Multiple Point Metrics to Assess Categorical Variable Models

Jeff B. Boisvert, Michael J. Pyrcz, and Clayton V. Deutsch

Geostatistical models should be checked to ensure consistency with conditioning data and statistical inputs. These are minimum acceptance criteria. Often the first and second order statistics, such as the histogram and variogram, of simulated geological realizations are compared to the input parameters to check the reasonableness of the simulation implementation. Assessing the reproduction of statistics beyond second order is often not considered because the 'correct' higher order statistics are rarely known. With multiple point simulation (MPS) geostatistical methods, practitioners are now explicitly modeling higher order statistics taken from a training image. This paper explores potential methods for extending minimum acceptance criteria to multiple point statistical comparisons between geostatistical realizations made with multiple point statistical simulation algorithms and the associated training image. The intent is to assess how well the geostatistical models have reproduced the input statistics of the training image; akin to assessing the histogram and variogram reproduction in traditional geostatistics. A number of metrics are presented to compare the input multiple point statistics of the training image with the multiple point statistics of the geostatistical realizations. These metrics are (1) first and second order statistics (2) trends (3) the multiscale histogram (4) the multiple point density function and (5) the missing bins in the multiple point density function. A case study using MPS realizations will be presented to demonstrate the proposed metrics; however, the metrics are not limited to specific MPS realizations. Comparisons could be made between any reference numerical analogue model and any simulated categorical variable model.

Introduction

Geostatistical algorithms are intended to reproduce input statistics within ergodic fluctuations, that is, the statistical fluctuations due to a limited model size relative to spatial correlation range (Deutsch and Journel, 1998). Yet, due to the complexity of geostatistical algorithms, workflows and the associated trade craft; errors may occur. These errors may result in significant bias in reserves estimates, fluid flow or mineral recovery prediction. These errors may include blunders (e.g. inflated correlation due to the application of the same random number seed for primary and secondary variables in collocated cosimulation), poor implementation (e.g. loss of long range correlation due to overly constrained search limits) or algorithm limitations (e.g. unreasonable short range variability inherent to the sequential indicator simulation method). To avoid such errors, it is essential to compare statistics of simulation output against input statistics.

Leuangthong, McLennan and Deutsch (2004) described the minimum acceptance criteria approach to variogram-based geostatistics. They demonstrated the need to check the expectation of first and second order moments, the property distribution functions and the spatial continuity as characterized by the semivariogram. They also propose the use of cross validation and accuracy plots as additional checks of local accuracy and uncertainty models. The goal of this paper is to go beyond the first and second order moments and present metrics that can assess the higher order statistics of geostatistical realizations; therefore, extend minimum acceptance to MPS methods.

The increasing popularity of MPS is due to the recognition that statistics beyond the second order variogram have a significant impact on resource models (Liu and others, 2004, Strebelle, 2002). Traditional geostatistics assumes a maximum entropy multiGaussian model for all statistics beyond the two point variogram. This results in maximum disconnection of high and low values and the reproduction of only linear spatial features (Deutsch and Journel, 1998) and often biased results in subsequent transfer functions such as flow simulation, mine design or reserve calculations. For example, in flow simulation maximum entropy spatial continuity results in the break up of barriers, baffles and conduits, resulting in dispersive flow patterns and biased estimates of sweep efficiency. MPS simulation mitigates the limitations of the multiGuassian distribution by inferring multiple point statistics beyond the variogram from a representative TI; however, if MPS algorithms are to be used to generate geostatistical realizations with the desired MPS, the reproduction of these MPS should be assessed. Literature on assessing the multiple point characteristics of realizations is lacking because of the inherently complex nature of a multiple point statistic. Consider the multiple point density function (Figure 1), often used as the input to MPS simulation algorithms. This density function counts the frequency of a particular facies configuration. The multiple point density function can be used to generate MPS realizations, but due to its high dimensionality it is not easily visualized and it cannot be modeled analytically, rendering it a poor criteria for judging the acceptability of a model. This paper will provide alternative metrics and summary statistics that can be used to compare the multiple point statistics of input training images and the associated realizations. With the greater degree of control of spatial heterogeniety available with multiple point geostatistics, checking these higher order statistics will be even more important to ensure a reasonable reproduction of input statistics.

