POROSITY AND PERMEABILITY MODELING USING PCA IN THE COAL BED RESERVOIR

Manasi Manjari Mohanty

Department of Mining Engineering, N.I.T, Rourkela, (India)

ABSRACT

The flow behavior of injected CO_2 is highly depends on the value of the porosity and permeability of the cleats and coal matrix. In this paper, the multi-point geo-statistical algorithm is applied for porosity and permeability modeling. Multiple-point geostatistical approaches (MPS) do not try to reproduce the training image itself, but they consist of extracting statistics of the variables from the training image statistics at multiple locations, and subsequently, using them to reproduce similar features to the training image. The principal component analysis (PCA) is applied to reduce the dimension of the pattern. The advantage of the PCA is that it can map the data in different principal components (PCs) which are normal to each other. The k-means clustering algorithm is applied for classifying the PCs of pattern database.

APrincipal component analysis (PCA) based clustering algorithm is proposed for pattern-based simulation. The dimensional reduction of the pattern database is performed using the PCA. The members of a pattern are more and less correlated. Therefore, a simulation algorithm can be developed more efficiently by taking these correlations into account and projecting them into an uncorrelated domain using the PCA. The main concern of PCA–based simulation is to select the optimum number of principal components (PCs) and optimum cluster numbers. In this paper, we have developed a PCA-based simulation where the number of PCs and number of cluster will be selected automatically.

Keywords: Multiple-Point Geo-Statistics, PCA, K-Means Clustering, Euclidian Distance

I. INTRODUCTION

Coal is a naturally fractured porous solid with a dual porosity consisting of both micropores and macropores (Kolesar et al, 1990). The microporosity of coal is contained within the macromolecular network of the coal matrix. The macroporosity of a coal seam consists of the naturally occurring fractures called cleats (Meyers, 1982). Coals also contain a range of microstructures of various shapes and sizes between the micropores and the cleats (Gamson et al, 1993). The storage of gas is dominated by adsorption within micropores; whereas, the cleat system provides the medium for mass transfer through the formation (Shi and Durucan 2008).

However in reality porosity and permeability value vary from location to location within the domain. Therefore, the location dependent porosity and permeability values are more realistic idea for reservoir modeling. In this report, the multi-point geo-statistical algorithm will be applied for porosity and permeability modelingpattern based simulation algorithms where the patterns are extracted from the training image. Pattern-based simulation algorithms consider the training image as a collection of patterns, from which a pattern can be selected to locally match as close as possible to the conditioning data (Arpat and Cares 2007). During simulation, multi-point conditioning data is compared with patterns of the training image and the most similar pattern corresponding to the conditioning data is selected from the training image and pasted at the simulated node.

A Principal component analysis (PCA) based clustering algorithm is proposed for pattern-based simulation. The dimensional reduction of the pattern database is performed using the PCA. The members of a pattern are more and less correlated. Therefore, a simulation algorithm can be developed more efficiently by taking these correlations into account and projecting them into an uncorrelated domain using the PCA. The main concern of PCA–based simulation is to select the optimum number of principal components (PCs) and optimum cluster numbers. In this paper, we have developed a PCA-based simulation where the number of PCs and number of cluster will be selected automatically. During simulation, the similarity of the classes with the conditioning data is calculated.

