

A New Method to Select Training Images in Multi-Point Geostatistics

Lixin Wang, Yanshu Yin, Wenjie Feng

School of Geosciences, Yangtze University, Wuhan, China Email: 201571323@yangtzeu.edu.cn

How to cite this paper: Wang, L.X., Yin, Y.S. and Feng, W.J. (2018) A New Method to Select Training Images in Multi-Point Geostatistics. *Open Journal of Yangtze Gas and Oil*, **3**, 112-129. https://doi.org/10.4236/ojogas.2018.32010

Received: December 20, 2017 **Accepted:** April 25, 2018 **Published:** April 28, 2018

Copyright © 2018 by authors and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0). http://creativecommons.org/licenses/by/4.0/

Abstract

Training images, as an important modeling parameter in the multi-point geostatistics, directly determine the effect of modeling. It's necessary to evaluate and select the candidate training image before using the multi-point geostatistical modeling. The overall repetition probability is not sufficient to describe the relationship of single data events in the training image. Based on the understanding, a new method was presented in this paper to select the training image. As is shown in the basic idea, the repetition probability distribution of a single data event was used to characterize the type and stationarity of the sedimentary pattern in the training image. The repetition probability mean value and deviation of single data event reflected the stationarity of the geological model of the training image; the rate of data event mismatching reflected the diversity of geological patterns in training images. The selection of optimal training image was achieved by combining the probability of repeated events and the probability of overall repetition of single data events. It's illustrated in the simulation tests that a good training image has the advantages of high repetition probability compatibility, stable distribution of repeated probability of single data event, low probability mean value, low probability deviation and low rate of mismatching. The method can quickly select the training image and provide the basic guarantee for multi-point geostatistical simulations.

Keywords

Training Image, Single Data Event Repetition Probability, Repetition Probability Mean Value, Multi-Point Geostatistics

1. Introduction

Multi-point geostatistics was proposed by Guardiano and Srivastava in 1993 [1],

which aimed to cope with the problem of the insufficient consideration of twopoint statistical information. The problem made it difficult to reproduce the shape of the simulated target more truthfully. By establishing a quantitative training image, the probability of determining different data events after scanning with a multi-point template was used to characterize the probability of occurrence of different data events. The objective of multi-point geostatistics is to recreate the geological patterns contained in the training images, so that training images can be considered as one of the key factors that determine the effect of simulation [2]-[9]. In recent years, in order to obtain effective training images, scholars have proposed different methods, including the target-based method [10] [11] [12], the method based on the deposition process [5] [13] [14], the method based on the process of imitation deposition [15] [16], and the method based on geological data transformation [17] [18], etc.

At present, there are so many methods of creating training images that a large number of different training images can be created through various methods and tools for a certain research area. However, as a geological understanding of training images, how to select one or more most-suitable training images for the actual research area from multiple (group) training images of different sources, different creation methods, different spatial structure characteristics and credibility before conducting multi-point modeling? It has become a problem that modelers have to face. Yet, the optimal selection methods for training images are very limited, which include the optimal selection method based on variogram, the method based on conditional probability [3] [6] [19], and the method based on similar distance [20].

The optimal selection method based on variogram can effectively obtain the two-point geostatistical information contained in the data volume, but it is limited by the two-point geostatistics of the variogram. It can only be used to compare the features of second-order space structure, but can not analyze and compare the higher-order geostatistical features. Ortiz and Deutsch first proposed a way to sort training images through high-level geostatistical information [19]. By the method, data events composed of a plurality of grid points in a single well can be obtained, and the training images can be scanned to obtain the distribution of the condition data events in the training images. The training images were sorted by comparing multiple distribution features. Boisvert further proposed a training image optimization method based on data event distribution and multi-point density equations [3]. The example tests showed that the above two methods can effectively sort the training images. However, these two methods can only be used to analyze and compare one-dimensional data extracted from a single well, but no effective high-level geological statistics can be obtained in the three-dimensional space. Then, Pérez proposed a training image optimization method based on three-dimensional data event repetition probability statistics [6], that is, the spiral search was conducted to obtain condition data events in the condition data, search the training data events of the spatial structure in the candidate training images, count the number of repetitions appearing in different training images, normalize all the repetitions obtained from each condition data event, and then obtain the average of the repetitions of each condition data event to get the compatibility between different training images and condition data events. However, this method simplified the calculation of data event disparity in data event search and matching degree calculation, and allocated the same weight to each point in the data event. In addition, this method cannot exactly reveal the true match between the training image and the condition data, and cannot differentiate and analyze a large number of training data events. And there is no direct relationship between the overall compatibility of training images with data events and the compatibility of training images with individual data events. Therefore, this method still cannot provide the absolute matching of different data events and training images in the condition data.