The proposed metrics are not limited to assessing MPS realizations, they could also be used to (1) compare MPS algorithms with other facies modeling techniques such as SIS or truncated Gaussian simulation, (2) rank realizations, (3) determine the fitness of realizations, (4) help select algorithmic input parameters or (5) assess the MPS characteristics of traditional geostatistical realizations. Previously, a subset of these metrics were applied to test training image and conditioning data consistency (Boisvert, Pyrcz and Deutsch, 2007).

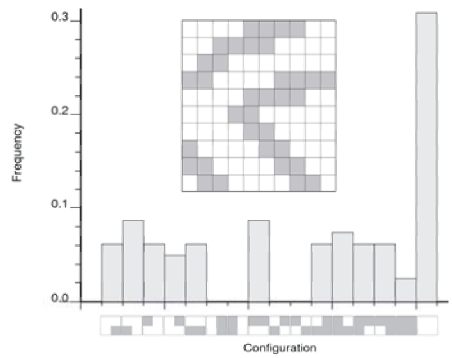

Figure 1: A multiple point density function of the example TI shown. Four point configuration with two indicators.

Background

Recognition of the limitations of the multiGuassian distribution lead to the initial work on MPS simulation (Deutsch, 1992 and Guardiano and Srivastava, 1993). Later, the development of a practical MPS simulation algorithm, SNESIM (Strebelle, 2002), led to the wide adoption of MPS simulation for reservoir facies simulation. Numerous published case studies are available to demonstrate the practice and advantages of MPS simulation (Strebelle, Payrazyan, and Caers, 2003, Caers, Strebelle, and Payrazyan, 2003, Harding and others, 2004, Liu and others, 2004 to name a few).

Multiple Point Statistics

MPS consider the relationship between more than two locations. MPS techniques are often compared to two point statistics. Two point statistics, such as the variogram, quantify linear spatial continuity, while MPS allow for the reproduction of curvilinear heterogeneity patterns and data ordering relationships.

The multiple point density function is commonly used in MPS algorithms. The traditional histogram counts the frequency of times a particular continuous variable falls in a bin, or counts the frequency of times that a particular categorical variable occurs. The multiple point density function counts the frequency of times a multiple point configuration occurs; in other words, the conditional probability of an outcome given the specific categories at surrounding locations. Consider a 4 point configuration that could take two different values, there are a total of 16 (2^4) unique configurations. The frequency of each configuration in the image constitutes the multiple point density function (Figure 1). Note that the ordering of the bins (Figure 1) is arbitrary, only order consistency is required. Considering a larger template or more categories significantly increases the number of bins in the multiple point histogram (number of bins is equal to the number of categories to the power of the number of points), making visualizing the multiple point histogram for large templates challenging to impossible.

Using Multiple Point Statistics

Multiple point histograms can be calculated along wells, but in general, a training image is required that provides a quantification of the patterns likely to be encountered.

Techniques available for MPS include: using the single normal equations (Guardiano and Srivastava, 1993 and Strebelle, 2002); using simulated annealing with the multiple point histogram as the objective function (Deutsch, 1992); updating conditional distributions with multiple point statistics (Ortiz and Deutsch, 2004); training neural networks on training images (Caers, 2001); using a GIBBS sampler (Lyster and Deutsch, 2008); and pattern-based approaches (Eskandari 2008, Arpat and Caers, 2007 or Wu, Zhang and Journel, 2008).

Methodology

The true distribution at any geologic site will remain unknown and it is impossible to completely validate or verify a numerical geological model (Oreskes, Shrader-Frechette and Belitz, 1994). Nevertheless, geological models can be subjected to a series of tests that increase their credibility rather than verify their correctness, this is deemed model *confirmation* (Oreskes, Shrader-Frechette and Belitz, 1994). A basic model confirmation would be a check on the first and second order statistical properties of the models (Leuangthong, McLennan and Deutsch, 2004). Models failing to reproduce the basic histogram/variogram should be rejected, unless there is an application specific reason to expect such a departure (i.e. biased sampling).

The primary confirmation of any geologic model is a visual assessment. In the case of MPS models the visual inspection should be an examination of artifacts specific to the algorithm used and for the reproduction of large/small scale features if appropriate. Multiple point statistic based techniques are implemented because there is a belief that non-linear features, such as channels or complex folding patterns, and ordering relationships are important to the transfer function. Frequently, the assessment of the reproduction of non-linear features and ordering relationships is best performed with a qualitative visual inspection rather than some type of quantitative summary statistic or multiple point statistic metric. This is the first level of model confirmation for MPS models.

