

Verifying the high-order consistency of training images with data for multiple-point geostatistics



Cristian Pérez ^{a,b,*}, Gregoire Mariethoz ^{c,d}, Julián M. Ortiz ^{a,b}

^a ALGES Laboratory – Advanced Mining Technology Centre (AMTC), Universidad de Chile, Santiago, Chile

^b Department of Mining Engineering, Universidad de Chile, Santiago, Chile

^c School of Civil and Environmental Engineering, The University of New South Wales, Sydney, Australia

^d National Centre for Groundwater Research and Training, Australia

ARTICLE INFO

Article history:

Received 6 November 2013

Received in revised form

30 May 2014

Accepted 4 June 2014

Available online 13 June 2014

Keywords:

Multiple-point statistics

Direct sampling

Sensitivity analysis

Training image

Inference

Selection

ABSTRACT

Parameter inference is a key aspect of spatial modeling. A major appeal of variograms is that they allow inferring the spatial structure solely based on conditioning data. This is very convenient when the modeler does not have a ready-made geological interpretation. To date, such an easy and automated interpretation is not available in the context of most multiple-point geostatistics applications. Because training images are generally conceptual models, their preparation is often based on subjective criteria of the modeling expert. As a consequence, selection of an appropriate training image is one of the main issues one must face when using multiple-point simulation. This paper addresses the development of a geostatistical tool that addresses two separate problems. It allows (1) ranking training images according to their *relative* compatibility to the data, and (2) obtaining an *absolute* measure quantifying the consistency between training image and data in terms of spatial structure. For both, two alternative implementations are developed. The first one computes the frequency of each pattern in each training image. This method is statistically sound but computationally demanding. The second implementation obtains similar results at a lesser computational cost using a direct sampling approach. The applicability of the methodologies is successfully evaluated in two synthetic 2D examples and one real 3D mining example at the Escondida Norte deposit.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Geological models are often built using deterministic techniques, meaning that their construction relies on the knowledge and experience of a specialist that assigns geological attribute values to a given volume. This practice is however not satisfying because it does not allow quantifying geological uncertainty (Refsgaard et al., 2012). With increasing frequency in recent years, geostatistical simulation has been used to construct stochastic models. Most of these techniques rely on the available sparse data (such as for example boreholes) to infer spatial continuity. This is accomplished using statistics based on relations between pairs of points, as for example variograms or correlograms (Caers, 2005; Deutsch and Journel, 1992; Isaaks and Srivastava, 1990; Journel and Huijbregts, 1978; Kitanidis, 1997). These conventional 2-point simulation methods are relatively simple and often show appropriate

results. However, they present certain inherent limitations. Among those:

- Curvilinear or complex structures can be poorly represented by Gaussian simulations and may require higher order statistics (Journel and Zhang, 2006).
- Variograms do not inform about contacts or contact geometries between different categories (e.g. Carle and Fogg, 1996).
- The reliance of variogram-based Geostatistics on the maximum entropy, multi-Gaussian distribution to model all statistics beyond the two-point statistics results in maximum disconnection of extremes and the reproduction of only linear spatial features (e.g. Boisvert et al., 2007; Zinn and Harvey, 2003).

As an alternative, multiple point simulation algorithms, have recently become an important point of focus, with a wealth of different methods developed in the last decade (Arpat and Caers, 2007; Gloaguen and Dimitrakopoulos, 2009; Guardiano and Srivastava, 1993; Honarkhah and Caers, 2010; Mariethoz et al., 2010; Parra and Ortiz, 2011; Straubhaar et al., 2011; Strebelle, 2002; Tahmasebi et al., 2012; Zhang et al., 2006). These algorithms

* Corresponding author at: Department of Mining Engineering, Universidad de Chile, Avenida Tupper 2069, Santiago, Chile. Tel./fax: +56 222772556.

E-mail address: criperez@ing.uchile.cl (C. Pérez).

model spatial continuity using higher order statistics, and therefore do not use variographic models to impose a spatial structure. As the amount of higher-order events contained in scattered data is usually insufficient, multiple point simulation algorithms require inferring the statistics of spatial patterns from a training image. Because training images are conceptual models, their preparation is often based on subjective criteria of the modeling expert. As a consequence, verifying the consistency of the training image with data is one of the main issues a modeling professional must face when using multiple-point simulation. It is even more important than variogram modeling for classical geostatistical simulation, as it controls higher-order, as well as second order spatial relations.

1.1. Background on training image selection

Two different types of data can be distinguished to serve as a base for the selection of training images. The first type is indirect state data such as flow and transport, which are typically integrated through inverse methods. The problem consists in evaluating the compatibility of a training image with dynamic outputs (for example time series of contaminant output). Approaches to select training images on this basis have been proposed in the framework of distance-based approaches (Suzuki and Caers, 2008). Another avenue in this direction is to consider each candidate training image as a prior model and to weight the different priors using a Bayesian mixture model (Khodabakhshi and Jafarour, 2013). A drawback of such approaches is that in general the selection of training images based on state data requires expensive forward model runs.

The other type of data, which is the specific focus of this paper, is static data. As for variograms, it consists in quantifying the consistency of a training image based on spatial statistics derived on both training image and data. To date, significant developments are lacking regarding objective criteria for verifying high-order training image consistency with the available scattered data. In this paper we focus on the consistency problem, and we leave aside some related issues as for example training image scaling issues (e.g. Ortiz et al., 2007).

One of the first approaches proposed for training image selection based on static data was initially proposed by Ortiz and Deutsch (2004). It consists in comparing the cumulative distribution of runs of the training image with the cumulative distribution of runs observed in 1D wells. Boisvert et al. (2007) proposed another method based on the comparison of multiple point histograms for vertical one-dimensional patterns. The training image and the conditioning data are scanned using a search template and the resulting statistics are compared.

A different approach is suggested by Eskandaridavand (2008), who proposes a spiral search method. It loops over all conditioning nodes, and for each of them, over all training nodes. If the node in the training image has the same value as the conditioning node, the method loops over the close conditioning nodes from the nearest to the farthest. The values are compared to the nodes in the training image that show the same spatial configuration relative to the central node. If both values have the same relationship, i.e. if both increase or decrease in the same manner, a counter for compatible nodes increases. The method allows obtaining a distribution of compatible training nodes for each conditioning node and a unique distribution of maximum compatible nodes. These distributions can be used for deriving a measure of consistency between the training image and the conditioning information.

As an alternative to the methods mentioned above, we note that spatial cumulants are promising because they offer a parametric description of the high-order spatial statistics (Dimitrakopoulos

et al., 2010). However, to this day cumulants have not been used in the context of training image selection.

This paper addresses the development of a geostatistical tool that provides two measurable criteria for selecting training images based on their consistency with given data. The first method, applicable in cases where more than one conceptual model is available, allows ranking training images according to their compatibility with a data set in terms of low and high order spatial structures. This represents a *relative compatibility* measurement. With this relative measurement, even a somewhat incompatible image can potentially be top ranked if all other available images are even less compatible. To overcome this limitation, an *absolute compatibility* measurement method is developed, which computes the probability of finding patterns in a given training image. These metrics are able to compare conditioning data with pattern statistics of training images across a range of statistical orders. It is important to remark that if modeling takes place in a deformed space, the data to be used should be considered after deformation. Three examples are addressed: the first one is a simple synthetic 2D example, in the second one our method is used to identify non-stationarity in a large synthetic 3D alluvial model, and the third one demonstrates the applicability on a real 3D mining example.

2. Methodology

The purpose of the proposed algorithm is to generate a ranking of several training images, according to their compatibility with conditioning data. In essence, the algorithm is given a conditioning data set and a series of training images. The method works by defining conditioning data events, which are patterns of spatially distributed data values, and computing their frequency of occurrence in the different training images. The training images that have a higher frequency of data events are deemed more coherent with the data. The result is a ranking of the training images according to their data consistency. The overall algorithm can be divided in a number of steps whose implementation is described below in details. We first present an algorithm that is statistically straightforward but computationally inefficient. In a second step, we present an equivalent alternative that yields similar results with a much lesser computational burden.

2.1. Conditioning data event definition

The first step of the method is to migrate the scattered data to a regular grid. The grid structure permits to accelerate future spatial searches of data events, therefore reducing computation times. To avoid scale issues, the nodes spacing in the user defined

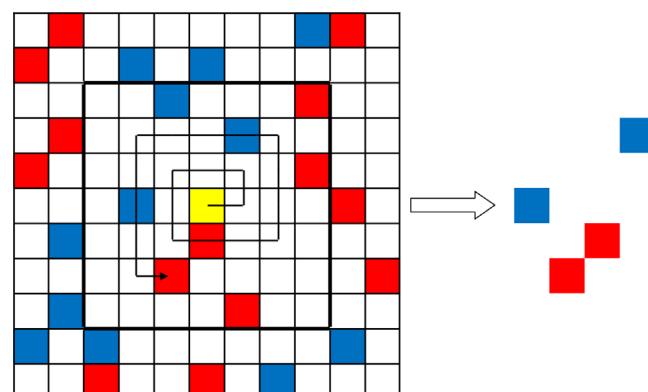


Fig. 1. Spiral search (left) and defined conditioning event (right).

migration grid must be the same as in the training images. The algorithm migrates data points to the closest node of the grid.

Once migration is completed, conditioning data events (\mathbf{CE}_i), to be used for comparing spatial statistics of the data and of the training images, are defined. At every position of the grid, regardless of whether it contains a migrated data point, a spiral search is started (Fig. 1). In order to avoid using very large data events, the user needs to define a maximum search radius in each direction and a minimum and maximum number of conditioning nodes that form the event, which corresponds to the minimum and maximum order of spatial structure consistency that is going to be analyzed. The spiral search goes from the nearest (starting at the center node) to the farthest point. The positions, defined as offsets to the central node and facies of the informed nodes, are saved in a list. If not enough data points are found in the prescribed radius (i.e. if the maximum search radius is reached without finding the minimum required amount of conditioning nodes), the data event is skipped.

The data events are then used to compare training images with data. It should be noted that data events with a small number of nodes can be used to reflect lower-order statistics, whereas data events with increasing nodes will represent a higher order spatial distribution. For example, using a set of patterns of 1 node allows representing the marginal distribution, patterns of 2 nodes can represent 2-point dependence (similar to a variogram), etc. The appealing aspect of this approach is that it allows for comparisons using multiple-order criteria. In the following, we will use equivalently the terms of “number of neighbors” and “data event order”.

2.2. Relative compatibility with exhaustive scanning of the training image

Data events are repeatedly extracted from the conditioning data, and similar patterns in the different training images are inventoried. Event occurrences are searched using a linear search path that runs over all the positions contained in the images. It is considered that an event found in the training image matches the conditioning data event if the positions and facies of all nodes of the conditioning event are matched by those in the training image. Since it can be difficult to find an exact match, a tolerance parameter t is introduced, which represents the proportion of nodes in the conditioning event whose facies are allowed to be not matched in the pattern of the training image. This parameter can take values between 0 and 1.