II. POROSITY AND PERMEABILITY MODELING USING MULTIPOINT STATISTICS

The flow behavior of injected CO_2 is highly depends on the value of the porosity and permeability of the cleats and coal matrix. Most of the studies it is assumed that these values are either scalar (porosity) or tensor (permeability) and constant over the geological formation. However in reality they vary from location to location within the domain. Therefore, the location dependent porosity and permeability values are more realistic idea for reservoir modeling. In this report, the multi-point geo-statistical algorithm will be applied for porosity and permeability modeling/There are a number of multi-point simulation techniques proposed in the literature i.e. simpat,snesim,filtersim,cumulant-based simulationand wavelet-based simulation (Arpat and Cares 2007;Strebelle 2002;Zhang et al. 2006; Wu et al. 2008;Mustapha and Dimitrakopoulos 2010;Chatterjee et al. 2012). Out of these multi-point simulation techniques, simpat, filtersim and wavelet-based simulation are pattern simulation algorithms. In pattern-based multiple-point geo-statistical algorithm, the patterns are extracted from the pattern from the training image and the most similar pattern corresponding to the conditioning data event is pasted at the simulated node. Simpat (Arpat and Cares 2007), considers the training image as a collection of patterns, from which a pattern can be selected to locally match as close as possible to the conditioning data event. The main advantage of this algorithm is that no conditioning data points from the conditioning data event are required to be deleted, however, the major limitation is that the entire pattern database will be searched to find the best match at each simulating node; therefore computational time will be extensively high. The filtersim (simulation using filter scores)(Zhang et al. 2006; Wu et al. 2008) algorithm overcomes SIMPAT's computing limitation. Like simpat, the main advantage of filtersim is that no conditioning data points need to be deleted from the conditioning data event for matching with the patterns from the pattern database. In filtersim, the scanning of the entire training image is performed using a given template to obtain patterns. Different filters are applied on patterns to obtain values of filter scores. The patterns in the pattern database are then grouped, based on their filter score values, into different classes. The classes are represented by their prototype, which is the average value of all patterns in a class. During simulation, the conditioning data event is compared with the class prototypes to find the closest matched class. Unlike simpat, filtersim does not need to search the entire pattern database. The algorithm is looking for 'best match' rather than 'exact match'; therefore, no elimination of conditioning data points from the data event is required. The main disadvantage of the filtersim algorithm is that it only considered only few numbers of filters to represent the pattern. These filters may not always able to classify the patterns in correct classes. Wavesim (Chatterjee et al. 2012) proposed an alternative algorithm to reduce the dimension of the patterns using wavelet decomposition. The performance of their approach is significantly better than filtersim; however some time it is computationally expensive. In this research work, the principal component analysis (PCA) is applied to reduce the dimension of the pattern. The advantage of the PCA is that it can map the data in different principal components (PCs) which are normal to each other. Thus, PCs are independent to each other. Since, the PCs are extracted by Eigen value decomposition; the first few principal components can capture the maximum data variability. Therefore, preserving only first few PCs and eliminating rest of the PCs can significantly reduce the dimension of the pattern database. It is noted that the difference of all pattern-based simulation algorithm is how the dimensionality of the patterns can be reduced. Fig. 3 represents the schematic diagram of the proposed PCA-based simulation algorithm. The training image is first scan using the selected template (9x9) and stored in the pattern database. The dimension of the pattern database is reduced by performing the PCA analysis. The first 4 PCs of pattern database are used for classification. It can be observed that the dimensions of the patterns are reduced from 81 to 4. The k-means clustering algorithm is applied for classifying the PCs of pattern database and each class is represented by their class prototype. Then simulating a point within the domain, the conditioning data event is extracted by placing the template at the node of simulation. The similarity of the conditioning data event and class prototype is calculated and a random pattern is selected from the best match class. This procedure is repetitive for simulation within the domain. For measuring the similarity between conditioning data event and class prototype, Euclidian distance is considered.