Based on Pérez's methodological analysis, this paper considered the issue that, in some cases, the overall probability of repetition may result in a high overall compatibility due to the repetition of a certain pattern in the training images, as no direct relation exists between the overall compatibility of training images and data events and the compatibility between training images and single data events. Furthermore, a new index was proposed in this paper, that is, statistical characteristic parameters of single data event repetition. These two ideas were combined to sort and optimize the training images. The synthetic theoretical model showed that the new method could better achieve the sorting and optimization of training images. The research provided a new method for multi-point geostatistical modeling core and key parameters, *i.e.* training image optimization. It promoted multi-point geological modeling to better serve the reservoir model establishment and laid the foundation for enhanced oil recovery. An accurate training image could improve the effect of modeling, making the multi-point modeling closer to the actual reservoir situation. [9] [17] [21] [22] [23].

2. Optimal Selection Method

2.1. Method Based on Overall Repetition Probability

Pérez (2014) proposed to optimize the training images by counting the repetition probability of the whole data event and computing the relative compatibility and absolute compatibility.

The relative compatibility is to normalize the repetition number of each data event and calculate the repetition probability $P_{i,j}$ of the *i*-th data event in each training image,

$$P_{i,j} = \frac{R_{i,j}}{\sum_{j=1}^{t} R_{i,j}}$$
(1)

 $R_{i,j}$ represents the repetition number of the *i*-th data event in the *j*-th training image, and then calculate the average repetition probability of the *n*-th data

events as the relative compatibility C_p

$$C_{j} = \frac{\sum_{i=1}^{n} P_{i,j}}{\sum_{i=1}^{r} \sum_{i=1}^{n} P_{i,j}}$$
(2)

Absolute compatibility is the occurrence of statistical events in the training image. If the *i*-th data event has appeared in the *j*-th training image, $Y_{i,j}$ is 1, otherwise $Y_{i,j}$ would be 0, then the proportion of data events contained in this training image is calculated, that is, absolute compatibility $M_{i,j}$

$$M_{j} = \frac{\sum_{i=1}^{n} Y_{i,j}}{n}$$
(3)

Through the relative compatibility characterizing the probability of occurrence of conditional patterns and characterizing the pattern matching rate in the training images by absolute compatibility, the overall characteristics of the training images can be reflected. However, as there is no direct relationship between the overall compatibility of training images with data events and the compatibility of training images with individual data events. In some cases, the overall repetition probability may result in a high overall compatibility due to the repetition of a pattern in the training image. As shown in **Figure 1**, there are three training images (the number of grids is $50 \times 50 \times 1$), and their geological features are similar. The condition data TIC3 is obtained from the training image T3. According to the method of Pérez (2014), the overall repetition probability is used to optimize the training images, but the result of the evaluation is not significant enough. When the number of condition points is more than 7, there will be a big difference (**Figure 2**). Based on this understanding,



Figure 1. Training image and condition data (Pérez, 2014). (a) Training image T1; (b) training image T1; (c) training image T3; (d) training image TIC3 (from T3).





a single data event repetition probability analysis based on its absolute compatibility and relative compatibility was proposed to make up for the shortcoming that the overall repetition probability does not reflect the distribution of individual data events within the training image.