Considering, $Tl_j(\vec{\mathbf{X}}_{i,n})$ the value of the facies in the training image j at the offsetting coordinate $\vec{\mathbf{X}}_{i,n}$ the facies value at that conditioning node n , the following indicator variable is used to quantify the match for each node in \mathbf{CE}_i .

$$Ind_{i,j,n} = \begin{cases} 0 & \text{if } Tl_j(\vec{\mathbf{X}}_{i,n}) = F_{i,n} \\ 1 & \text{if } Tl_j(\vec{\mathbf{X}}_{i,n}) \neq F_{i,n} \end{cases} \quad (1)$$

It should be mentioned that the relative compatibility measurement algorithm can also be applied to continuous variables. For this kind of applications another acceptance threshold ϵ is needed. The threshold that should be applied is then:

$$Ind_{i,j,n} = \begin{cases} 0 & \text{if } |Tl_j(\vec{\mathbf{X}}_{i,n}) - F_{i,n}| < \epsilon \\ 1 & \text{otherwise} \end{cases} \quad (2)$$

Considering N_i the order of \mathbf{CE}_i a pattern analyzed in an image is considered a match if:

$$\sum_{n=1}^{N_i} Ind_{i,n} \leq t \quad (3)$$

Every time a match is found the value of the counter $M_{i,j}$ is increased, which corresponds to the number of matches for each data event in each training image. Once all occurrences are counted, the relative frequency of this specific data event in each image ($P_{i,j}$) is computed:

$$P(\mathbf{CE}_i \in \mathbf{Tl}_j | \mathbf{CE}_i \in \mathbf{Tl}) = P_{i,j} = \frac{M_{i,j}}{\sum_{j=1}^J M_{i,j}} \quad (4)$$

$P_{i,j}$ can be interpreted as the proportion of occurrences of the conditioning event \mathbf{CE}_i in the training image \mathbf{Tl}_j , relative to its occurrence in the set of analyzed training images \mathbf{Tl} .

Finally, the relative compatibility between the training images and the conditioning data set is calculated by normalizing the sum of frequencies of each conditioning event in a given training image. This relative compatibility is calculated with the following equation:

$$C_j = \frac{\sum_{i=1}^I P_{i,j}}{\sum_{i=1}^I \sum_{j=1}^J P_{i,j}} \quad (5)$$

This relative compatibility value between the data set and each training image is the final result of the algorithm. As compatibilities are calculated through normalization, the sum of the components of \mathbf{C} equals 1. It is important to note that this compatibility measurement is relative, as these frequencies are subject to finding the conditioning event in the analyzed image set. Therefore C_j would change if new training images are added to the set of images \mathbf{Tl} , and it does not give an absolute ranking of whether an individual training image is good or bad.

2.3. Relative compatibility with alternative direct sampling method

The previously proposed method is statistically sound as it checks the occurrence of conditioning patterns in every position of the ranked training images. However, because of the high amount of positions that need to be analyzed, the CPU cost is important. A faster alternative approach inspired from Direct Sampling (Mariethoz et al., 2010) is proposed.

Scattered conditioning data points migration and conditioning event definition remain unchanged. However the occurrence count of conditioning events in the images and compatibility calculations are achieved differently. In this case, the search for occurrences of a conditioning pattern in the training images is done using a random path that goes through the different training images. Considering the tolerance parameter t , if one training image j matches the conditioning event at the checked position, a counter (L_j) for that image is incremented and the search is stopped. The algorithm then carries on with the next node in the conditioning data grid. Note that in the algorithm described in Section 2.2, the search would continue in order to count all matching occurrences. In contrast, here we stop at the first occurrence, allowing for computational gain. In case more than one image matches the conditioning event simultaneously, the counter of every matching image is incremented.

Compared to the previous method consisting in scanning the entire training image, the user needs to define the maximum fraction f of the training image that can be scanned. This parameter corresponds to the fraction of total positions (fraction of total nodes) of the training image that are analyzed in the search of a match, before stopping the loop. If a fraction f is reached without a match, it is considered that the searched pattern does not occur in the training image and the algorithm passes on to the next node in the conditioning data grid.

The compatibility between the conditioning data set and the training images is finally calculated using the following equation:

$$C_j \approx \frac{L_j}{\sum_{j=1}^J L_j} \quad (6)$$

If the amount of conditioning events and size of the training images is high, this relative compatibility value should equal the value of relative compatibility.

2.4. Absolute compatibility measurement

In several cases it can be desirable to have an absolute measure of the compatibility between training images and data. For example, it may not be relevant to establish that one training image is slightly better than another, if both training images are grossly incompatible anyway. We therefore propose an absolute compatibility measurement, which is represented by the proportion of patterns contained in the data that are matched by the training image. This absolute compatibility is computed by analyzing each training image separately, as this absolute measure depends only of the analyzed training image and is independent from other images. It fluctuates between values of 0 and 1: if each scanned conditioning event is found in the training image, which would represent the best possible case, the absolute compatibility is 1. Conversely, if no data event from the data is found in the training image, the value is 0.

To calculate this measure, an indicator (Y_{ij}) is introduced. It takes a value of 1 if \mathbf{CE}_i is contained in the \mathbf{TI}_j , and a value of 0 if it is not, considering the tolerance t . Unlike M_{ij} , Y_{ij} indicates if a pattern is present or absent in the training image whereas M_{ij} counts how many times it occurs in the training image. The absolute matching patterns proportion (MP) can be calculated using the following equation,

$$MP_j = \frac{\sum_{i=1}^I Y_{ij}}{I} \quad (7)$$

where I is the total number of patterns searched. As for the relative measure, we implemented a fast DS-based approach, with a training image scanned fraction of 1 ($f=1$). In this case the absolute compatibility A_j of training image j with data can be calculated as

$$MP_j \approx A_j \approx \frac{L_j}{I} \quad (8)$$

3. Application examples

This section presents three application examples of the proposed method. In all examples, 0 is used for tolerance t , and categorical variable is considered. For a given conditioning event order, the results only consider those of exactly that order (i.e. we

Table 1
TiGenerator parameters, example 1.

Training image	Geometry	Orientation	Structure characteristics			Proportion
1	Ellipsoid	$60^\circ \pm 15^\circ$	Max rad	Med rad	Min rad	0.302
4		15	50	5	0.302	
2	Sinusoid	$0^\circ \pm 10^\circ$	Amplitude	Wavelength	Width	0.304
5		15	3	1	1	0.331
3	Cuboid	0°	Width	Height		0.320
6		3	1			0.300

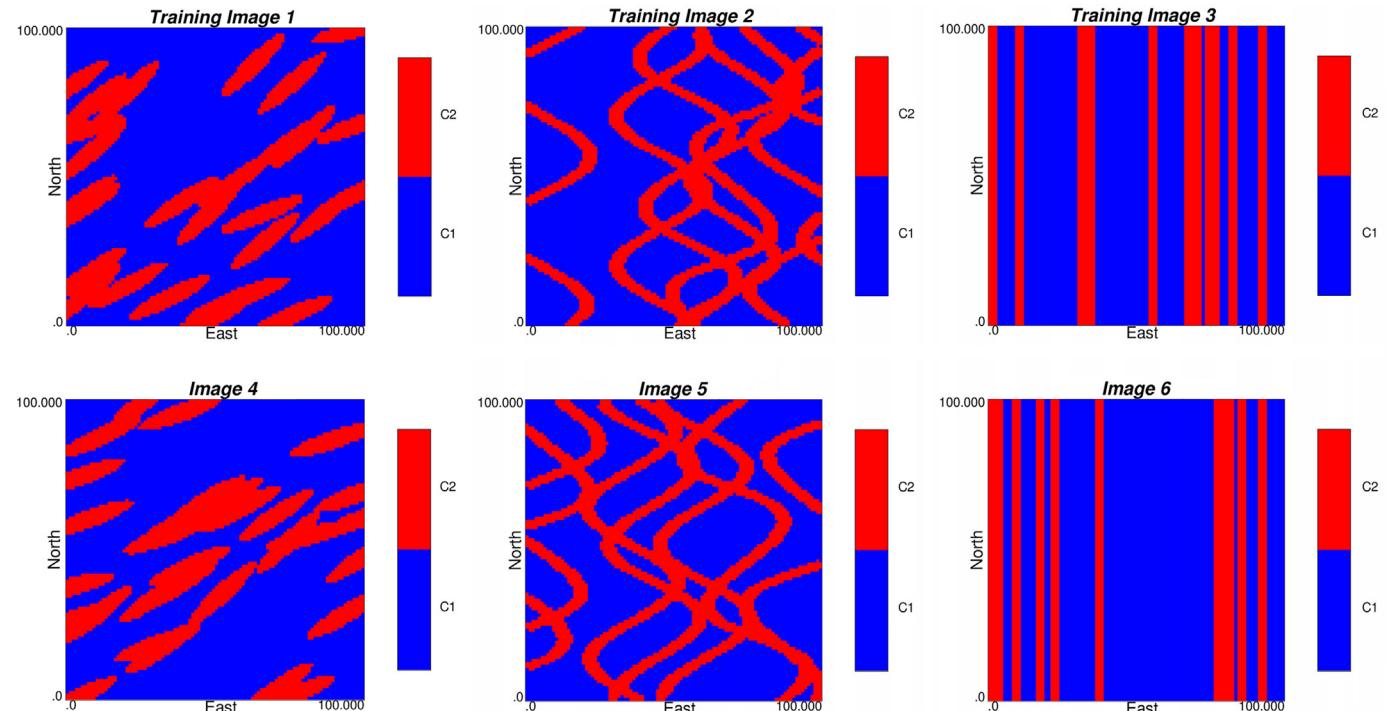


Fig. 2. Binary models used as training images (upper row) and for random conditioning data extraction (lower row) for example 1.

