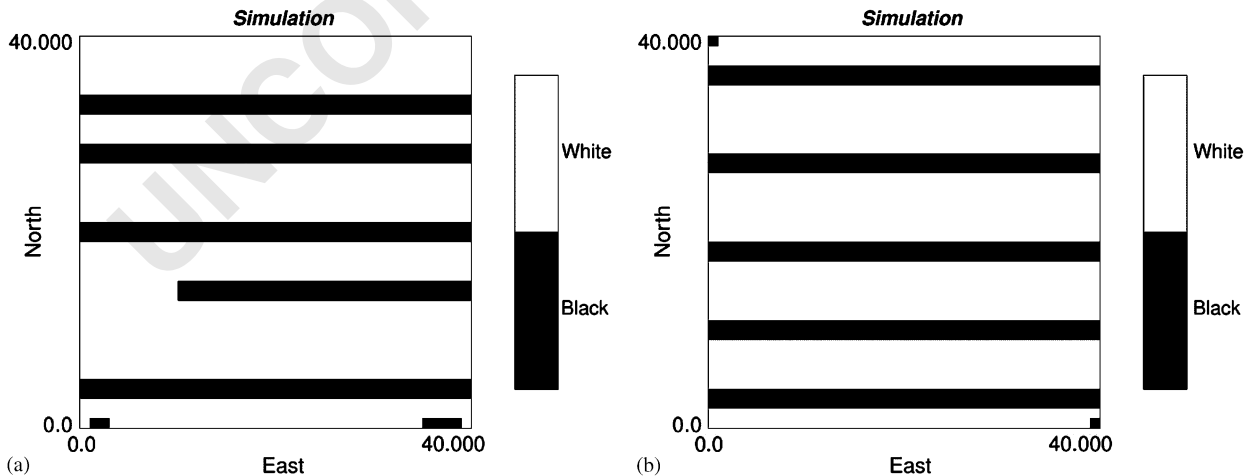


Fig. 9. Two realizations that honor the original proportions, made considering a 4 by 4 nodes pattern.

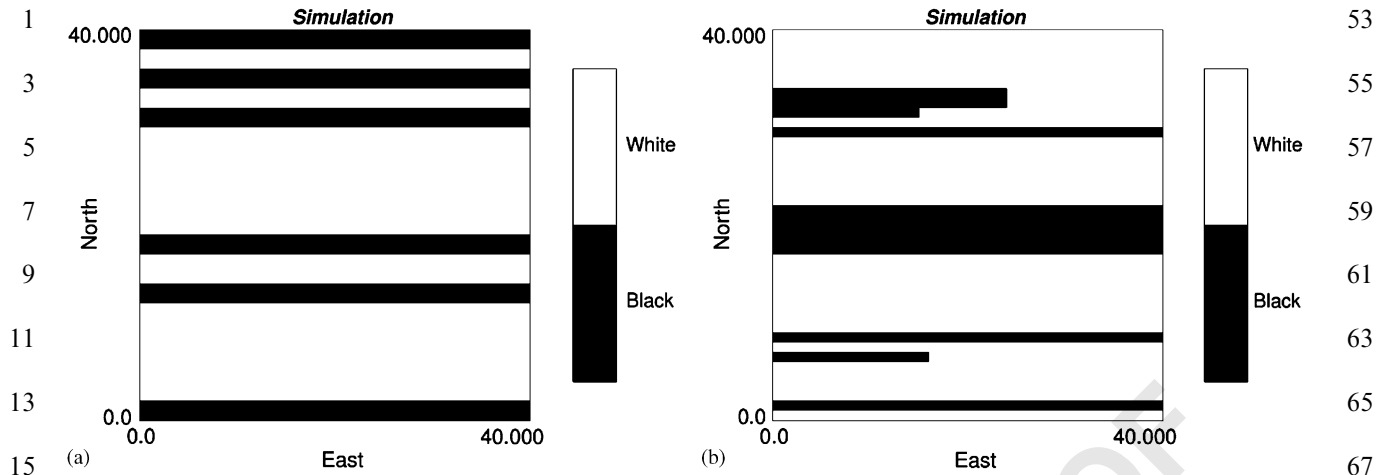


Fig. 10. Two realizations that honor the original proportions, made considering statistics extracted from a 2 by 2 (left) and a 3 by 3 nodes pattern (right).