2.2. Statistical Characteristic of Single Data Event Repetition

The single data event repetition probability is designed to reflect the distribution characteristics of data events within a certain training image. It uses the conditional probability as the evaluation data and selects a suitable search range and the number of conditional points involved in evaluation to weight the grid points within the search range. It also finds the number of occurrences of this mode in the training images and records the number of repetitions for each mode. That is, for the *t*-th candidate training images, the set of the *n* data events CE is obtained by scanning the condition data with the specified template, and the number of occurrences of the *i*-th data event CE_i in the *j*-th training image is denoted as R_{i,i}. Then, the distribution statistics of data events in each training image are calculated, so as to select a better training image. The statistical characteristics of these distributions include: single data event repetition probability distribution, single data event repetition probability average, single data event repetition probability deviation and data event mismatch rate. With single data event repetition probability distribution, single data event repetition probability average and single data event repetition probability deviation, the stability of data events in the training image can be reflected. And with data event mismatch rate, the diversity of training image patterns can be highlighted. The repetition rate of a single data event is the repetition probability of a single data event in the repetition of all data events of a training image, that is,

$$PT_{i,j} = \frac{R_{i,j}}{\sum_{i=1}^{n} R_{i,j}}$$
(4)

Data events with $PT_{i,j}$ being 0 mean no matching event in the training image. If there is no match found in the training image, it will be marked as 1, otherwise 0, then no match will be calculated, where,

$$UNR_{i,j} = \begin{cases} 1; & PT_{i,j} = 0\\ 0; & PT_{i,j} \neq 0 \end{cases}$$
(5)

$$UNP_{j} = \frac{\sum_{i=1}^{n} UNR_{i,j}}{n}$$
(6)

 $UNR_{i,j}$ is the index of mismatch events, and UNP_j is the mismatch rate. When establishing statistical distribution probability for one-event repetition probability $PT_{i,p}$ without considering data events without matching, the effective data event repetition probability $PT_{i,j}$ is calculated by interval, and the distribution probability average and deviation are calculated. The training images with lower data event mismatch rate, even single data event repetition probability distribution and smaller single data event repetition probability average and single data event repetition probability deviation are closer to the real geological features. Aiming at the poor performance of the above training images, the probabilistic characteristics of single data events are statistically analyzed when the five conditional points are taken (Figure 3). It can be clearly seen from Figure 3 that the single data event repetition probability deviation and data event mismatch rate of single data events are obviously lower. Aided by a single data event indicator and combined with the overall repetition probability indicators, it will be able to more directly filter out the training images in line with the actual geological features.



Figure 3. Statistical characteristics of the single data event repetition probability. (a) Repetition probability deviation; (b) repetition probability average; (c) repetition probability distribution; (d) mismatch rate.

2.3. Process of the Method

Through the programming, the method of combining the overall repetition probability and the single data event statistical index is proposed to select the optimal training image. By meshing the work area with known condition data, a random search path is established. At the same time, the search range of the template is sorted by weight. For any node location, the search sequentially matches the condition data event exactly from the nearest condition point to the farthest condition point. Once the perfect match pattern is found, the number of repetitions for this pattern increases until all data points in the data model are searched across the training image, which returns the number of repetitions $R_{i,j}$ that exactly matches the condition data event, and calculates the normalized

probability $P_{i,j}$ and the single-event repetition probability $PT_{i,j}$. According to the normalized probability, the relative compatibility and absolute compatibility of the whole training image are calculated. According to the single event repetition probability, the distribution proportion, the distribution mean and the distribution deviation are calculated: The specific steps are as follows (**Figure 4**):

1) Determine the search template, then create a search template weight ranking, and determine the pseudo-random path to find data events according to the distribution of condition data.

2) Scan training images to look for patterns that match the data events. If the data event condition points find an exact match in the training image, the event repetition number $R_{i,j}$ is incremented by 1 until the training image search is completed.

3) Jump to the next data event, repeat step 3) until all data events have been searched.



Figure 4. The flow chart of training image evaluation.

4) Select the next training image, repeat steps 2) - 4) until all the training images have been scanned.

5) Get the normalized probability $P_{i,j}$ and the single event probability $PT_{i,j}$ to calculate relative compatibility C_p absolute compatibility M_p . Single data event repetition probability average and Single data event repetition probability deviation and Data event mismatch rate of single data events UNP_p .