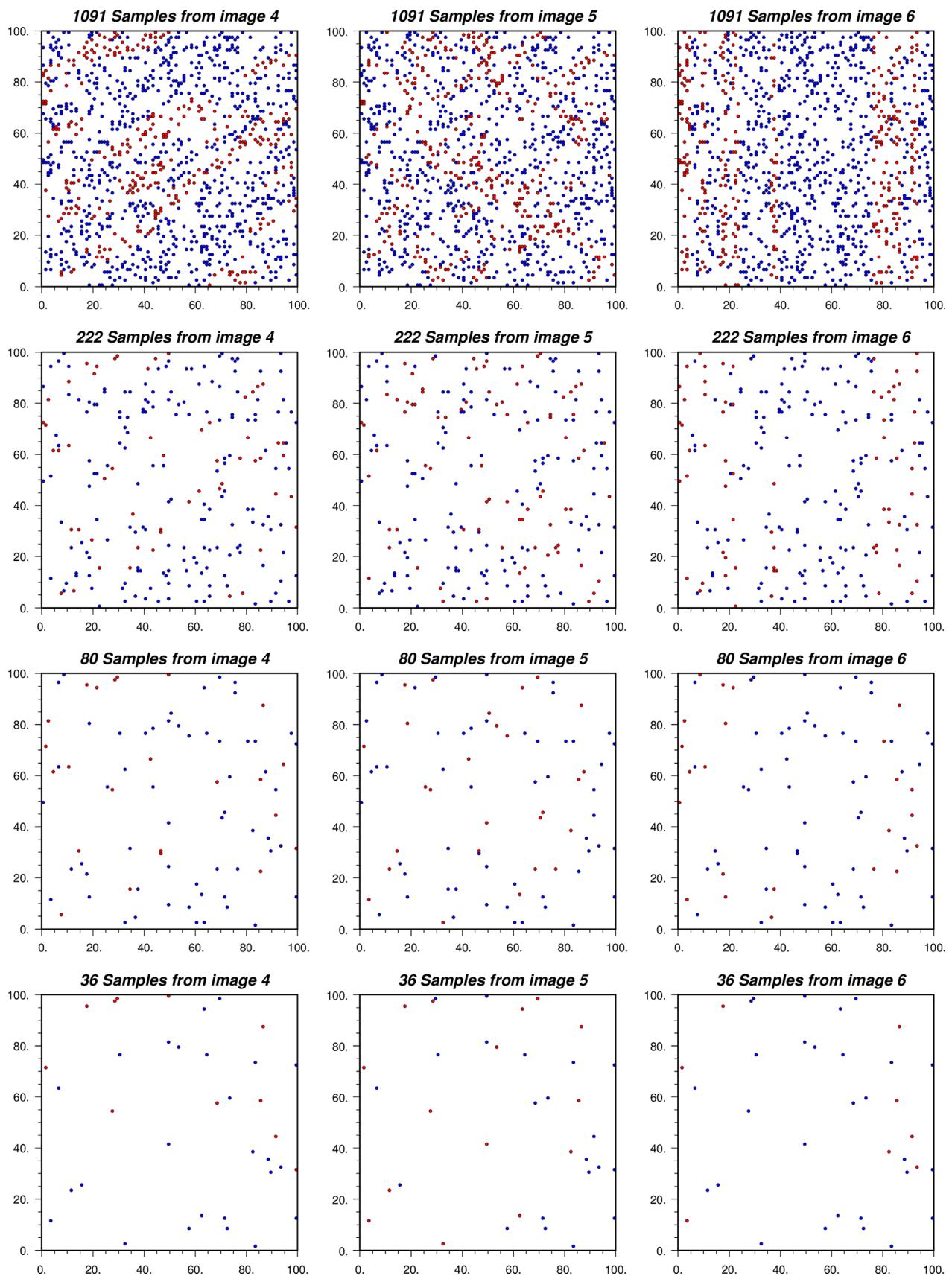


Fig. 3. Randomly selected conditioning data samples for example 1.

do not include smaller order data events when the maximum search radius is reached).

3.1. Example 1

In this example, 3 different parameter sets were used to construct binary models using the TiGenerator tool implemented in SGeMS (Boucher et al., 2010; Maharaja, 2008), as presented in Table 1. All parameters were set such as to obtain the same proportion of 0.3 for facies 1. For each parameter set, 2 models of 100 by 100 nodes were constructed: the first model is used as training image, and the second model is used to extract data (Fig. 2). Since both models are obtained with the same parameters, the data extracted from one model should show a high level of compatibility with the other.

Four sample sets were obtained, consisting of 1091, 222, 80 and 36 random samples, and then used for determining training image

compatibility with all three images corresponding to the first model in each example (Fig. 3).

Compatibilities were obtained considering conditioning events of orders: 1, 2, 3, 4, 5, 10, 15, 20, 25 and 30 (Fig. 4) and using a 25×25 nodes search radius. It is interesting to observe how the relative compatibility values behave for low order (≤ 5) conditioning events. Since the analyzed images have approximately the same facies proportions, the relative compatibility shows no preference for any image when assessing first order conditioning events. As the event order increases, a better compatibility measure with the expected images is observed. This trend is visible even in examples with a small number of conditioning data (bottom row of Fig. 4), although it is less marked in these examples.

This is explained by a poor and non-representative inference of such high order spatial structures in poorly informed scenarios. This insufficient amount of data for proper inference can be

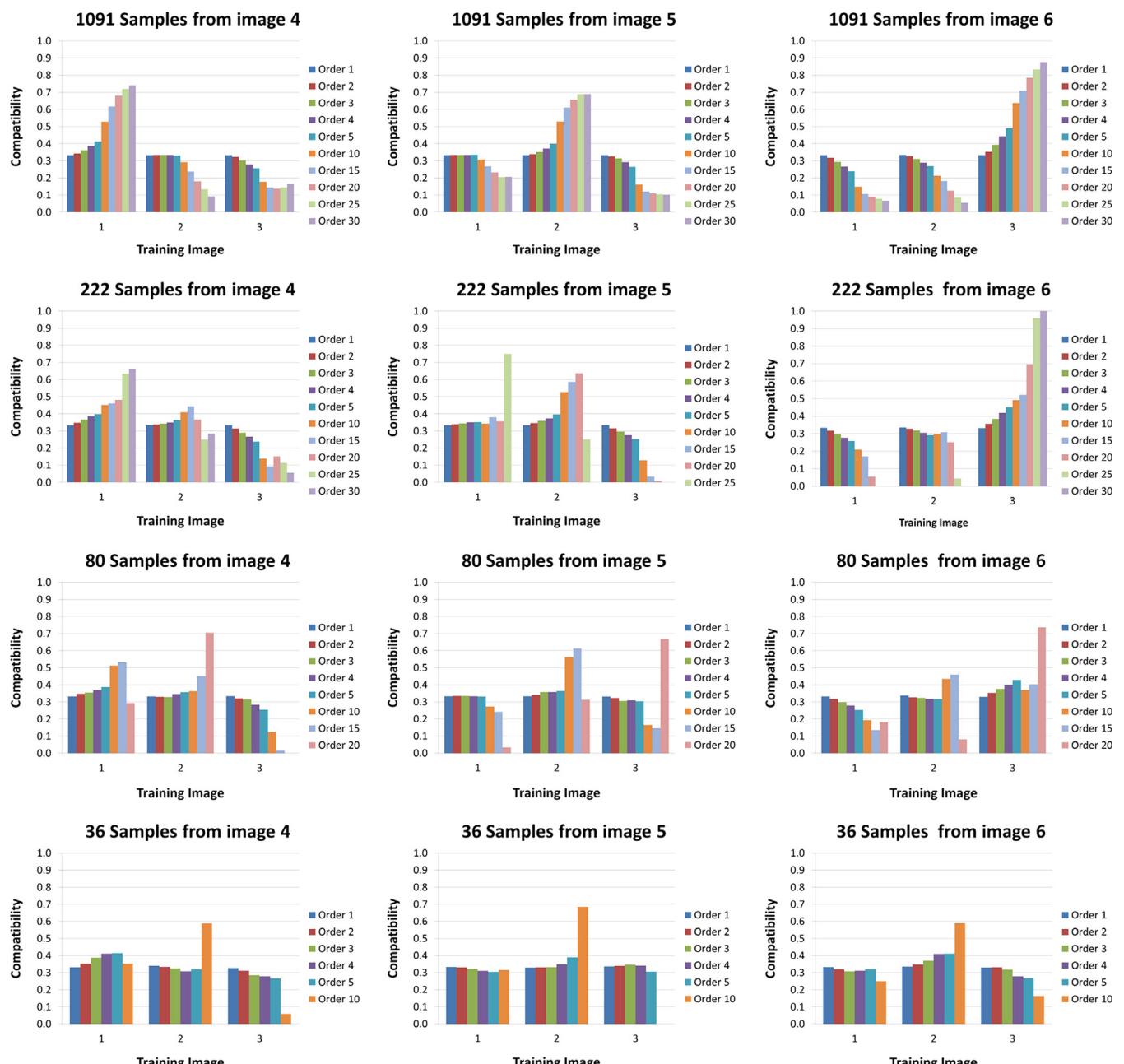


Fig. 4. Relative compatibilities obtained with exhaustive scanning, example 1.

diagnosed by analyzing the change in the amount of patterns of a given order found in the data grid. This value is reported in the output file of line 8. The results of this analysis are shown in Fig. 5.

This allows deciding whether the inference can be considered representative or if the inference is impossible given the amount of available data. In the 222 and 80 samples examples, it is possible to see an important drop in the amount of patterns found when the order of conditioning events rises, while in the 36 samples example no patterns of order 15th or higher can be found given the search neighborhood used. Furthermore, in the 36 data example, no clear differentiation between training images is observed before the results become erratic. This example shows that the selection tool is not able to differentiate effectively training images in scarcely informed scenarios.

The same example is used to test the direct sampling based compatibility measurement algorithm. For each example, the same parameters and inputs considered for the exhaustive scanning method are used. Additionally a training image scan proportion of $f=1$ is used (up to the entire training image can be scanned by the direct sampling). The results are summarized in Fig. 6. It is important to note that parameters required by this method are essentially the same parameters used in the DS multiple-point simulation algorithm (Meerschman et al., 2013).

The results are similar to those obtained with the exhaustive scanning. The main advantage of this method is the speed up of the analysis. A comparison of the computing times is shown in Fig. 7.

The computing time reduction is substantial when the analyzed data event is of a low order. For high order events, the time reduction is not as important, but still significant, especially for high number of samples. The time reduction depends on the frequencies of data patterns in the analyzed training image set. When large scale event occurrences are rare or even nonexistent, the direct sampling method needs to analyze all positions in the images to find a matching occurrence (when the image checking proportion f is set to 1). The order of CPU time gain is not analyzed in this paper, however it can be assured that the number of iterations required by the direct sampling based algorithm to find a match will be at most the number of nodes contained in the training image multiplied by f . This means that in the worst scenario, in which not a single data pattern is contained in the analyzed training image set and the checking proportion f is set to 1, the iterations carried by the direct sampling alternative will be the same as the iterations carried out by the exhaustive scanning algorithm.

The particular behavior of the computing times in the 80 samples example is due to the drop in the number of conditioning events found (see Fig. 5).

