

# A Python Based Multi-Point Geostatistics by using Direct Sampling Algorithm

Edwin Brilliant<sup>1</sup>, Sanggeni Gali Wardhana<sup>1</sup>, Alissa Bilqis<sup>1</sup>, Alda Ressa Nurdianingsih<sup>1</sup>,  
Rafif Rajendra Widya Daniswara<sup>1</sup>, and Waskito Pranowo<sup>1,2,\*</sup>

<sup>1</sup>Department of Geophysical Engineering, Universitas Pertamina, Jakarta, Indonesia

<sup>2</sup>Center for Geosciences Artificial Intelligence and Advanced Computing, Universitas Pertamina, Jakarta

\*Email: [waskito.pranowo@universitaspertamina.ac.id](mailto:waskito.pranowo@universitaspertamina.ac.id)

*Submit: 2020-11-27; Revised: 2021-01-14; Accepted: 2021-01-17*

**Abstract:** Kriging is one of the main options when estimating the distribution of reservoir properties. In estimating, kriging uses a variogram which contains information on the spatial relationship between data. However, in some cases, kriging has drawbacks. In a similar configuration, the sampled data will produce the same variogram so that the kriging results can be similar. To overcome the shortcomings of kriging, Multi-Point Geostatistics (MPS) can be an alternative choice of complementary methods. MPS is a geostatistical method to estimate the value of unknown points by simultaneously utilizing several data points around it. In MPS, spatial relationships are presented as the training image (TI). TI is a geological conceptual model of the research area which is used as the initial model for estimating the unknown point. In this study, the MPS method was used using a Python-based direct sampling algorithm. The data used is synthetic data derived from snippets from TI. The TI used is in the form of a braided channel. By sampling data from TI, it can be seen that MPS's ability to reconstruct the value compared to the indicator kriging. The results of this study indicate that MPS can reconstruct braided channel cases better than indicator kriging.

**Keywords:** multi-point geostatistics, direct sampling, training image, Python

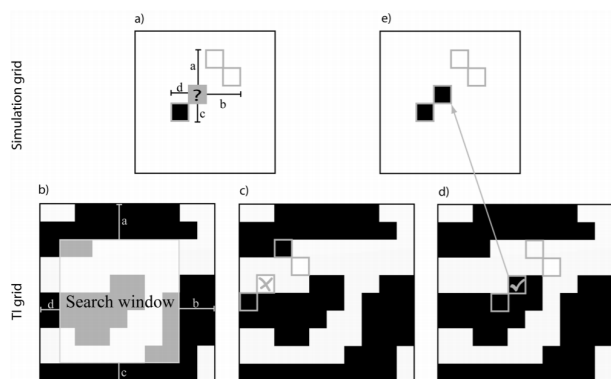
**Abstrak:** Kriging merupakan salah satu pilihan utama ketika melakukan estimasi penyebaran properti reservoir. Dalam melakukan estimasi, kriging menggunakan variogram yang berisi informasi hubungan spasial antar data. Namun, untuk beberapa kasus, kriging memiliki kelemahan. Pada konfigurasi sampled data yang mirip akan menghasilkan variogram yang sama, sehingga hasil kringingnya dapat mirip. Untuk mengatasi kekurangan kriging, Multi-Point Geostatistics (MPS) dapat sebagai alternatif pilihan metode pelengkap. MPS adalah metode geostatistika yang digunakan untuk mengestimasi nilai dari titik yang belum diketahui, dengan memanfaatkan beberapa data titik di sekitarnya secara bersamaan. Pada MPS, hubungan spasial disajikan sebagai training image (TI). TI adalah model konseptual geologi daerah penelitian yang digunakan sebagai model awal untuk estimasi titik yang belum diketahui tersebut.