Model confirmation for MPS based techniques can also include higher order tests because of access to the correct MPS in the form of the TI. High order statistics are explicitly contained in the TI and *should* be reproduced in the MPS realizations. Visualization and a direct comparison of input and output high order statistics is not possible, but the proposed metrics provide a relative measure of high order statistical reproduction from the TI.

Reasons for not reproducing the exact MPS of the TI include; (1) realizations are often conditioned to data which imposes a constraint on the possible multiple point statistics of the realizations, (2) MPS algorithms are often non-exact and even iterative and may become trapped in a sub-optimum solution and (3) ergodic fluctuations. It is necessary to assess how well the high order statistics are reproduced in MPS realizations; such tests lend support to the use of the numerical simulation results.

Seven metrics will be presented to assess the reproduction of input statistics in multiple point realizations: (1) first and second order checks (2) global trends (3) local trends (4) multiscale histograms (5) behavior of distribution tails (6) the multiple point density function (7) missing bins in the multiple point density function. The specifics of each metric will be discussed with the following case study.

The Case Study

The proposed metrics will be demonstrated with a simple case study. For this study a TI was generated by a simple binary object-based training image generator with channel objects in background overbank. Channel objects with concave upward bases and flat tops where generated with stochastic geometries (width 40 m, width to depth ratio 10:1 and sinuosity 1.6 parameterized by constant amplitude and wavelength. These channels were positioned by a Poisson point process until the global proportion of 27% was matched in a model with extents 2,000 m x 1,000 m and 80 m vertically. The grid cells are 10 m in the horizontal and 1 m in the vertical. Unconditional geostatistical realizations were generated on the same grid used for the training image with the version of SNESIM presented in Strebelle (2007). Conditional realizations could have been generated, but for demonstrating the methods for checking the multiple point characteristics, unconditional realizations are sufficient.

To assess the proposed metrics, 10 realizations were generated using various multiple point template sizes (1, 2, 8, 16, 24, and 56) in the simulation. A simple trend model was constructed with polygons and a filter as described by Strebelle (2007). The trend is intended to represent a fair way with greater concentration of channel elements. The global average of the trend model was set to the target channel proportion in the MPS realizations to ensure consistency. This resulted in MPS realizations with varying consistency with the training image (Figure 2).

These realizations are checked against the input training image. The following metrics are applied; (1) first and second order statistics, (2) trends, (3) multiscale histogram, (4) multiple point density function and (5) missing bins in the multiple point density function.

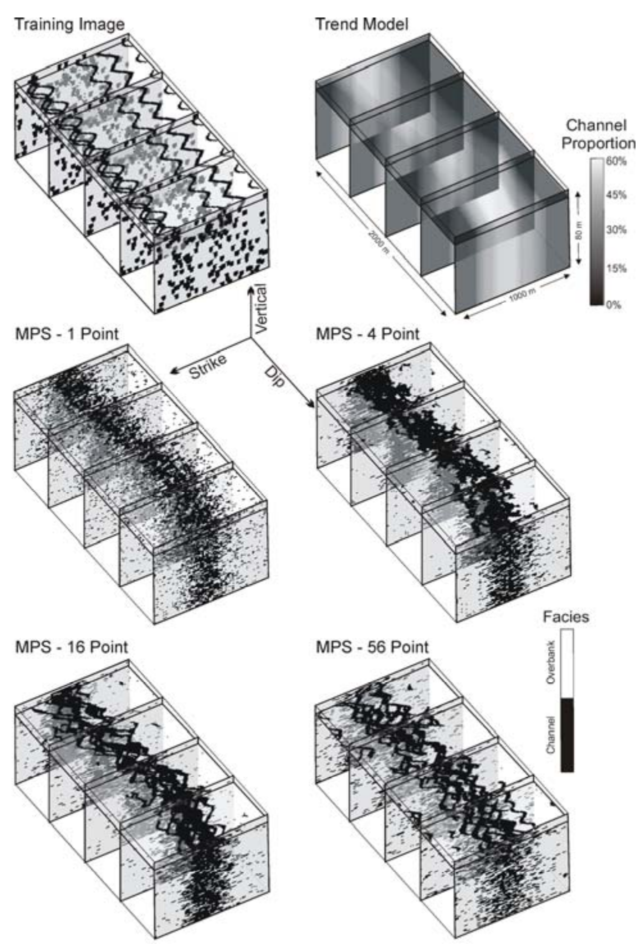

Figure 2: Visualizations of the input training image, trend model and realizations with 1, 4, 16 or 56 point templates, excluding the origin. All models are identically sized. TI and realizations contain only two facies.