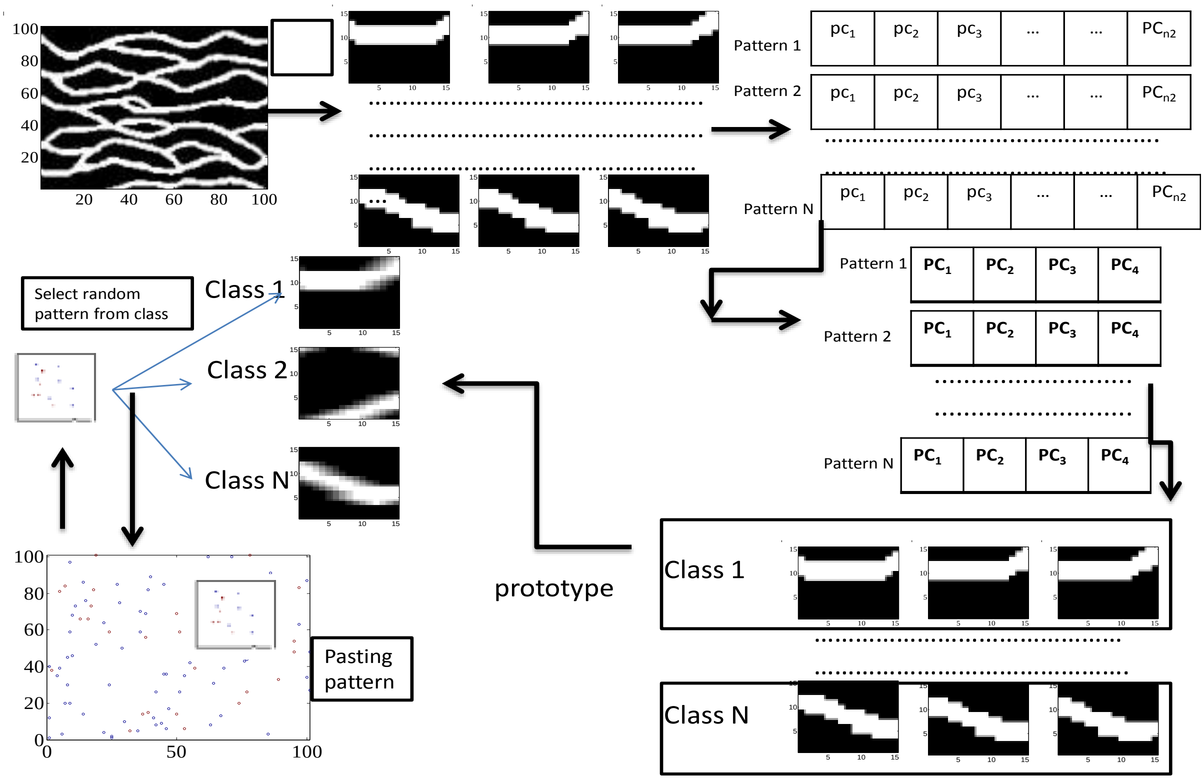
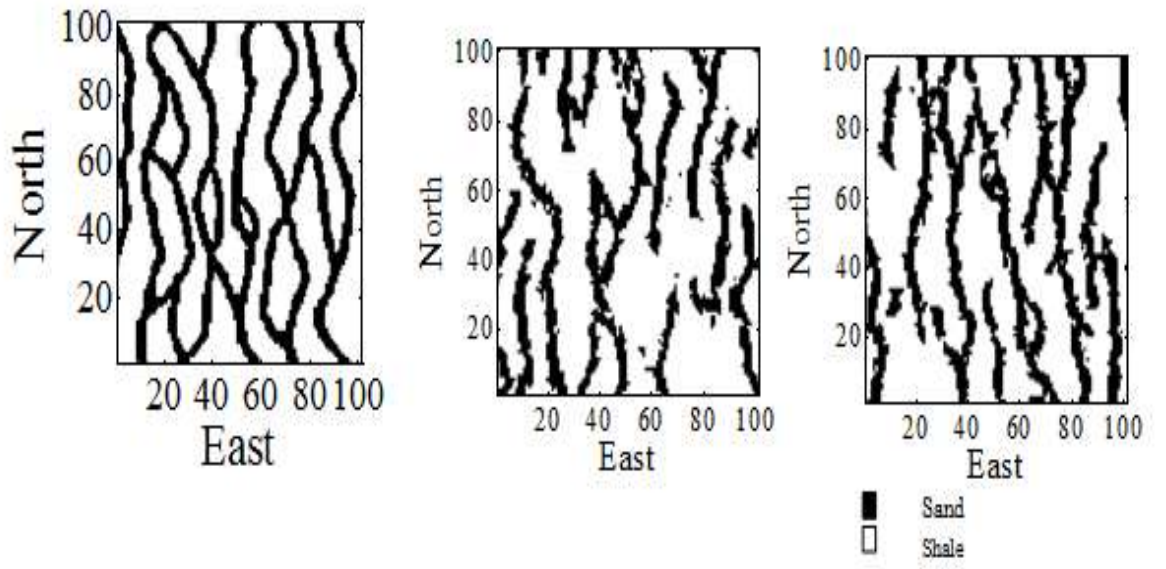