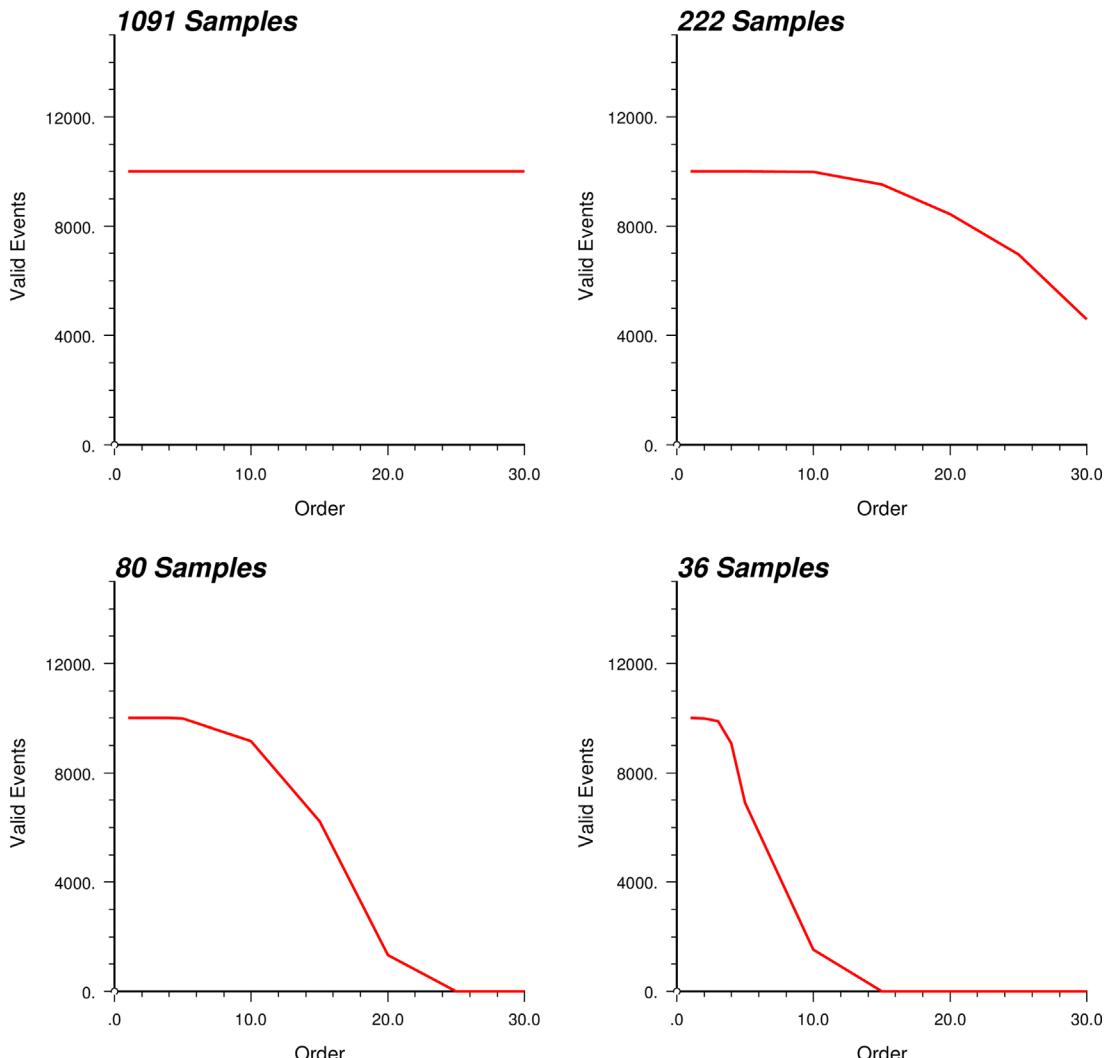


Fig. 5. Valid conditioning events by sample number and event order.

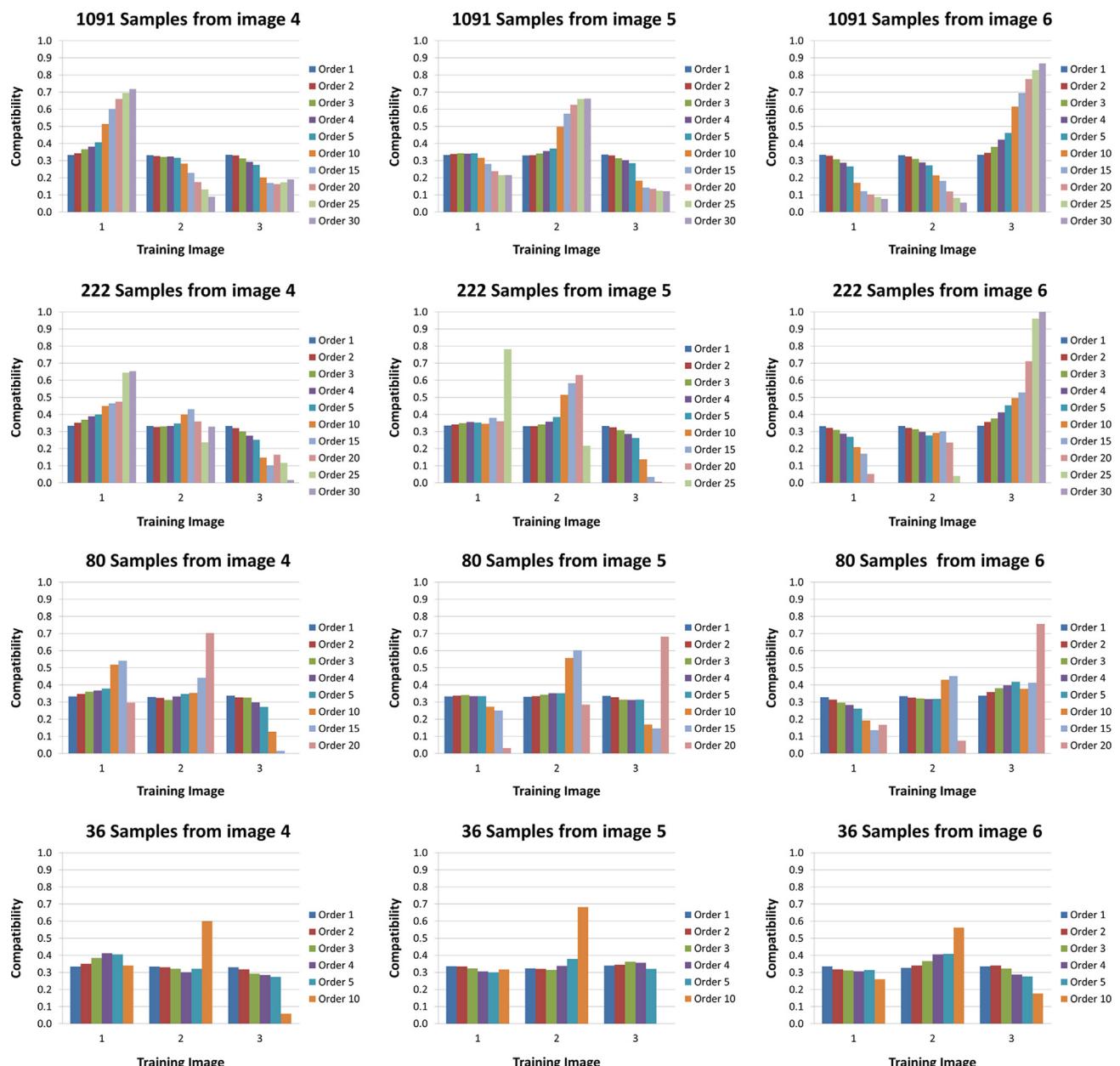


Fig. 6. Relative compatibilities obtained with direct sampling approach, example 1.

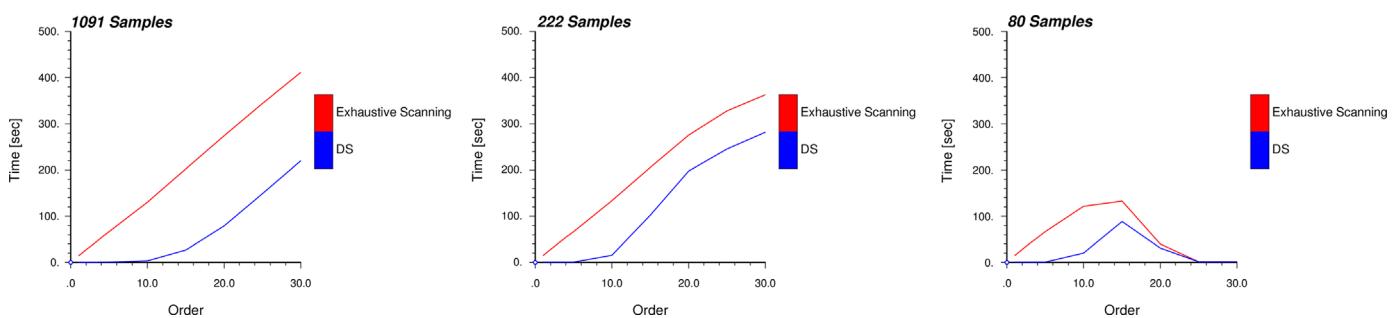


Fig. 7. Time comparison exhaustive scanning versus direct sampling, example 1.

Absolute compatibility is obtained using the same setting. Results regarding the 1091 samples set are summarized in Fig. 8. As expected, higher absolute compatibilities with data sets are

obtained when analyzing the training image with the same spatial characteristics (i.e. samples from TI4 are similar to TI1, samples from TI5 are similar to TI2, samples from TI6 are similar to TI3). It is

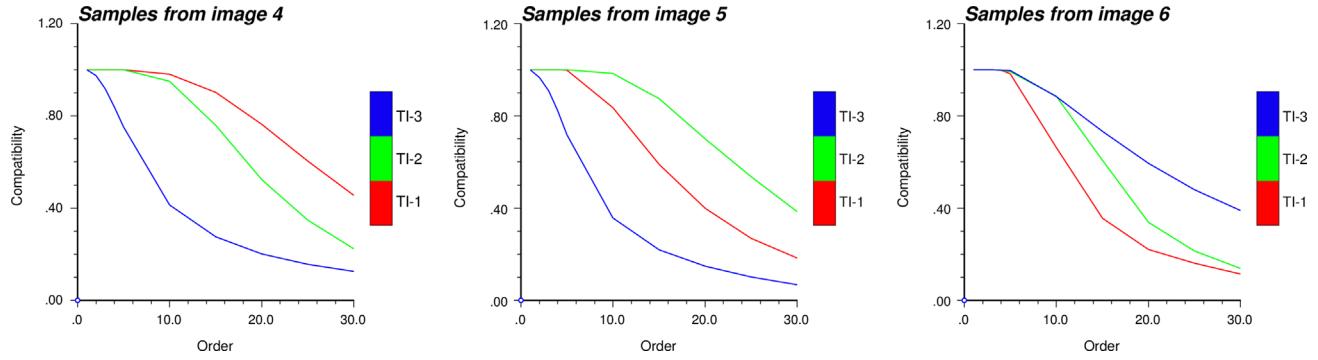


Fig. 8. Absolute compatibilities example 1 for samples coming from training images 4, 5 and 6.