First and Second Order Statistics

While the goal of MPS simulation is to reproduce higher order statistics, the lower order checks proposed by Leuangthong, McLennan, and Deutsch (2004) should still be preformed:

- Ensure that the exact data values are reproduced at data locations.
- Histogram and variogram reproduction. Within acceptable ergodic fluctuations these statistics should be well reproduced.
- Cross validation and accuracy plots to check local accuracy and uncertainty model.

The purpose of this paper is to supplement (not replace) these basic tools with metrics that can better assess the multiple point characteristics of geostatistical models.

Trends

All geological modeling techniques assume a form of stationarity. This allows the practitioner to 'pool' their data and infer meaningful statistics for modeling. In the context of MPS, the multiple point density function is often assumed stationary that is, the multiple point density function of the TI is assumed to apply everywhere in the modeling domain. As in traditional geostatistical techniques, a trend can be used to enforce non-stationary proportions of specific facies that cannot be captured in the TI. If a trend is used in geostatistical modeling, its reproduction in subsequent realizations should be checked. The reproduction of a trend will be assessed in two ways; (1) globally for a single realization or (2) locally over a set of realizations.

Evaluating the reproduction of the trend for a single realization ensures that the trend is reproduced in expected terms over the realization without concern for local accuracy. For example, all locations where the trend map indicates a proportion between 0.2-0.3 can be examined (Figure 3). For a single realization the proportion in the shaded area should be about 0.25. The expected trend proportion from the trend model can be compared to the actual proportion in the realization (Figure 4). Using a different multiple point template in SNESIM generates a range of results. Using 2, 4, 8, 16 or 24 points seems to generate models where too much emphasis is placed on the trend; where the trend proportion is high the realization proportion is even higher, where the trend proportion is low the realization proportion is even lower. A sufficiently large template (in this case at least 32 points) must be selected for reasonable trend reproduction. Alternatively, the trend may not be given sufficient weight (Figure 5).

Assessing the local trend ensures that on average the realizations honor a local trend at a specific location. The average proportion over all realizations at each location is compared to the actual trend proportion (Figure 6). A randomized plot close to zero is desirable, as shown by the realizations with MP templates larger than 16 points. Areas that are consistently high or low do not reproduce the trend well (upper plots on Figure 6).

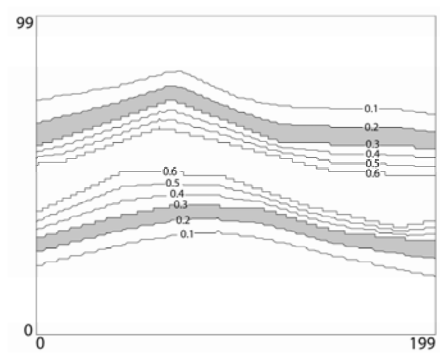

Figure 3: Proportion map. The shaded region indicates areas where the realization proportions should total between 0.2-0.3.

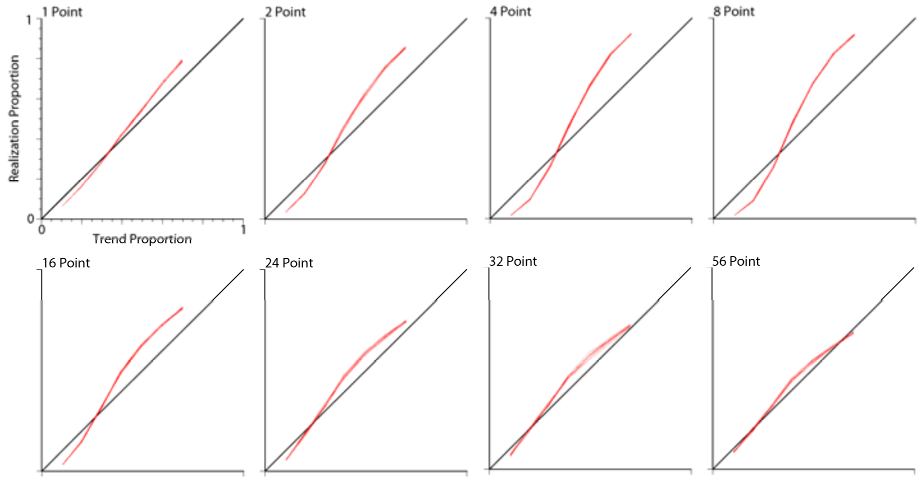

Figure 4: Global trend check. Each plot represents a different set of 10 realizations (lines) using 1,2,4,8,16,24,32 and 56 point statistics in SNESIM. All axis are identical and indicated in the first plot.