Pada penelitian ini, dilakukan penggunaan metode MPS dengan menggunakan algoritma direct sampling berbasis Python. Data yang digunakan adalah data sintetik yang berasal dari cuplikan dari TI. TI yang digunakan berbentuk braided channel. Dengan mencuplik data dari TI, maka dapat dilihat kemampuan MPS dalam merekonstruksi ulang nilai dibandingkan dengan indicator kriging. Hasil dari studi ini menunjukkan bahwa MPS dapat merekonstruksi kasus braided channel lebih baik daripada indicator kriging.

**Kata kunci:** multi-point geostatistics, direct sampling, training image, Python

## 1 INTRODUCTION

In the oil & gas exploration and production phase, geostatistical methods are required to see the distribution of reservoir property data in the area of interest. The covariance-based geostatistical method, such as kriging, is the standard method in the industry. This method is a two-point statistical method that relies on the variogram model to form a spatial relationship. In real data, different spatial models may produce the same variogram model. Multi-Point Geostatistics (MPS) method that can perform training data using all wells data simultaneously to experiment with the distribution of reservoir properties. In MPS, the geological concept can be inputted by using the training image (TI). The training image acts as a spatial relationship in a certain area, so different areas may have unique spatial relations. In this project, the MPS technique is performed by using the Direct Sampling (DS) algorithm (Mariethoz, 2009) that requires a training image that represents the geological conditions of the interest zone. The prediction by using MPS or covariance-based geostatistical methods should be seen as complementary approaches (Mariethoz, 2018). The two methods are complementary because they can solve various types of problems distinguished based on the nature and amount of information available. Apart from making predictions using the MPS with the Direct Sampling algorithm, this project will also compare the MPS method with the Indicator Kriging method. This comparison is carried out to show the best reconstruction of reservoir property on the synthetic sample data used in this project.



**Figure 1.** Illustration of Direct Sampling Method (Mariethoz et al., 2010). (a) The simulation grid or the area that we want to estimate, (b) the searching window on TI, (c) the process the window on TI in searching estimation value, (d) the searching window gets the best estimation value, and (e) the estimation value is taken to the simulation grid.

## 2 DATA AND METHODOLOGY

### 2.1 MPS

MPS is performed by defining a model based on different training image data from the variogram that is determined by the relationship between the data. Training image is a collection of data in the form of a geological conceptual model in the research area that is used as a basis or reference for the estimation process using available data. The use of the training image itself as a framework for an algorithm is to estimate values for the unsampled location. The estimation algorithm will use a training image as an input model to obtain certain patterns to estimate values at the unsampled locations. The unsampled and sampled locations are located in an area called the simulation grid. Sampled locations are the position/location which already has data, while unsampled locations are the location with no data and expected to be estimated.

### 2.2 Direct Sampling Algorithm

One of the methods used by MPS in retrieving data is a direct sampling (Figure 1), from Mariethoz (2009) with an illustration as followed:

- (i) Take the point to estimate, called an unsampled location, on the simulation grid. Two white and black pixels represent the sampled nodes, while the 'question mark' is the pixel or point of unsampled location. This point will be simulated or estimated.
- (ii) Determine the search window on the TI grid by matching dimensions in figures a, b, c, and d (Figure 1a). The data at the sampled locations in the search-window will be called neighbor data.
- (iii) Scan the search window starting from a random location on TI,
- (iv) The simulation data events are matched until a satisfactory result is obtained. The criterion of the match is if the configuration's error below a threshold.
- (v) Take the value on TI if there is a matched configuration.

Then, place the value on the simulation grid. This value is the estimation of the unsampled location.

This algorithm's advantage is that when you find a suitable configuration, the simulation point's value will be directly filled in by the train data, without calculating any probabilities. However, if you do not find suitable data, the distance/error concept is used, and it is taken with the smallest error value. The equation used in the direct sampling algorithm (Mariethoz et al., 2010) is as follows,

$$d = \frac{1}{n} \sum_{i=1}^n a_i \quad (1)$$

$$a_i = \begin{cases} 1, & \text{if } Z(x_i, y_i) \neq TI(X_i, Y_i) \\ 0, & \text{if } Z(x_i, y_i) = TI(X_i, Y_i) \end{cases} \quad (2)$$