3. Test of the Method

3.1. Two-Dimensional Test

The two-dimensional test grid adopted were the training images published by Pérez (2014) with a grid size of $100 \times 100 \times 1$ (**Figure 5**). From the real images TI4, TI5, TI6, 1091 conditional points were randomly selected, corresponding to TIC4, TIC5, TIC6. Based on the condition data, the training images of candidate T1, T2, T3 were tested and sorted, and the training image was optimized.



TIC4: The condition data form TI4; TIC5: The condition data form TI5; TIC6: The condition data form TI6; T1, T2, T3: The training image for MPS; For the condition data TIC4, T1, T2 and T3 are used as the modeling parameter. The same as TIC5 and TIC6.

Figure 5. Training image and condition data.

The maximum search range in the test was set to $31 \times 31 \times 1$, and the number of upper limit condition points was to 35. The absolute compatibility and the relative compatibility were calculated respectively for the number of repetitions when searching for 5, 10, 15, 20, 25, 30, 35 condition points within the search range (**Figure 6**). It can be seen that as the condition points increased, the relative compatibility of the training images close to the original geological model tended to increase, while the absolute compatibility was higher than that of other training images. For the data events when 15 conditional points were considered, the Single data event repetition probability distribution, Single data event repetition probability average, Single data event repetition probability deviation and data event mismatch rate were calculated (**Figure 7**). And it is not difficult to find that, with better training images, there comes more stable repetition probability distribution, lower repetition probability average and deviation and mismatch rate.



Figure 6. Statistical characteristics of the overall repetition probability. (a) Absolute compatibility; (b) relative compatibility.

Based on the above parameters, the training image T1 was preferably selected based on the condition point TIC4, the training image T2 was preferably selected based on the condition point TIC5, and the training image T3 was preferably selected based on the condition point TIC6. According to the multi-point simulation with three training images and three sets of condition data (**Figure 8**), with the template size of $5 \times 5 \times 1$, it can be concluded that the optimal training images corresponding to condition points TIC4, TIC5 and TIC6 were T1, T2 and T3 respectively, indicating that the results of multi-point simulation were in good agreement with the training images.



Figure 7. Statistical characteristics of the single data event repetition probability. (a) Repetition probability deviation; (b) repetition probability average; (c) repetition probability distribution; (d) mismatch rate.



Figure 8. Multi-point simulation results.

3.2. Three-Dimensional Test

With the three-dimensional test grid with the size of $60 \times 60 \times 10$, three different specifications (**Table 1**) of the river phase model TI4, TI5, TI6 and 900 corresponding to the point data were established, and at the same time, three training images T1, T2 and T3 were selected (**Figure 9**). For three different data conditions, the test tried to find their appropriate training images. For multi-point modeling, the maximum conditional point is 35. The grid size is $20 \times 20 \times 4$ meters. It can be seen that the width of T1 is the largest, the thickness of T3 is the smallest, and the thickness of T2 is the largest while its width is moderate.





DOI: 10.4236/ojogas.2018.32010



DOI: 10.4236/ojogas.2018.32010



Figure 9. Training image and condition data. (a) Geologic model T4 and condition data TIC4; (b) geologic model T5 and condition data TIC5; (c) geologic model T6 and condition data TIC6; (d) training image T1, T2, T3.

The maximum search range in the test was set to $31 \times 31 \times 9$, and the number of upper limit condition points was to 35. The absolute compatibility and the relative compatibility were calculated respectively for the number of repetitions when searching for 5, 10, 15, 20, 25, 30, 35 condition points within the search range (Figure 10). It can be seen that when the condition point TIC4 or the condition point TIC5 was not available for the training images T1 and T2, the condition point TIC6 could better select the training image T3 with similar geological parameters. For the data events when 15 conditional points were considered, the Single data event repetition probability distribution, Single data event repetition probability average, Single data event repetition probability deviation and data event mismatch rate were calculated (Figure 11). And it is not difficult to find that, with better training images, there comes more stable repetition probability distribution, lower repetition probability average and deviation and mismatch rate. Because single data event analysis presented the distribution of internal patterns of training images, it directly revealed the distribution of single data events rather than replacing the local probability



Figure 10. Statistical characteristics of the overall repetition probability. (a) Absolute compatibility, (b) relative compatibility.