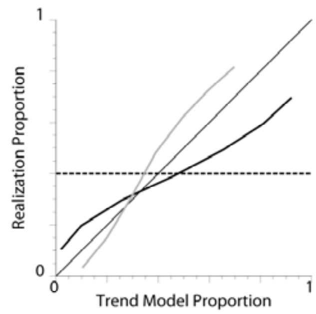

Figure 5: Solid black line – trend is not receiving enough weight. Solid gray line – trend is receiving too much weight. Dashed black line – trend is ignored, global proportion is used for generating the realizations.

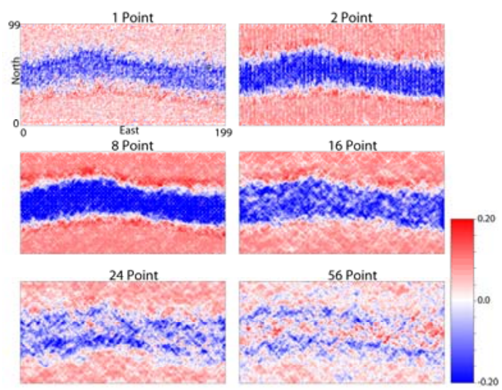

Figure 6: Local trend check. Plotted is the (*trend map*) – (*average realization proportion*).

Multiscale Histograms

Geostatistical realizations are commonly used to generate an average value at a scale larger than the modeled block size (e.g. selective mining unit in mineral resources and upscaled flow simulation grid in reservoirs). In this application the behavior of the realizations as they are scaled to the desired volumetric support is critical. One possibility is to visually assess and compare the behavior of the scaled histogram of the TI to that of the realizations. Ideally, the histograms of the TI and the MPS realization at different scales would be identical.

The behavior of the multiscale histogram is dependent on the underlying spatial structure of the training images and realizations. In this case study the upscaled realizations average to the global proportions quicker than the TI (Figure 7). This indicates more structure (i.e. connected channels) in the TI than in the MPS realizations. Less structured models quickly scale to the global average/proportion (Figure 7). More structured models will remain bimodal at larger scales.

The behavior of the tails of the multiscale distribution is critical to resource calculation. How these tails change as the support of the histogram increases is a function of the multiple point distribution of the realization. If the tails in the MPS realizations *disappear* or average out at a smaller volume than the TI there is too little structure in the realizations (Figure 8). Specifically, this indicates that the highs and/or lows in the MPS realizations are too disconnected and the practitioner should assess if this will affect the desired upscaled properties. Conversely, the realizations may contain additional structure, in which case the tails of the TI will disappear more quickly. In this case the practitioner should examine the cause of the increased connectivity of extremes in the MPS realizations and again assess the importance. This could be due to a trend model that does not appear in the TI or unwarranted structure added by the simulation algorithm. MPS simulation parameters (e.g. template size) can be adjusted to better match the scaling properties of the TI.