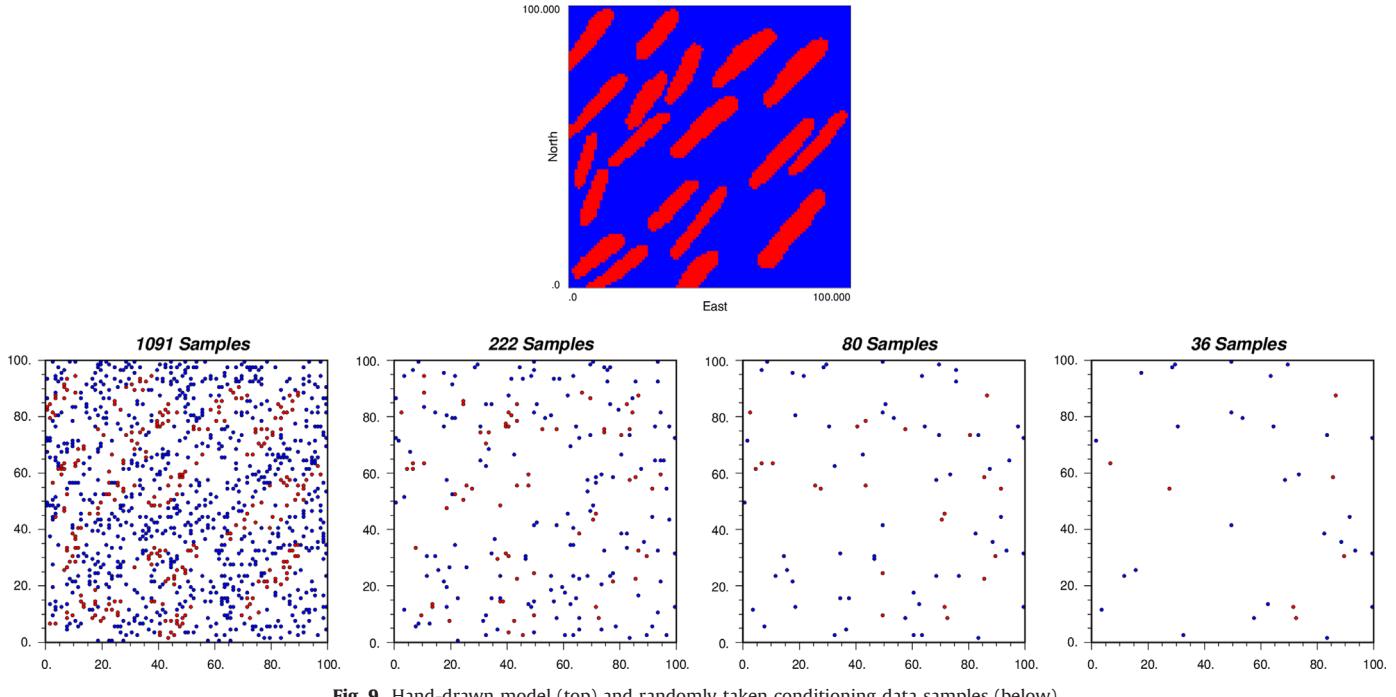


Fig. 9. Hand-drawn model (top) and randomly taken conditioning data samples (below).

important to remember that this compatibility takes a value of 1 if every analyzed conditioning event is contained in the training image. The maximum value is reached for every analyzed image when low order events are considered. However, this indicator should be considered an aid for training image quality measurement only in terms of high order event compatibility, as there are better tools available to measure first and second order spatial compatibility through histogram and variogram comparisons.

The results shown above were obtained by considering training images and models for data extraction, both generated through the use of the same method (TiGenerator). In order to further test the training image selection tool, new samples are extracted from a hand-drawn model (Fig. 9) and relative and absolute compatibilities are computed with the same training images as previously (Fig. 2). Results are shown in Figs. 10 and 11.

It can be observed that the hand-drawn model has similar structural characteristics to those observed in training image 1. Results are as expected, showing little differences for low orders and an increasing differentiation in favor of image 1 when higher order patterns are analyzed. The behavior of relative compatibility results remains erratic when the available conditioning data are scarce and the complexity of the analyzed patterns is high.

3.2. Example 2

The model consists in a realistic alluvial formation obtained with process-based simulation (Chugunova and Hu, 2006; Lopez, 2003), consisting of 1385 layers numbered from zero (bottom) to 1384 (top). The structures in the model, which contains 6 different facies, show a smooth transition from lobes in the lower layers to channelized structures in the upper layers (see Fig. 12). These 2-D layers were used to get training images and conditioning data. The layers selected as training images were chosen every 140 to 150 m in the vertical direction (Fig. 12). Randomly located scattered conditioning data of 1093 and 222 samples were extracted from layers 5 m below the layers used as training images. Each pair consisting of a training image and the associated scattered data are termed a data set, and each data set is attributed a number (Table 2). We use our method to try and assemble the correct pairs conditioning data/training image.

Relative compatibilities between training images and 1093 and 222 samples data sets were calculated using the direct sampling approach. Results are summarized in Fig. 13.

In example 1, similar facies proportions in the training images caused relative compatibilities to become noticeably different only when analyzing conditioning events of higher order. In this

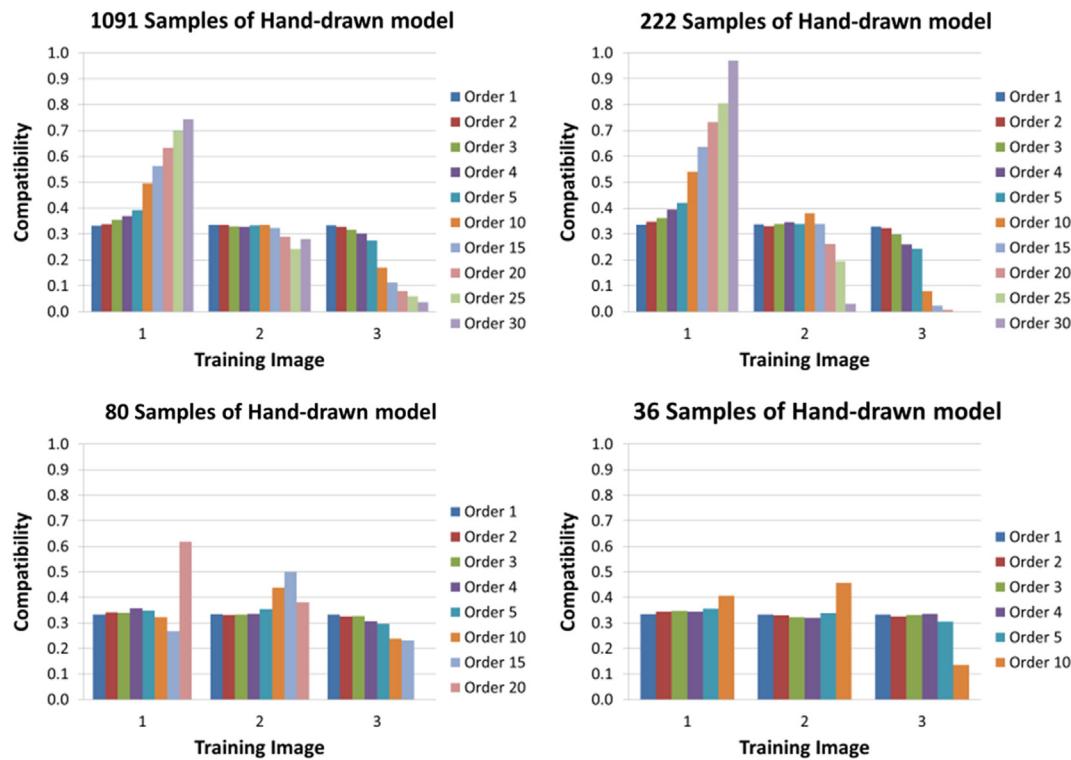


Fig. 10. Relative compatibilities obtained with direct sampling approach, hand-drawn model.

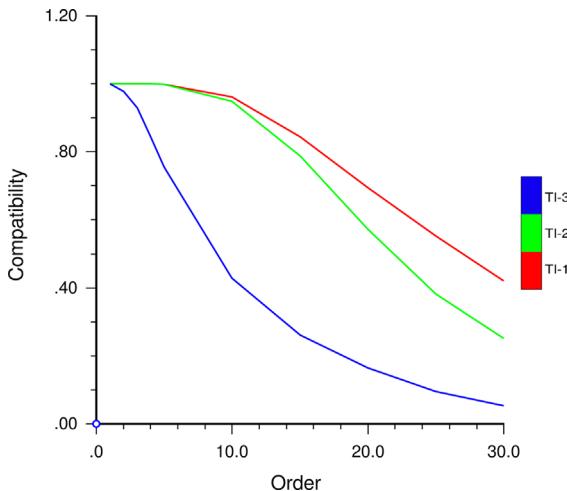


Fig. 11. Absolute compatibilities example 1 for samples coming from hand-drawn model.

example, large differences between images in terms of spatial structures and proportions result in higher variations of relative compatibility, even when low order events are considered and a lower amount of samples is available. Care should be taken when analyzing the results, as these can be driven by the differences in proportions rather than in spatial structure, especially when low order events are considered. Spikes observed in data set #5 can be explained by a low amount of high order conditioning events found.

This example shows that the method allows obtaining the most consistent training image of a 3D vertical non-stationary alluvial formation. This can be particularly useful when modeling non-stationary models by assigning a different training image for each zone of the model (de Vries et al., 2009; Wu et al., 2008).

3.3. Example 3

This example is used to address the applicability of the developed tool on a real mine-scale example. In this example, the relative compatibility of different 3-D geological models of the Escondida Norte deposit with the available drill-hole dataset is measured.

Minera Escondida is the largest copper mining operation in the world reaching an annual production of 1 086 700 t in 2010. The operation is located in northern Chile, 170 [km] southeast of the city of Antofagasta at an altitude of 3100 [m] above mean sea level, and since 2005 the company operates Escondida Norte, the deposit addressed in this example, which is located 5 [km] from the main pit.

Available information corresponds to the current deterministic geological model of the lithologies at Escondida Norte open pit area and to 534 drill-holes. Drill-hole data are composited at 5 [m] length, which means that a total of 26 926 drill-hole data are available. The geological model corresponds to a block model with blocks of $25 \times 25 \times 15 \text{ m}^3$ in size. The location and size of the model are detailed in Table 3. The geological model was constructed through wireframing in order to convert interpreted sections into 3D volumes (Duke and Hanna, 2001). These interpretations were based on surface information and drill-hole information at depth.

The spatial location of the available information and the lithologies codification are shown in Fig. 14.