Figure 11. Statistical characteristics of the single data event repetition probability. (a) Repetition probability deviation; (b) repetition probability average; (c) repetition probability distribution; (d) mismatch rate.

distribution with the overall repetition probability. Therefore, the training images with similar parameters can be optimized by using the single-event repetition probability for the case that relatively good training images could not be selected by relative compatibility and absolute compatibility.

Multiple simulations were performed based on the three training images and the three sets of condition data (**Figure 12**). The differences between the three river phase models in terms of width and thickness were acceptable from the point of view of multipoint simulation. However, the optimality is the best. It is obvious that condition point TIC4 with training images T1, condition point TIC5 with training images T2 and condition point TIC6 with training images T3 produced the best simulation effect.

Based on the two-dimensional model and three-dimensional model test, it can be seen that the relative compatibility, the absolute compatibility and the absolute compatibility in the overall repetition probability can improve the optimal selection evaluation for the training image with significant difference. And for the training images whose structural features are close to each other, the overall repetition probability will give a better evaluation of the training images in the event of partial data events with a high number of repetitions. However, the single data event repetition probability starts from the distribution of single data event repetition number, and takes the stability of data events, which is evaluated with the Single data event repetition probability average, Deviation and Mismatch rate, as the optimal selection index of training images. Combined with the overall repetitive probability of data events, the training images can be more fully optimized.

4. Conclusions

The training image is equivalent to a geological pattern library for multi-point simulation, where data events are the embodiment of geological model. The advantages and disadvantages of the training images depend on the matching degree of the conditional patterns. It is an effective way to train the images by analyzing the data events.

The overall repetition probability of data events optimizes the overall pattern of training images through relative compatibility and absolute compatibility, which can reflect the matching degree of the geological patterns in the training images as a whole to the condition data. The higher relative compatibility and absolute compatibility have generally evaluated the training images. However, the lack of credibility of the condition data for a single data event would result in an additive effect of the individual significant data event on the overall repetition probability, and that training images that are not faithful to the condition data also be selected. Single data event repetition probability can make up for the overall repetition probability of a single data event description of the deficiencies and evaluate the stability of the distribution of individual data events.

In the steady reservoirs modeling, training image selected by this method can



Figure 12. Multi-point simulation results.

match with the actual geologic pattern, a good result in the actual modeling can be achieved, but for the optimization of training image in non-stationary reservoir modeling still, it needs to add some new control factors.

Acknowledgements

The work presented in the paper was financially supported by the National Natural Science Foundation of China (No. 41572081), the National Science and Technology Major Project (NO:2016ZX05031002-001 and 2016ZX05015001-001) and the Natural Science Foundation of Hubei Province Innovation Project Group.

References

- [1] Guardiano, F.B. and Srivastava, R.M. (1993) Multivariate Geostatistics: Beyond Bivariate Moments. Springer, Netherlands.
- [2] Arpat, G.B. and Caers, J. (2007) Conditional Simulation with Patterns. *Mathematical Geology*, 39, 177-203. <u>https://doi.org/10.1007/s11004-006-9075-3</u>
- Boisvert, J.B., Pyrcz, M.J. and Deutsch, C.V. (2007) Multiple-Point Statistics for Training Image Selection. *Natural Resources Research*, 16, 313-321. https://doi.org/10.1007/s11053-008-9058-9
- [4] Hu, L.Y. and Chugunova, T. (2008) Multiple-Point Geostatistics for Modeling Subsurface Heterogeneity: A Comprehensive Review. *Water Resources Research*, 44, 2276-2283. <u>https://doi.org/10.1029/2008WR006993</u>
- [5] Mariethoz, G. and Caers, J. (2014) Multiple-Point Geostatistics: Stochastic Modeling with Training Images.
- [6] Pérez, C., Mariethoz, G. and Ortiz, J.M. (2014) Verifying the High-Order Consistency of Training Images with Data for Multiple-Point Geostatistics. *Computers & Geosciences*, 70, 190-205. https://doi.org/10.1016/j.cageo.2014.06.001
- [7] Strebelle, S.B. and Journel, A.G. (2001) Reservoir Modeling Using Multiple-point Statistics. In: SPE Annual Technical Conference and Exhibition, Society of Petroleum Engineers. <u>https://doi.org/10.2118/71324-MS</u>
- [8] Zhang, T., Switzer, P. and Journel, A. (2006) Filter-Based Classification of Training Image Patterns for Spatial Simulation. *Mathematical Geology*, 38, 63-80. <u>https://doi.org/10.1007/s11004-005-9004-x</u>
- [9] Yin, Y., Zhang, C., Li, J., et al. (2011) Progress and Prospect of Multiple-Point Geostatistics. Journal of Palaeogeography, 13, 245-252.
- [10] Haldorsen, H.H. and Chang, D.M. (1986) Notes on Stochastic Shales; from Outcrop to Simulation Model. *Reservoir Characterization*, 445-485. https://doi.org/10.1016/B978-0-12-434065-7.50020-4
- [11] Lantuéjoul, C. (2002) Geostatistical Simulation. Models and Algorithms. *Minerva Ginecologica*, **39**, 503-510.
- [12] Deutsch, C.V. and Tran, T.T. (2002) FLUVSIM: A Program for Object-Based Stochastic Modeling of Fluvial Depositional Systems. *Computers & Geosciences*, 28, 525-535. https://doi.org/10.1016/S0098-3004(01)00075-9
- [13] Pyrcz, M.J., Boisvert, J.B. and Deutsch, C.V. (2009) ALLUVSIM: A Program for Event-Based Stochastic Modeling of Fluvial Depositional Systems. *Computers & Geosciences*, 35, 1671-1685. <u>https://doi.org/10.1016/j.cageo.2008.09.012</u>