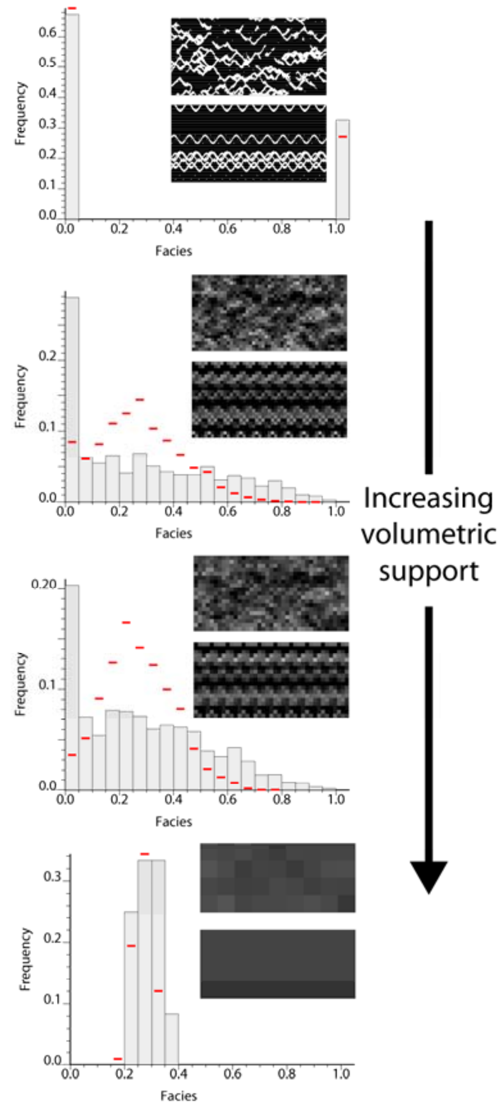

Figure 7: Histogram changing with block averaging. The main histogram represents the TI. A 16 multiple point simulation scaled histogram is shown as horizontal dashes. Overlain are slices for each volumetric support for the TI (below in each histogram) and MPS realization (above in each histogram).

Multiple Point Density Function

Akin to checking the variogram reproduction with traditional geostatistical approaches the reproduction of the multiple point density function should be assessed.

Checking the multiple point density function is complicated by its high dimensionality; even small templates are difficult to visualize because of the large number of bins. Moreover, the lack of an intuitive ordering to the MPS bins negates any reduction in dimensionality (Figure 1). A solution to this problem was proposed by Boisvert, Pyrcz and Deutsch (2007) where a difference measure is calculated between two multiple point density functions. This can be applied to compare MPS realizations to the input TI (Figure 9). The difference measure is the sum of the absolute difference between each bin in the multiple point density function. Ideally this would be near zero, indicating similar multiple point histograms. This metric can be used to tune input parameters; for example, in the case study using more than 24 points in the MPS simulation does not provide significant gains based on this metric (Figure 9).

While this comparison is useful for assessing multiple point characteristics, it is important to realize this difference measure reduces a very high order statistic to a single value. Much information contained in the multiple point density function is lost.

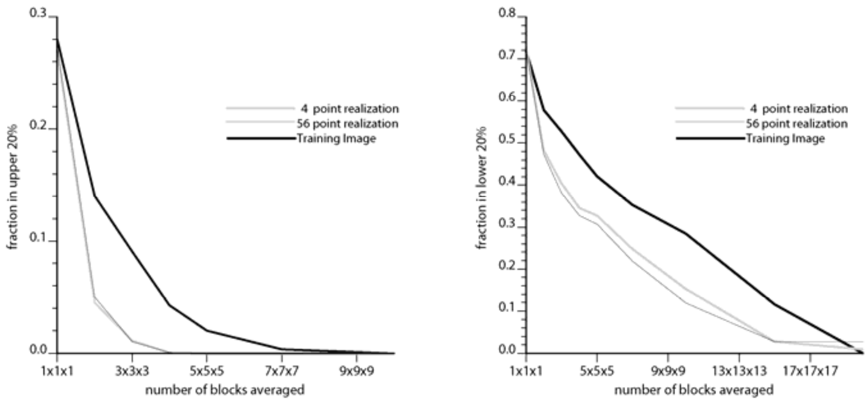

Figure 8: Checking the fraction in the tails when block averaging. The TI and realizations are binary; therefore, the tails represent the only indicators present.