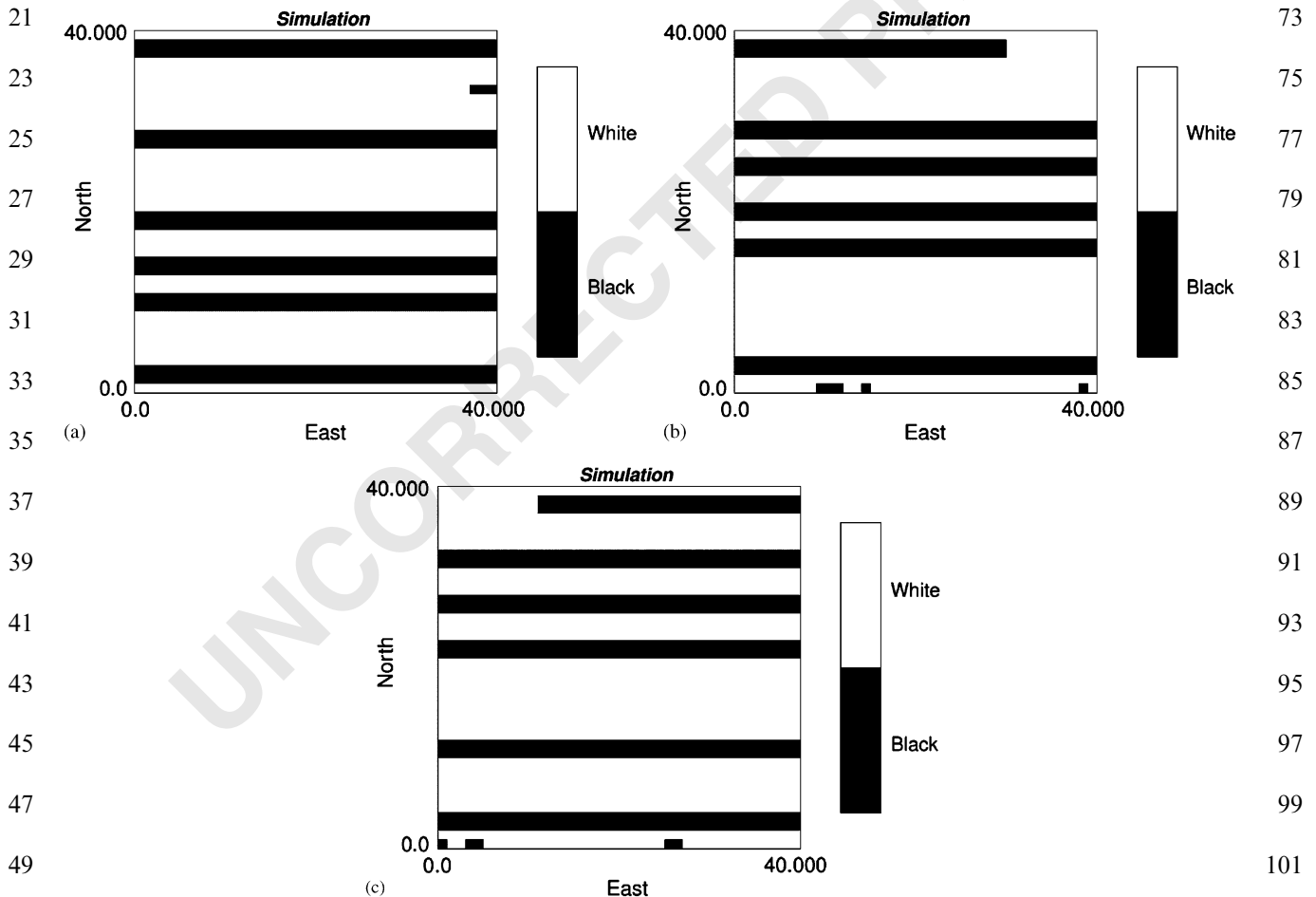


Fig. 11. Three realizations considering the statistics from a 3 by 3 nodes pattern and scaled proportion of 30% for the black facies and 70% for the white background.

there are more squares in the domain, but these are also slightly larger than the ones in the training image. The relevant features of the reference image are then captured in the statistics.