where  $d$  is the distance,  $n$  is the number of neighbor data in a search window,  $a_i$  is the similarity between neighbor data on the simulation grid and data on training image,  $Z(x_i, y_i)$  is neighbor data in the search window on simulation grid,  $TI(X_i, Y_i)$  is data in the search window on the training image. The estimated value at an unsampled location is obtained if

$$d < t \quad (3)$$

where  $t$  is the error threshold. After the estimated value at an unsampled location is obtained, the searching window will stop, and the algorithm will continue to estimate the next unsampled location. This simulation is carried out by paying attention to several parameters that must be adjusted, including:

- Simulation size Simulation size is the size or area of the desired simulation results.
- Template size Template size is a window search of the DS algorithm. If the window is small, fewer neighbor data will be involved in the estimation.
- Fraction of TI TI's fraction is the maximum number of iterations or searches for data from the training image. For example, if the TI points are 300 data, and the fraction of TI is 0.3, then the searching window will search pattern on TI as many as 90 points of search. As a consequence, if the fraction of TI value increases, the iteration will increase, requiring a long-running time but with better results.
- Error threshold The threshold is the error tolerance of the distance (Equation 1 and 2). If the threshold is high, data with less similarity on TI will be considered as estimated data. However, if it is too small, it may not get the desired estimated value. If this happens, DS will take the lowest distance along with the iterations even though it is higher than the threshold.

### 2.3 Indicator Kriging

Indicator Kriging (IK) as a technique for spatial-estimation was introduced by Journel (1983). Based on Switzer (1977), Indicator Kriging is an algorithm that can estimate a local uncertainty by decreasing the local cumulative distribution function (CDF). This Indicator kriging technique is a non-linear geostatistical technique used in the mineral and oil

and gas industries, for instance, to estimate facies distribution (Kelkar et al., 2002; Caceres et al., 2010). The essence of the Indicator Kriging is the coding of the binomial data to be 1 or 0. It depends on boundary value or commonly referred to as the threshold value ( $x_t$ ) for a given value  $x(\vec{u}_j)$ . The following is the equation used in performing indicator transform (Kelkar et al., 2002). In the continuous case, the following equation applies:

$$I(\vec{u}_j, x_t) = \begin{cases} 1, & \text{if } x(\vec{u}_j) \leq x_t \\ 0, & \text{if } x(\vec{u}_j) > x_t \end{cases} \quad (4)$$

where  $I(\vec{u}_j, x_t)$  and  $x(\vec{u}_j)$  are the indicator value and the original data at  $\vec{u}_j$  position respectively, and  $x_t$  is the threshold. Based on Equation 4, the data's non-linear transformation will be a value of 0 or 1. A data with a value greater than the threshold value will get an indicator value of 0, and if the data value is smaller than the threshold, it will have an indicator value of 1. Thus, the transformation of data indicators is a relatively effective way of limiting the effect of very high values. In the overall process of indicator kriging after getting the indicator value, the next step is to estimate the indicator value at an unsampled location. Indicator kriging uses ordinary or simple kriging formulas to estimate the values at unsampled locations. The following is the equation used to perform indicator kriging by using ordinary kriging formula (Kelkar et al., 2002),

$$I^*(\vec{u}_0, x_t) = \sum_{i=1}^n \lambda_i I(\vec{u}_i, x_t) \quad (5)$$

where  $I^*(\vec{u}_0, x_t)$  is the estimated value at an unsampled location,  $I(\vec{u}_i, x_t)$  is the neighbor data, and  $\lambda_i$  is the weight of neighbor data. Basically, indicator kriging is ordinary or simple kriging with the input data transformed by using indicator transform. Therefore,  $\lambda_i$  are obtained by solving the ordinary or simple kriging process. The results of indicator kriging will be ranging between 0 and 1 for each point estimate.