At the beginning, the available deterministic model is used as training image to build 3 groups of 10 realizations with the Direct Sampling simulation algorithm. Each group is constructed using the different parameter sets shown in Table 4. The simulations are conditioned to all the available drill-hole data information. The non-stationarity of the training image is conserved by MPS realizations due to the amount of available conditioning data. It should be noted that considering MPS realizations as training images is not the approach to be used in practice. Here the

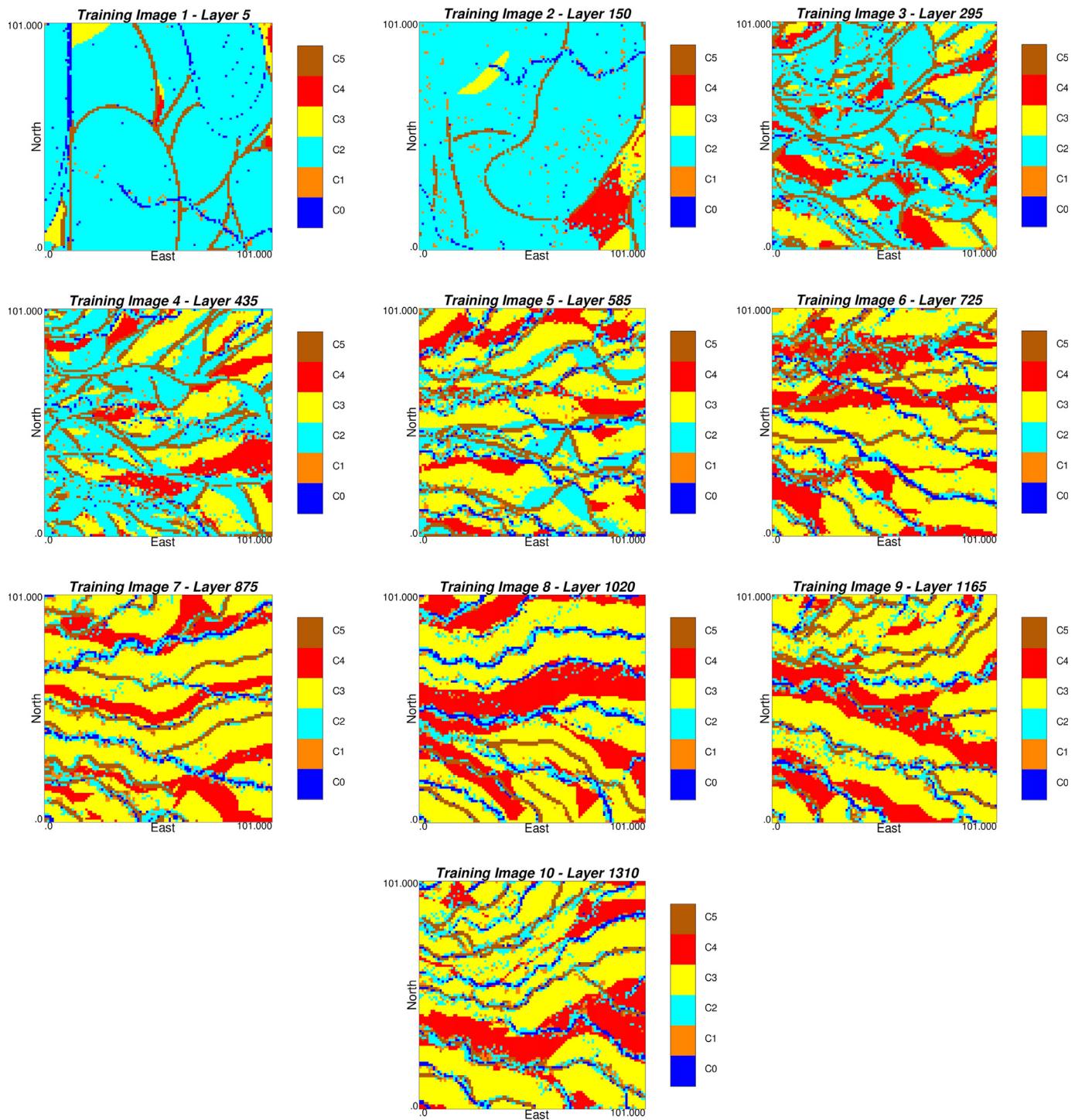


Fig. 12. Layers from synthetic model used as training images for example 2.

Table 2
Summary of sampled and training layers, example 2.

Data set #	Training image layers	Sampled data set layers
1	5	0
2	150	145
3	295	290
4	435	430
5	585	580
6	725	720
7	875	870
8	1020	1015
9	1165	1160
10	1310	1305

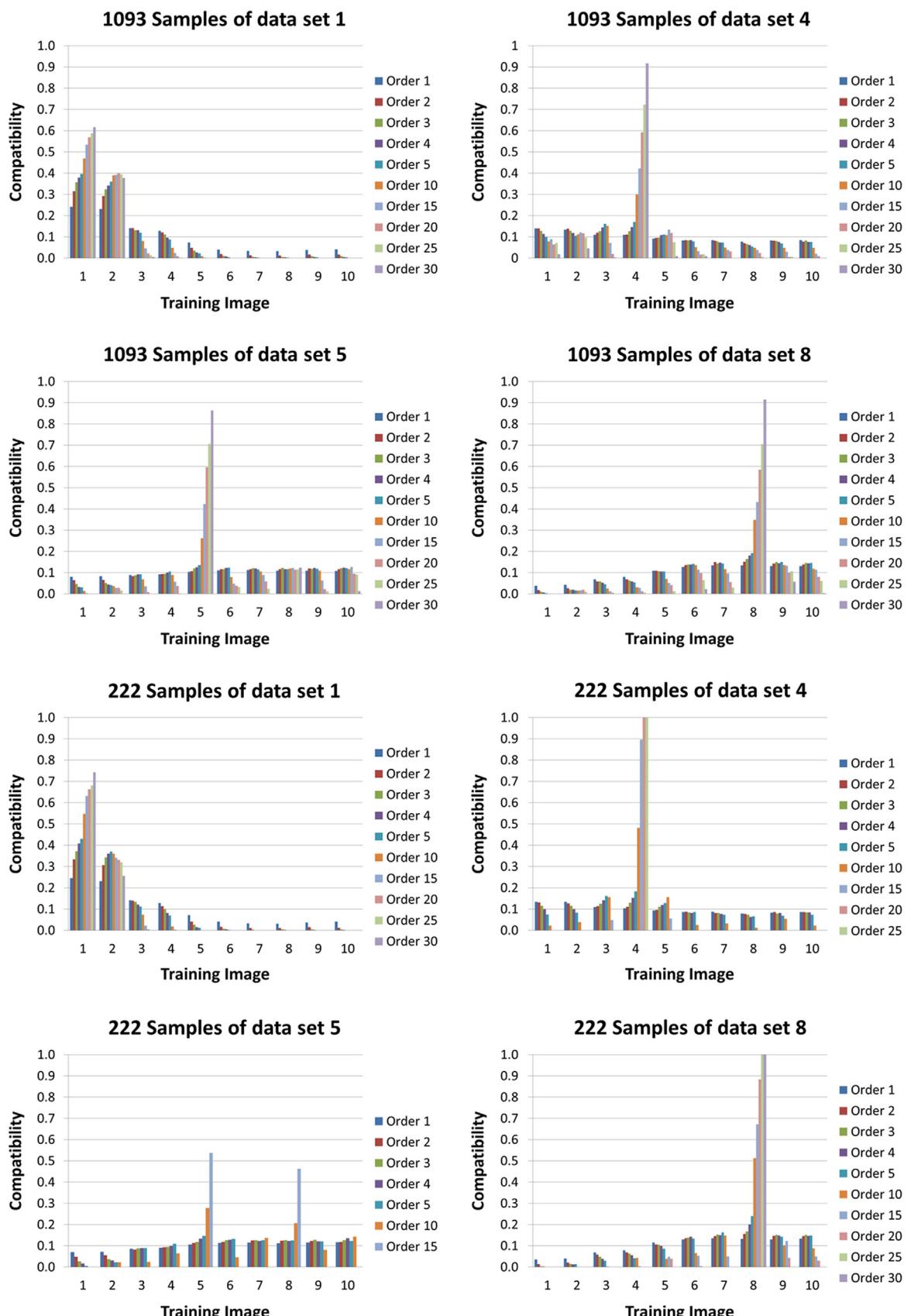


Fig. 13. Relative compatibilities obtained with direct sampling approach, example 2.

objective is only to show the applicability of the proposed method in a real mining 3D case study.

Next, the simulations are grouped 3 by 3, each group consisting of one realization obtained with each parameter set. We have in each group one simulation constructed with a low-distance threshold (good reproduction of the training patterns), one realization constructed with a medium and one realization with a high

distance threshold (poor reproduction of the training patterns). Realizations with each setting are shown in Fig. 15. These groups of realizations are reconsidered as training image sets and used for illustrating the validity of the training image selection tool.

Interesting features of the grouped realizations are obtained through the use of this methodology. Among these, each realization honors the conditioning data, all have the same lithologies distribution, and variograms computed over the different realizations do not vary significantly (see Fig. 16). This means that all these models have a very similar 2-point spatial structure. Therefore, it would not be possible for a modeler to select the best training image of one of these sets based solely on first or second order statistics based criteria. However, the training image selection tool may be able to achieve this goal as the patterns contained in each training image should be different: realizations constructed with a low distance threshold should better reproduce high-order statistics than realizations constructed with higher distance thresholds.

Table 3
Spatial characteristics of the geological model.