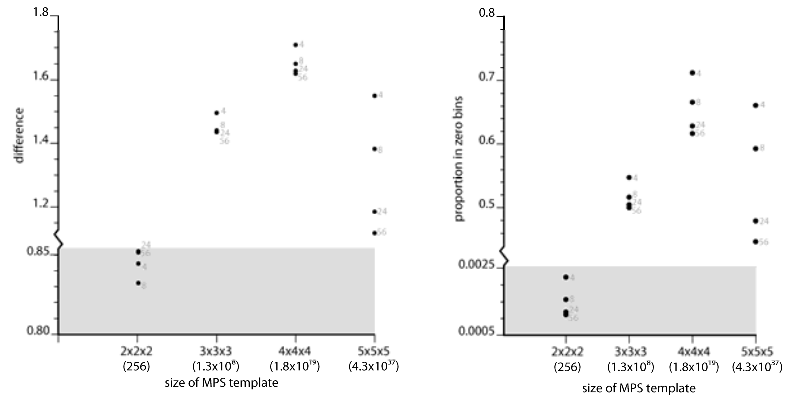

Figure 9: Left: Sum of the absolute difference between the multiple point density function of the TI and the realizations. Right: Proportion of implausible configurations. Points represent a realization using either 4, 8, 24 or 56 MP statistics. The number of bins in the MP density function are plotted in brackets on the x-axis to emphasize the nonlinearity of the scale. Note the broken scale to show the 2x2x2 configuration.

Missing or Zero Bins

Of concern in a practical MPS application are the multiple point bins that have zero frequency. An assumption is made that the TI is representative of the modeling domain; therefore, if a configuration does not exist in the TI it is implausible and should not exist in the realizations. In practice, these implausible configurations do exist in MPS realizations. Consider the significant proportion of implausible configurations found the case study for large template sizes (Figure 9). This check should be preformed for each MPS realization, with preference given to realizations that have a lower proportion.

Discussion

This paper has presented seven metrics to assess the multiple point statistical characteristics of geostatistical realizations (Table 1). Each metric attempts to summarize a high level statistical relationship in a way that can be compared, either quantitatively or qualitatively, to a reference statistic provided by the TI. Not all metrics will be useful in all applications. Poor performance of a model on one metric is not reason to reject the model if the metric is not relevant to the intended purpose of the model; however, good performance on any metric increases the credibility of the model and a failure should raise suspicions as to the models appropriateness.

Table 1: Summary of assessment metrics.

Measure	Pros	Cons	Appropriateness
First and second order statistics	-easily inferable from data -well understood metrics	-only captures linear correlation between data	-should always be checked
Global trend	-ensures algorithm is reproducing the trend globally	- no measure of local accuracy	-should be checked if using a trend
Local trend	-ensures algorithm is reproducing the trend locally	- require a significant number of realizations to check	-should be checked if using a trend
Multiscale histogram	-important for upscale calculations of reserves -can be used to tune algorithmic parameters	-not explicitly reproduced with typical simulation algorithms -qualitative comparison of spatial structure at multiple scales	-use if upscaled properties are of interest
Fraction in tails	-a good summary of the multiscale histogram -can be used to tune algorithmic specific parameters -important for many resource volume calculations	-may be influenced by non-stationarity (trends)	-use if upscaled properties are of interest
Multiple point density function difference	-summarizes the multiple point density function to a single value -can be used to tune algorithmic specific parameters	-some critical information may be lost due to oversimplification	- use for a relative comparison of multiple implementations
Missing or zero bins	-can be used to tune algorithmic specific parameters	-zero bins may actually be plausible but not represented in the TI	-zero bins should not appear in realizations

MPS model confirmation cannot be addressed without discussing the inference of input statistics. All metrics discussed in this paper compare the models to the input statistics (e.g. global proportions, trend and TI), assuming appropriate input statistics. Selection of an inappropriate TI will produce inappropriate geological models and such errors will *not* be discovered by the proposed metrics. Pyrcz and others (2006) discuss methods for inference of representative input statistics (one and two point and trends) and Pyrcz, Deutsch and Boisvert (2007) attempted to provide objective measures to select a TIs for use in MPS simulation; however, literature on this topic is sparse. The use of MPS in geological modeling is in its infancy and it is likely that as the field matures more emphasis will be placed on practical issues such as TI selection, rather than the current focus on MPS algorithmic development. It must be realized that the proposed metrics do not lend credibility to the choice of input statistics, the histogram, semivariogram, TI or trend model, rather they lend confirmation to the ability of the MPS algorithm to reproduce the desired statistics in the final geological models.

Conclusions

The place of the proposed metrics in the overall model confirmation process should be explicitly stated. A visual inspection of the model is required and histogram and semivariogram reproduction are necessary checks. The appropriate metrics (Table 1) should then be applied to assess the reproduction of higher order multiple point statistics. In the face of significant deviation from the input statistics, adjustments to the implementation workflow and parameters should be made and contradiction between input statistics and conditioning should be addressed. Finally, a more intensive assessment of the models credibility should be undertaken using accepted methods such as hold out analysis (Davis, 1987), performance measuring or history matching. The proposed metrics are one step in model confirmation as they are intended to assess algorithmic performance rather than model performance. Model performance should be tested with application specific transfer functions such as: recoverable reserves; flow simulation; effective permeability readings; connectivity of extremes; volume above a cutoff value.