Fig. 1 PCA-Based Multiple Point Simulation

III. MULTIPLE-POINT SIMULATION ALGORITHM

The PCA-based multi-point geo-statistical algorithm proposed in this paper is validated by conditional and unconditional simulation of known image. A categorical unconditional simulation is performed for two-category (Sand and Shale) data sets. The binary training image consists of sand and shale materials. This training image represents complex channels present in a deposit (Fig. 2). The template size is selected using optimum template selection algorithm and for this training image the optimum template size is 9 x 9. The numbers of PCs are selected based on the cumulative amount of data variability preservation of the pattern database. In this work, the percentage of the data variability is fixed at 80%. Therefore, only those first PCs are selected where cumulative data variability is 80%. After PCA analysis, it is observed that with 8 first PCs, the 80% of the pattern database variability can be captured. Therefore, only 8 PCs are used for pattern classification. It is noted that the filtersim used 6 filters for pattern classification, so computational time of the proposed algorithm is marginally more than filtersim algorithm. The unconditionally simulated realisations generated using the proposed PCA-based algorithm is presented in Fig. The results revealed that the continuities of the channels are nicely reproduced by the proposed algorithm. The proportion of shale and sand in training image are also reproduced by the simulated realisations. When, the results were compared with filtersim, it is observed that the proposed algorithm is performed better than filtersim.

To perform the conditional simulation, using the proposed algorithm, two conditional data points at location (45, 47) and location (54, 49) are considered. The first conditioning point consists of shale and second point consists of sand. The same template size, clusters numbers are selected for conditional simulation algorithm as selected in unconditional simulation. Two different realizations are presented in the Fig 3 (a). Results reveal that the continuity of channels is reproduced. The first and second order statistics are also reproduced using the proposed algorithm. The algorithm performed better than filtersim for channel reproduction. The conditional capability of the proposed algorithm is tested using the ensemble map. The ensemble map of 60 conditional simulated realizations is presented in Fig 3 (b). The ensemble map revealed that the proposed algorithm honoured the conditioning data nicely.



(a) (b)

Fig. 2 (A) Training Image; (B) Two Unconditionally Simulated Realizations Generated Using PCA-Based Approach