Direction	Origin coordinate	Block size [m]	Number of blocks
X	17500	25	90
Y	113050	25	65
Z	2682.5	15	42
		Total below surface	225176

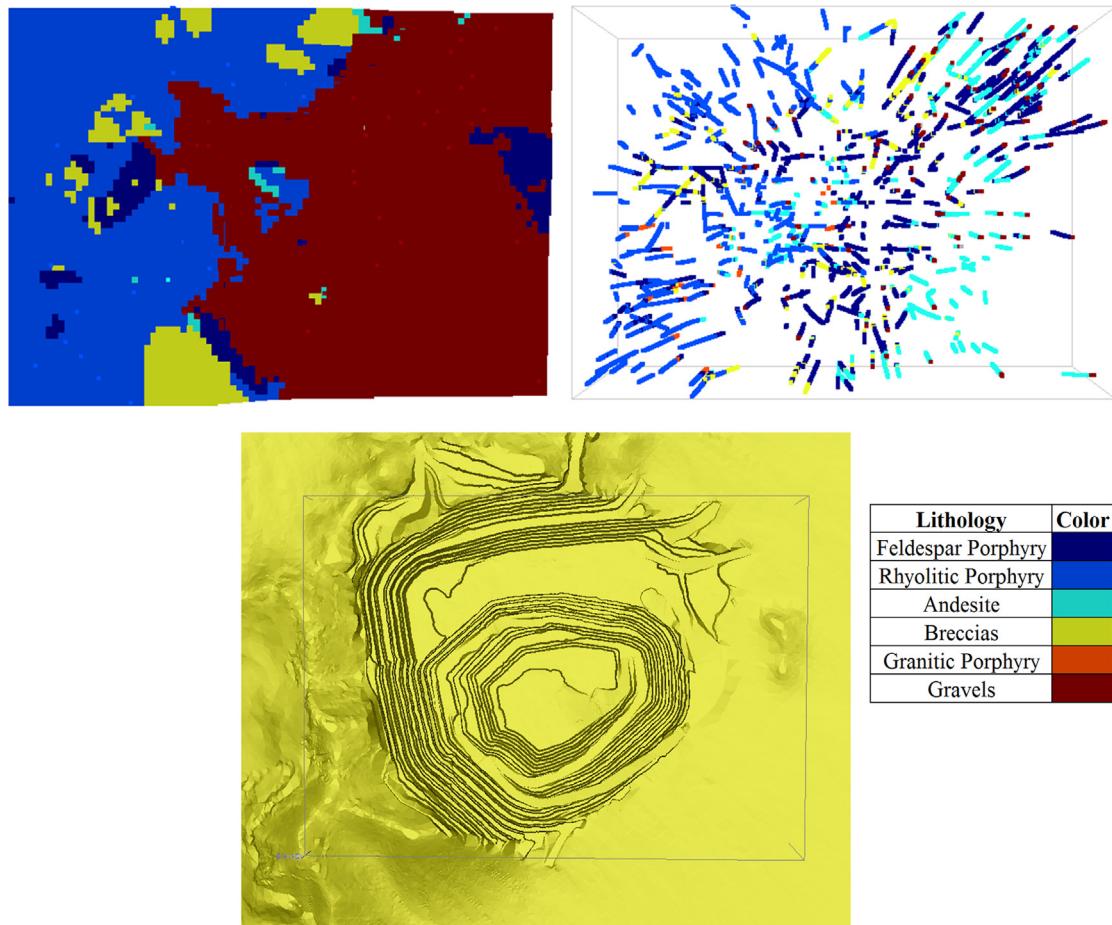


Fig. 14. Geological model (upper left), drill-hole data (upper right) and data location within the Escondida Norte open pit area (lower).

Table 4
DS simulation parameters.

Parameter	High-threshold	Medium-threshold	Low-threshold
Distance threshold	0.2	0.1	0
Maximum number of points in neighborhood	30	30	30
Search radius (x,y)	20 [nodes]	20 [nodes]	20 [nodes]
Search radius (z)	10 [nodes]	10 [nodes]	10 [nodes]
Checked TI fraction	1	1	1

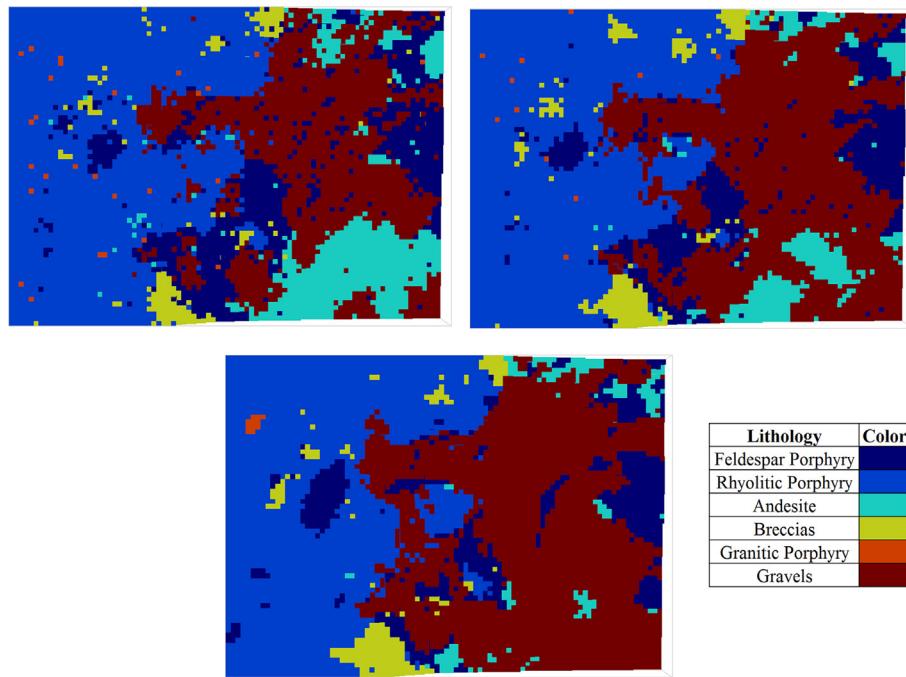


Fig. 15. First TI-set: DS realization 1 with high-distance threshold (upper-left), DS realization 1 with medium-threshold (upper-right) and DS realization 1 with low-distance threshold (lower).

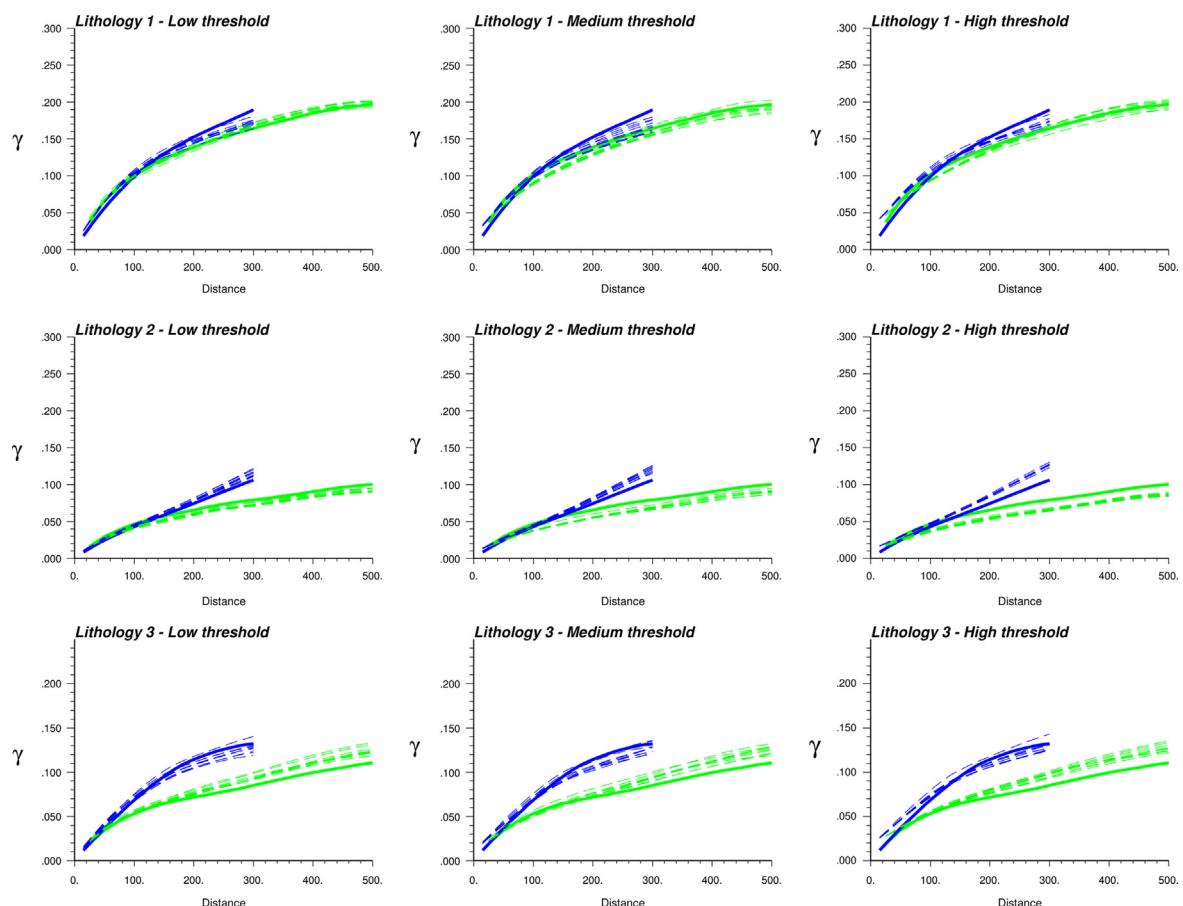


Fig. 16. Indicator variograms of the deterministic geological model of Escondida Norte (thick line) and stochastic realizations in the horizontal (green) and vertical (blue) directions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The parameters used to perform the training image selection are summarized in Table 5. To keep CPU times reasonable, only a direct sampling search strategy with a low training image fraction check is used.

Relative compatibilities were computed for each of the 10 training image sets. Four different conditioning data sets were used to test the method: the first consisted of all the available drill-holes, a second which considers only a randomly selected 50 [%] of the drill-holes, while the third and fourth conditioning data sets consisted of 15 and 5 randomly selected drill-holes. The mean relative compatibilities obtained by image is shown in Fig. 17.

Results show that the training image selection method is able to select the more compatible training image in terms of high-order spatial structure if 15 or more drill-holes are available. As in example 1, when compatibility in terms of low-order spatial structure is analyzed the method shows no particular preference for a given training image since all the images have the same lithologies proportions (1st order) and very similar 2nd order spatial structure (variograms). However, when the order of the analyzed patterns increases, the preference for the expected training image becomes noticeable. This is not clearly observable in the 5 drill-hole data set due to a lack of high-order patterns.

Table 5
Training image selection parameters, example 3.

Parameter	Value
Search radius (x,y)	20 [nodes]
Search radius (z)	10 [nodes]
Tolerance (t)	0
Analyzed fraction of the TI	0.1

This confirms the limitations of the proposed method when the available conditioning information is scarce.

4. Conclusion

This paper discusses two criteria (relative and absolute compatibility) to compare a data set with candidate training images, and presents a corresponding computer code. One advantage of our comparison approach is that it simultaneously considers different sizes and orders of data events.

The relative compatibility measure is especially useful as a training image selection tool in cases when more than a single conceptual model is available as training image. However, it does not allow assessing if all training images are incompatible with the available data. This is palliated by the absolute comparison criterion.

For both relative and absolute comparison criteria, two calculation approaches are tested, exhaustive scanning and direct sampling. Both show similar results, but considerably lower computation times are obtained when using the direct sampling approach, especially when lower order events compatibility is assessed.