A second example has been designed to illustrate the use of the scaling procedure proposed. A binary two-dimensional training image has been created with black horizontal stripes two-units thick in a white background. The training image is shown in Fig. 8. The original proportions are $p_{\text{black}} = 0.25$ and $p_{\text{white}} = 0.75$.

Multiple-point statistics have been inferred from the training image and unconditional realizations have been computed for different proportions of black (stripes) and white (background) facies. A simulated annealing program to match the multiple-point statistics was prepared. Different pattern sizes imply control over different scales of the phenomenon. For example, Fig. 9 shows two realizations

considering a 4 by 4 nodes pattern. Notice that with this pattern size the general appearance of the resulting models is close to the training image. If a different pattern is considered, the long range features may not be properly captured. Fig. 10 shows realizations made with different pattern sizes to illustrate this effect.

Fig. 11 shows three realizations constructed using the multiple-point statistics for a 3 by 3 nodes pattern, scaled to match the following facies proportions: 30% black and 70% white. These proportions were then changed to 50% for the black facies and 50% for the white background, and realizations with statistics from 2 by 2, 3 by 3, and 4 by 4 nodes patterns considered. These results are shown in Fig. 12.

The scaling procedure was used to scale up the proportion of the black facies. The initial proportions (25% black, 75% white) were changed and the