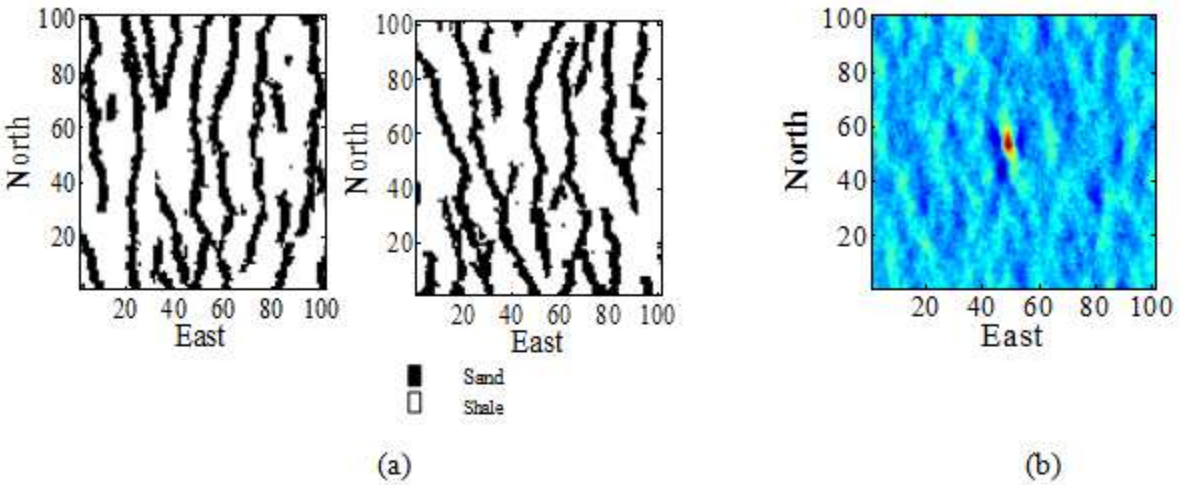


Fig. 3 (A) Two Conditionally Simulated Realizations Generated Using PCA-Based Approach; (B) Ensemble Map