In the context of testing algorithm performance the proposed metrics provide necessary, but not sufficient, confirmation tools. One important aspect of geological modeling is the diverse range of applications for each model created. Every application demands a set of confirmation tests that assess the high order statistical properties relevant to the models intended use. The tools provided were created to test common uses of geological categorical models.

References

Arpat, G.B. and Caers, J., 2007, Conditional simulation with patterns: Mathematical Geology, v. 39, no. 2, p. 177-203.

Boisvert, J. B., Pyrcz, M. J. and Deutsch, C. V., 2007, Multiple-point statistics for training image selection: Natural Resources Research, v. 16, no. 4, p. 313-321.

Caers, J., 2001, Geostatistical reservoir modeling using statistical pattern recognition: Journal of Petroleum Science and Engineering, v. 29, no. 3, p. 177-188.

Caers, J., Strebelle, S. and Payrazyan, K., 2003, Stochastic integration of seismic and geological scenarios: a submarine channel saga: The Leading Edge, v. 22, no. 3, p. 192-196.

Davis, B. M., 1987, Uses and abuses of cross-validation in geostatistics: Mathematical Geology, v. 19, no. 3, p. 241-248.

Deutsch, C. V., 1992, Annealing techniques applied to reservoir modeling and the integration of geological and engineering (well test) data: PhD thesis, Stanford University, Stanford, CA.

Deutsch, C.V. and A.G. Journel, 1998, GSLIB - geostatistical software library and user's guide: New York, Oxford University Press, 369 p.

Eskandari, K., 2008, Growthsim: A complete framework for integrating static and dynamic data into reservoir models: Ph.D. Thesis, University of Texas at Austin, 199 p.

Guardiano, F. and Srivastava, M., 1993, Multivariate geostatistics, beyond bivariate moments, *in* Soares, A. ed., Geostatistics Troia '92: v. 1, p. 133-144.

Harding, A., Strebelle, S., Levy, M., Thorne, J., Xie, D., Leigh, S., Preece, R. and Scamman, R., 2004, Reservoir facies modeling: New advances in MPS, *in* Leunangthong, O. and Deutsch, C. V., eds., Proceedings of the Seventh International Geostatistics Congress: Banff, Alberta, 10 p.

Leuangthong, O, McLennan, J. A. and Deutsch, C. V., 2004, Minimum acceptance criteria for geostatistical realizations: Natural Resources Research, v. 13, no. 3, p. 131-141.

Liu, Y., Harding, A., Abriel, W. and S. Strebelle, 2004, Multiple-point statistics simulation integrating wells, seismic data and geology: AAPG Bulletin, v. 88, no. 7, p. 905-921.

Lyster, S. and Deutsch, C.V., 2008, MPS simulation in a gibbs sampler algorithm, *in* 8th International Geostatistics Congress Chile, Awaiting publication.

Oreskes, N., Shrader-Frechette, K. and Belitz, K., 1994, Verification, validation, and confirmation in the earth sciences: Science, v. 263, no. 5147, p. 641-646.

Ortiz, J. M. and Deutsch, C. V., 2004, Indicator simulation accounting for multiple point statistics: Mathematical Geology, v. 36, no. 5, p. 545-565.

Pyrcz, M. J., Deutsch, C.V. and Boisvert, J.B., 2007, A library of training images for fluvial and deepwater reservoirs and associated code: Computers and Geosciences, v. 34, no. 5, p. 542-560.

Strebelle, S., 2002, Conditional simulation of complex geological structures using multiple-point statistics: Mathematical Geology, v. 34, no. 1, p. 1-22.

Strebelle, S., Payrazyan, K. and Caers, J., 2003, Modeling of a deepwater turbidite reservoir conditional to seismic data using multiple-point geostatistics: SPE Journal, v. 8, no. 3, p. 227-235.

Strebelle, S., 2007, Simulation of petrophysical property trends within facies geobodies: Petroleum Geostatistics 2007.

Wu, J., Zhang, T. and Journel, A., 2008, Fast FILTERSIM simulation with score-based distance: Mathematical Geosciences, DOI 10.1007/s11004-008-9157-5.