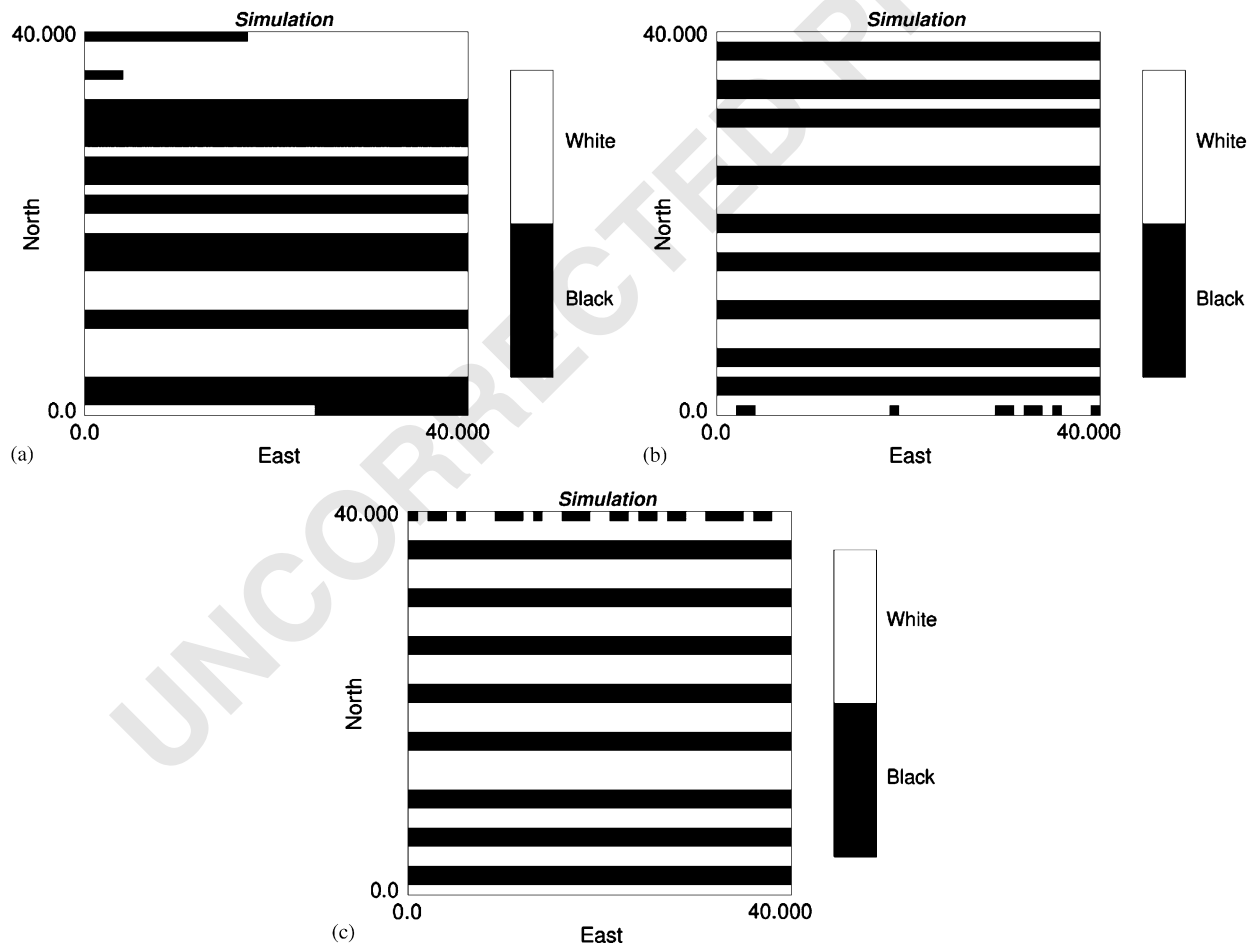


Fig. 12. Three realizations with multiple-point statistics scaled to 50% black facies and 50% white background, with increasing pattern size: 2 by 2 (left), 3 by 3 (middle), and 4 by 4 (right).

multiple-point statistics were scaled. These were then used to simulate realizations with the intent of generating realizations that kept the essential features of the training image, but honoring the facies proportions imposed to the new models.

It can be seen from the realizations displayed above that, depending on the pattern size, two situations can arise: the stripes may become wider if the pattern is small enough not to capture its entire width, or the stripes may become more abundant if the pattern is big enough to capture the relationships between the width of the black facies objects and the background, as is the case when 3 by 3 or 4 by 4 patterns are used.

5. Conclusion

Scaling multiple-point statistics can improve the current use of multiple-point geostatistical simulation techniques by allowing locally varying proportions of facies to be reproduced still honoring the relationships captured by multiple-point statistics. Another important use of a multiple-point scaling technique is the use of a representative training image that does not exactly match the target simulation proportions.

An iterative approach to scale multiple-point statistics was developed based on the way that multiple-point statistics change for a purely random field. This approach can be shown to provide reasonable values of the statistics that lie in between two extreme cases that are easy to show in the case of objects: an increase in the number of objects or an increase in the size of the objects.

The scaling procedure has the potential of being incorporated in a nonstationary simulation algorithm that honors the features provided by a training image, through the local updating of multiple-point statistics.

Acknowledgments

The authors would like to acknowledge the Chair in Ore Reserve Evaluation at the University of Chile sponsored by Codelco Chile and the industry sponsors of the Center for Computational Geostatistics at the University of Alberta for supporting this research. The comments by two anonymous reviewers helped improve the original manuscript and are acknowledged by the authors.

References

- Caers, J., Journel, A.G., 1998. Stochastic reservoir simulation using neural networks trained on outcrop data. In: 1998 SPE Annual Technical Conference and Exhibition, New Orleans, LA. Society of Petroleum Engineers. SPE paper No. 49026, pp. 321–336.
- Deutsch, C.V., 1992. Annealing techniques applied to reservoir modeling and the integration of geological and engineering (well test) data. Ph.D. Thesis, Stanford University, Stanford, CA, 306 pp.
- Guardiano, F., Srivastava, M. (Eds.), 1993. Multivariate Geostatistics: Beyond Bivariate Moments, In: Soares, A., Geostatistics Tróia'92, vol. 1. Kluwer Academic Publication, Dordrecht, pp. 133–144.
- Ortiz, J.M., 2003. Characterization of high order correlation for enhanced indicator simulation. Ph.D. Thesis, University of Alberta, Edmonton, AB, Canada, 255 pp.
- Ortiz, J.M., Deutsch, C.V., 2004. Indicator simulation accounting for multiple-point statistics. *Mathematical Geology* 36 (6), 545–565.
- Ortiz, J.M., Emery, X., 2005. Integrating multiple-point statistics into sequential simulation algorithms. In: Leuangthong, O., Deutsch, C.V. (Eds.), *Geostatistics Banff 2004*, vol. 2. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 969–978.
- Serra, J., 1982. *Image Analysis and Mathematical Morphology*, vol. 1. Academic Press, London, 600pp.
- Strebelle, S., 2002. Conditional simulation of complex geological structures using multiple-point statistics. *Mathematical Geology* 34 (1), 1–21.
- Strebelle, S., Journel, A.G., 2000. Sequential simulation drawing structures from training images. In: *Proceedings of the Sixth International Geostatistics Congress, Cape Town, South Africa, Geostatistical Association of Southern Africa, 2000*.