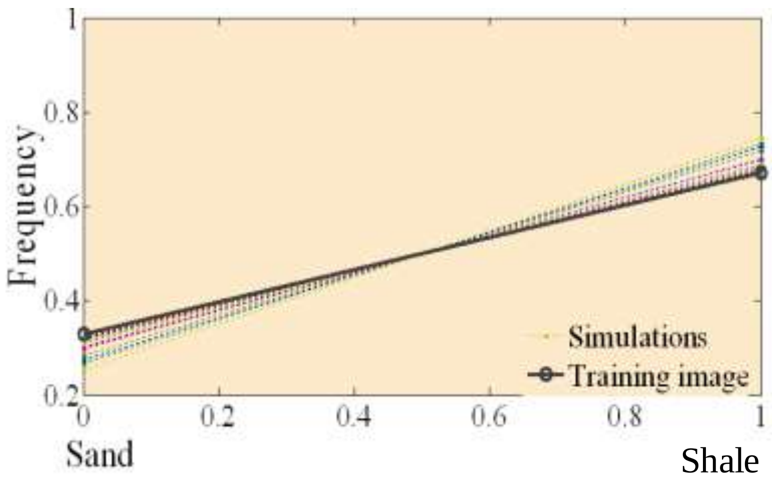


Fig.4 Reproduction of Proportion of Shale and Sand Using the Proposed Method

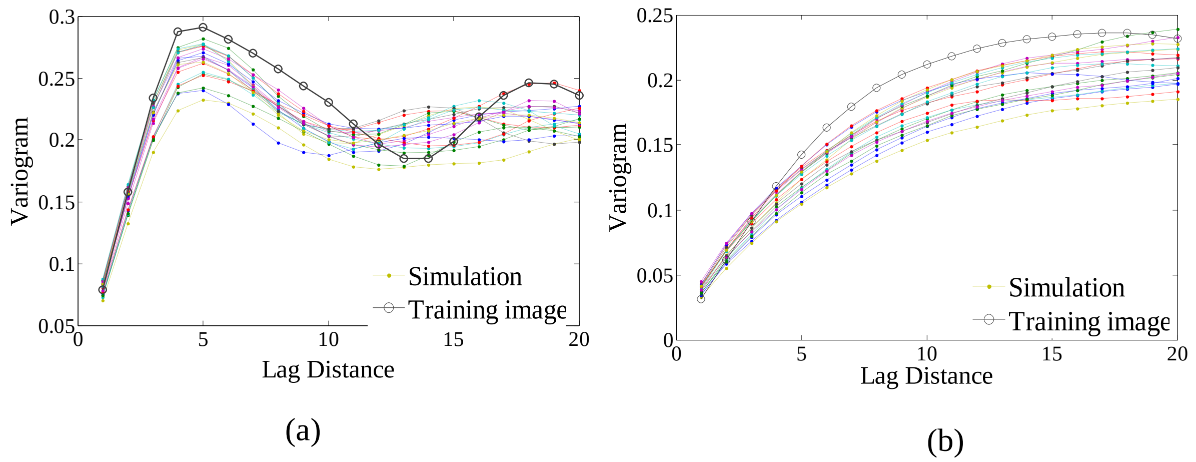


Figure.5 Reproduction of Proportion of Shale and Sand Using the Proposed Method

Figure 4 represents the proportion map of shale and sand for both training image and simulations. Figure shows that the proportion of shale and sand in training image are reproduced by the simulated realisations. The two-directional variograms of training image and simulated realizations are presented in Figure5. Variogram figures show that the directional variograms of simulated realisations are closely matched with the training image variograms. The reproduction of statistics and reproduction of continuity of channels revealed that proposed algorithm can reproduced both the statistics and complex structure.Figure6 represents the training image for this study. The size of the training image is 100×128. However, the simulation domain size selected for this study is 128 x 128. The reason for selecting the different simulation domain size from training image size is to show that the proposed algorithm can simulate different domain size from the training image size.The template size used in this study is selected using an automatic template selection algorithm (13×13). The pattern database is generated by scanning the entire training image. The pattern database is then mapped in an orthogonal space using PCA. In this example, first 10 PCs are selected which captured more than 80% pattern database variability. The *k*-mean clustering is applied to classify the PCA based reduced pattern database. The optimum cluster number selected for this study using gap statistics is 187. Two different realizations are generated using our proposed approach with optimum cluster number and presented in Figure7. In the continuous case, it is difficult to compare pattern-based simulation results; however, by visual inspection, a qualitative comparison can be made. It is observed from simulated realisations that channels of training image are reproduced. Moreover, from visual observation, it shows that the proportions of high, medium, low values of training image are also reproduced by simulated realisations. The histogram demonstrates that the simulated realizations reproduced the one-point statistics of the training image. Directional variograms of the training image and simulated realisations also show that the two-point statistics are also reproduced by the proposed algorithm. However, due to space limitation histogram and variograms are not shown here.

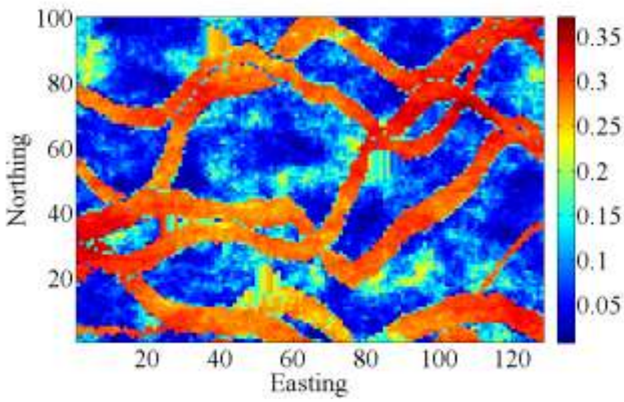


Fig.6 Continuous Training Image

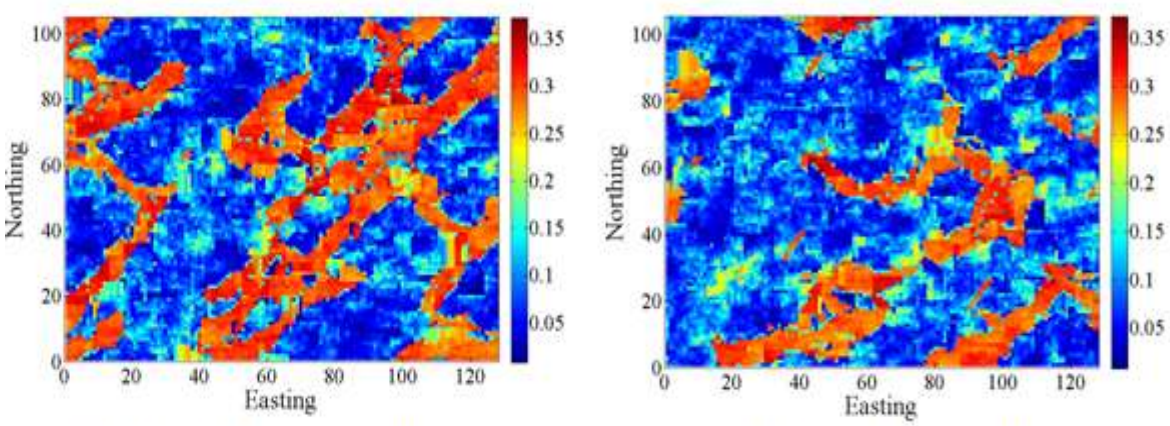


Fig.7 Two Different Unconditionally Simulated Realizations Using the Proposed Method