The proposed methods are successfully applied in three examples. The second example shows that the method can also be used as an aid to geological interpretation since it allows identifying the origin of data taken at different depths in a vertically non-stationary 3D volume. The third example demonstrates the applicability of the method to assess training image spatial compatibility with data in real cases. In this sense, a proper Bayesian training image inference method could represent a future research direction. The method would consist in proposing alternate training

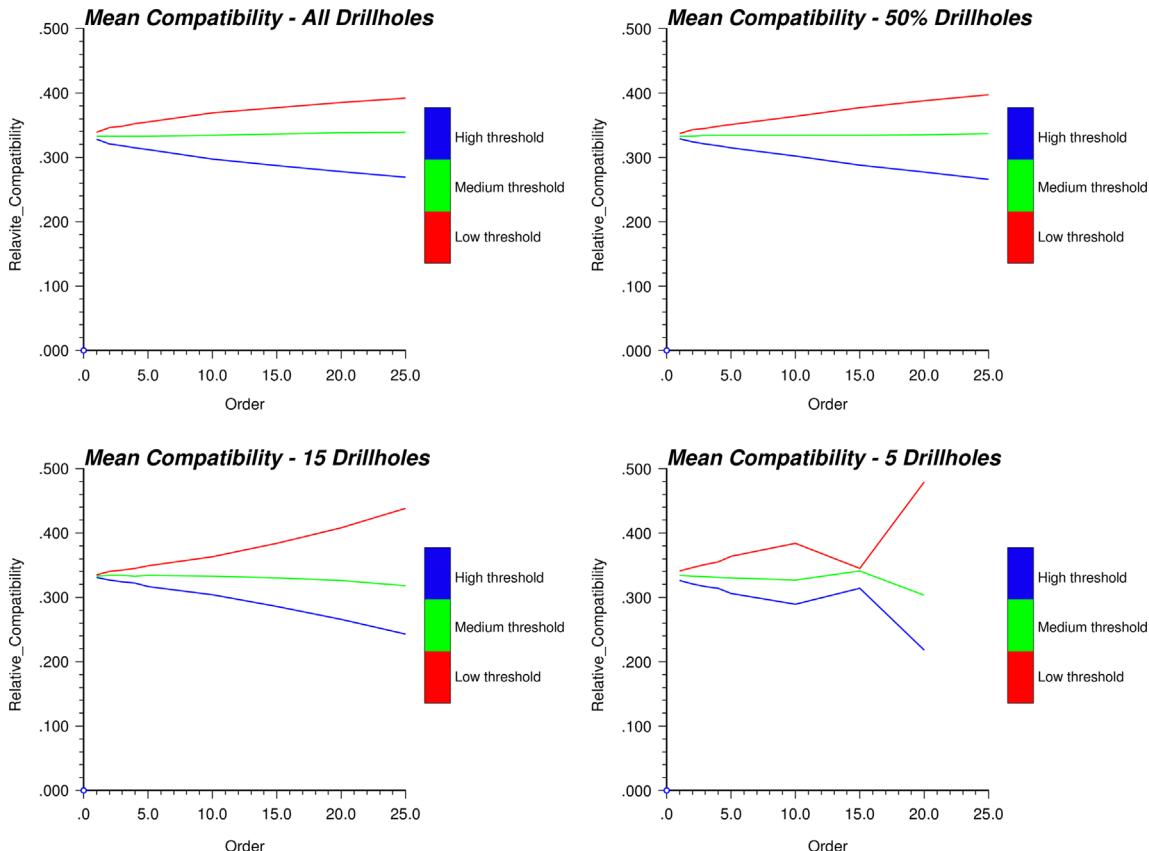


Fig. 17. Mean relative compatibilities, example 3.

images with a prior probability based on geological expertise, and then rejecting them based on compatibility measures.

In addition to offering the ability to verify training image consistency with data, our method could also be used to assess the validity of spatial parameterization in other contexts, for example in object-based-models, to select sets of rules for the placement of objects.

In case of non-stationary MPS simulation, our method is still applicable. When a zonation approach is used with a different training image for each zone, training image consistency should be applied separately for each zone (de Vries et al., 2009). When more complex non-stationarity modeling techniques are used such as involving data events transformation, the comparison should not be based on the training image, but on unconditional realizations (Mariethoz and Kelly, 2011; Strebelle 2002).

Our absolute compatibility metric is presented in a single way (data patterns occurrences are searched in training images), but could also be done looking at the reverse. If every data event is found in the training image, this does not mean that every training image pattern is found in the data. Looking at this reverse problem represents another interesting future research direction, as this measure of compatibility would also consider the completeness of data for the characterization of the training image. If data are scarce and structural patterns complex, this index would be low.

Acknowledgments

The work presented in this paper was financially supported by the National Fund for Science and Technology of Chile (FONDECYT) as part of the Project number 1090056 and by the National Centre for Groundwater Research and Training (Australia). The support of the ALGES laboratory at the Advanced Mining Technology Centre (Chile) is also acknowledged. ALGES would like to thank BHP Billiton Base Metals for allowing the use of the lithology model and data of Escondida Norte for testing the tools.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.cageo.2014.06.001>.

References

- Arpat, B., Caers, J., 2007. Conditional simulations with patterns. *Math. Geol.* 39 (2), 177–203.
- Boisvert, J.B., Pyrcz, M.J., Deutsch, C.V., 2007. Multiple-point statistics for training image selection. *Nat. Resour. Res.* 16 (4), 313–321.
- Boucher, A., Gupta, R., Caers, J., Satija, A., 2010. Tetris: a training image generator for SGeMS, paper presented at Proceedings of 23nd SCRF Annual Affiliates Meeting, Stanford University, May 5–6, 2010.
- Caers, J., 2005. Petroleum Geostatistics. Society of Petroleum Engineers, Richardson p. 88.
- Carle, S.F., Fogg, G.E., 1996. Transition probability-based indicator geostatistics. *Math. Geol.* 28 (4), 453–476.
- Chugunova, T., L. Hu, 2006. Multiple-point statistical simulations constrained by continuous auxiliary data, paper presented at International Association for Mathematical Geology Xlth International Congress, Université de Liège, Belgium.
- de Vries, L., Carrera, J., Falivene, O., Gratacos, O., Slooten, L., 2009. Application of multiple point geostatistics to non-stationary images. *Math. Geosci.* 41 (1), 29–42.
- Deutsch, C., Journel, A., 1992. *GSLIB: Geostatistical Software Library*. Oxford Univ. Press, New York p. 340.
- Dimitrakopoulos, R., Mustapha, H., Gloaguen, E., 2010. High-order statistics of spatial random fields: exploring spatial cumulants for modeling complex non-Gaussian and non-linear phenomena. *Math. Geosci.* 42 (1), 65–99.
- Duke, J.H., Hanna, P.J., 2001. Geological interpretation for resource modelling and estimation. *Mineral Resource and Ore Reserve Estimation-The AusIMM Guide to Good Practice*, 147–156.
- Eskandaridavand, K., 2008. Growthsim: A Complete Framework for Integrating Static and Dynamic Data into Reservoir Models. University of Texas, Austin.
- Gloaguen, E., Dimitrakopoulos, R., 2009. Two-dimensional conditional simulations based on the wavelet decomposition of training images. *Math. Geosci.* 41, 679–701 (6 SPEC. ISS.).
- Guardiano, F., Srivastava, M., 1993. Multivariate geostatistics: beyond bivariate moments. In: Soares, A. (Ed.), *Geostatistics-Troia*. Kluwer Academic, Dordrecht, pp. 133–144.
- Honarkhah, M., Caers, J., 2010. Stochastic simulation of patterns using distance-based pattern modeling. *Math. Geosci.* 42 (5), 487–517.
- Isaaks, E., Srivastava, M., 1989. *An Introduction to Applied Geostatistics*. Oxford University Press, New York.
- Journel, A., Huijbregts, C., 1978. *Mining Geostatistics*. The Blackburn Press, Caldwell, New Jersey p. 600.
- Journel, A., Zhang, T., 2006. The necessity of a multiple-point prior model. *Math. Geol.* 38 (5), 591–610.
- Khodabakhshi, M., Jafarour, B., 2013. A Bayesian mixture-modelling approach for flow-conditioned multiple-point statistical facies simulation from uncertain training images. *Water Resour. Res.* 49 (1), 328–342.
- Kitanidis, P., 1997. *Introduction to Geostatistics: Applications in Hydrogeology*. Cambridge University Press, Cambridge.
- Lopez, S., 2003. *Modélisation de Réservoirs Chenalisés Méandriformes, Approche Génétique et Stochastique*. Paris School of Mines.
- Maharaj, A., 2008. TiGenerator: object-based training image generator. *Comput. Geosci.* 34 (12), 1753–1761.
- Mariethoz, G., Renard, P., Straubhaar, J., 2010. The direct sampling method to perform multiple-point geostatistical simulations. *Water Resour. Res.* 46 (11), <http://dx.doi.org/10.1029/2008WR007621>.
- Mariethoz, G., Kelly, B.F.J., 2011. Modeling complex geological structures with elementary training images and transform-invariant distances. *Water Resour. Res.* 47, 7.
- 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. *Comput. Geosci.* 52, 307–324.
- Ortiz, J.M., Deutsch, C.V., 2004. Indicator simulation accounting for multiple-point statistics. *Math. Geol.* 36 (5), 545–565.
- Ortiz, J.M., Lyster, S., Deutsch, C.V., 2007. Scaling multiple-point statistics to different univariate proportions. *Comput. Geosci.* 33 (2), 191–201.
- Parra, A., Ortiz, J.M., 2011. Adapting a texture synthesis algorithm for conditional multiple point geostatistical simulation. *Stoch. Environ. Res. Risk Assess.* 25 (8), 1101–1111.
- Refsgaard, J.C., Christensen, S., et al., 2012. Review of strategies for handling geological uncertainty in groundwater flow and transport modeling. *Adv. Water Resour.* 36, 36–50.
- Straubhaar, J., Renard, P., Mariethoz, G., Froidevaux, R., Besson, O., 2011. An improved parallel multiple-point algorithm using a list approach. *Math. Geosci.* 43 (3), 305–328, <http://dx.doi.org/10.1007/s11004-011-9328-7>.
- Strebelle, S., 2002. Conditional simulation of complex geological structures using multiple-point statistics. *Math. Geol.* 34 (1), 1–22.
- Suzuki, S., Caers, J., 2008. A distance-based prior model parametrization for constraining solutions of spatial inverse problems. *Math. Geosci.* 40 (4), 445–469.
- Tahmasebi, P., Hezarkhani, A., Sahimi, M., 2012. Multiple-point geostatistical modeling based on the cross-correlation functions. *Comput. Geosc.* 16 (3), 779–797, <http://dx.doi.org/10.1007/s10596-012-9287-1>.
- Wu, J., Boucher, A., Zhang, T., 2008. A SGeMS code for pattern simulation of continuous and categorical variables: FILTERSIM. *Comput. Geosci.* 34 (12), 1863–1876.
- Zhang, T., Switzer, P., Journel, A., 2006. Filter-based classification of training image patterns for spatial simulation. *Math. Geol.* 38 (1), 63–80.
- Zinn, B., Harvey, C., 2003. When good statistical models of aquifer heterogeneity go bad: a comparison of flow, dispersion, and mass transfer in connected and multivariate Gaussian hydraulic conductivity fields. *Water Resour. Res.* 39 (3), WR00146.