- [14] Shi, S., Hu, S., Feng, W., et al. (2012) Building Geological Knowledge Database Based on Google Earth Software. Acta Sedimentologica Sinica, 30, 869-878.
- [15] Michael, H.A., Li, H., Boucher, A., et al. (2010) Combining Geologic-Process Models and Geostatistics for Conditional Simulation of 3-D Subsurface Heterogeneity. *Water Resources Research*, 46, 1532-1535. https://doi.org/10.1029/2009WR008414
- [16] Comunian, A., Jha, S.K., Giambastiani, B., et al. (2014) Training Images from Process-Imitating Methods. *Mathematical Geosciences*, 46, 241-260. https://doi.org/10.1007/s11004-013-9505-y
- [17] Zhang, W., Duan, T., Zheng, L., et al. (2015) Generation and Application of Three-Dimensional MPS Training Images Based on Shallow Seismic Data. Oil & Gas Geology, 36, 1030-1037.
- [18] Fadlelmula, F.M.M., Killough, J. and Fraim, M. (2016) Ti Converter: A Training Image Converting Tool for Multiple-Point Geostatistics. *Computers & Geosciences*, 96, 47-55. <u>https://doi.org/10.1016/j.cageo.2016.07.002</u>
- [19] Ortiz, J.M. and Deutsch, C.V. (2004) Indicator Simulation Accounting for Multiple-Point Statistics. *Mathematical Geology*, **36**, 545-565. https://doi.org/10.1023/B:MATG.0000037736.00489.b5
- [20] Feng, W., Wu, S., Yin, Y., et al. (2017) A Training Image Evaluation and Selection Method Based on Minimum Data Event Distance for Multiple-Point Geostatistics. *Computers & Geosciences*, 104, 35-53. <u>https://doi.org/10.1016/j.cageo.2017.04.004</u>
- [21] Ding, F., Lu, Y., Duan, D., et al. (2017) Complex Superimposition Type Sandbody Characterized by Multiple-Point Geostatistics Modeling Method. Complex Hydrocarbon Reservoirs, 10, 34-38.
- [22] Liu, K., Hou, J., Liu, Y., *et al.* (2016) Application of Multiple-Point Geostatistics in 3D Internal Architecture Modeling of Point Bar. *Oil & Gas Geology*, **37**, 577-583.
- [23] Chen, G., Zhao, F., Wang, J., et al. (2015) Regionalized Multiple-Point Stochastic Geological Modeling: A Case from Braided Delta Sedimentary Reservoirs in Qaidam Basin, NW China. Petroleum Exploration and Development, 52, 638-645. <u>https://doi.org/10.1016/S1876-3804(15)30065-3</u>