IV. CONCLUSION

In this paper, the PCA-based dimensional reduction algorithm is proposed for pattern-based geo-statistical simulation. The PCA helps to reduce the dimensions of the patterns by mapping them in orthogonal space. The main advantage of PCA over other dimensional reduction algorithms is that with the limited number of PCs maximum pattern database variability can be captured. Four different examples were presented in this paper to show the performance of the proposed approach. The results revealed that reproduction of statistics as well as continuity of the curvilinear structure can be well reproduced. The parameter like cluster number which has strong influence on the performance of other pattern-based simulation algorithms is selected automatically; therefore, sensitivity issues can be minimised. The main limitation of the proposed PCA-based simulation algorithm is that the projections using PCs are performed linearly which considers that the members of a pattern are linearly separable. The proposed algorithm may fail when pattern members are not linearly separable. The kernel-PCA which is more generalised algorithm for high-dimensional mapping can overcome the linearity issue PCA.

REFERENCES

[1] Honarkhah M and Caers J (2010) Stochastic simulation of patterns using distance-based pattern modelling. Mathematical Geosciences 42: 487-517.

[2] Mustapha H, Dimitrakopoulos R (2010) High-order stochastic simulations for complex non-Gaussian and non-linear geological patterns, Mathematical Geosciences, 42 (5):

[3] Strebelle S (2002) Conditional simulation of complex geological structures using multiplepoint statistics. Mathematical Geology 34 (1): 1-21.

[4] Wu J, Zhang T, Journal A (2008) Fast FILTERSIM simulation with score-based distance, Mathematical Geosciences 40(7): 773-788.

[5] Zhang T, Switzer P, and Journal A (2006) Filter-based classification of training image patterns for spatial simulation. Mathematical Geology 38(1): 63–80.

[6] Chatterjee S and Dimitrakopoulos R (2011) Multi-scale stochastic simulation with a wavelet-based approach. Computers & Geosciences, doi:10.1016/j.cageo.2011.11.006.

[7] Guardiano FB, and and Srivastava RM (1993) Multivariate gostatistics, beyond bivariate moments, Quantitative Geology and Geostatistics 5:133-144.

[8] Journel AG (1997) Deterministic geostatistics, A new visit In, Baafy E. and Shofield, N. (eds.), Geostatistics Woolongong '96', Kluwer, Dordrecht, 213-224.

[9] Chatterjee S, Dimitrakopulos R, and Mustafa H (2012) Dimensional reduction of pattern-based simulation using wavelet analysis, Mathematical Geosciences, Vol. 44, pp. 343-374.

[10] Mustafa H, Chatterjee S, and Dimitrakopulos R, Efficient Pattern-cdf Spatial Simulation Through Decomposition of Cumulative Distribution Functions of Transformed Spatial Patterns, Mathematical Geosciences, DOI 10.1007/s11004-013-9490-1.

[11] Jolliffe, I Principal Component Analysis, Springer-Verlag, New York, 1986.

[12] Mustapha H, and Dimitrakopoulos R (2010) High-order stochastic simulations for complex non-Gaussian and non-linear geological patterns, Mathematical Geosciences, 42 (5).

[13] Strebelle S (2002) Conditional simulation of complex geological structures using multiplepoint statistics, Mathematical Geology 34 (1): 1-21.

[14] Wu J, Zhang T, and Journel A (2008) Fast FILTERSIM simulation with score-based distance, Mathematical Geosciences 40(7): 773-788.