### 3 RESULTS: ESTIMATING VALUES BY MPS AND ITS COMPARISON WITH INDICATOR KRIGING

There are two inputs in this research:

- Training Image (TI) In this study, the training image in the form of a braided channel (Figure 2) was taken from Mariethoz (2009).
- Sampled data There are 36 points that are sampled (Figure 3) from the training image in Figure 2. We choose the sample data as many as possible. So, we can evaluate whether the two methods can reconstruct the channel again or not.

In the initial experiment, the MPS simulation is using a simulation size of  $80 \times 80$ , a searching window of  $30 \times 30$ , a fraction of TI of 0.03, and a threshold of 0.01 (1% of error). The values of fraction of TI and threshold are obtained from one test run. We choose the values that give the best estimation. After we got the parameters, the estimations are

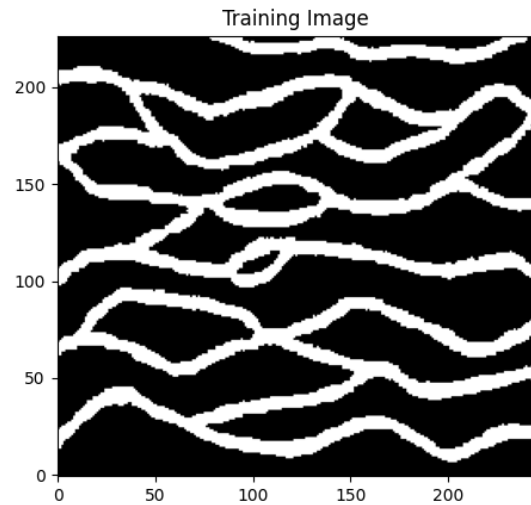


Figure 2. Training Image (Mariethoz, 2009). White color and the value of one represents sand, while black color and the value of zero is representing shale.

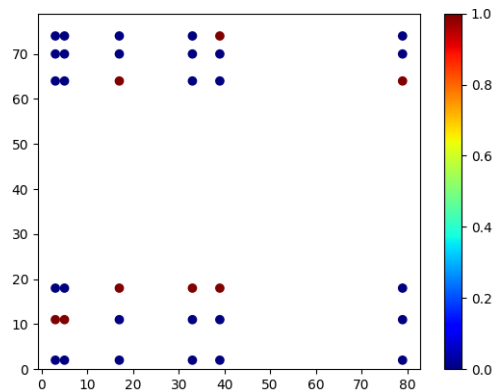


Figure 3. Sampled data. One represents the channel (blue color), while zero represents the non-channel (red color).

run in several simulations. The more simulation will give a better result. In this paper, it is chosen 50 simulations. One of MPS simulation results is shown in Figure 4. It is important to emphasize that TI's size could be different from the size of the estimation area. It can be bigger or smaller.

After performing 50 simulations, the simulation results are averaged (Figure 5a). The value after averaging will be ranging between 0 and 1. The highest value, 1, may indicate the highest probability of channel occurrence, while 0 may indicate the lowest probability of channel occurrence (or zero probability).

### 4 DISCUSSION AND CONCLUSION

As a comparison, we also perform indicator kriging. We choose indicator kriging because it produces probability

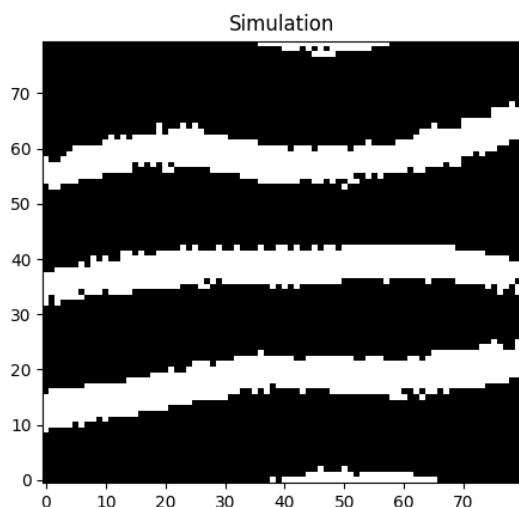


Figure 4. An example of a simulation result.

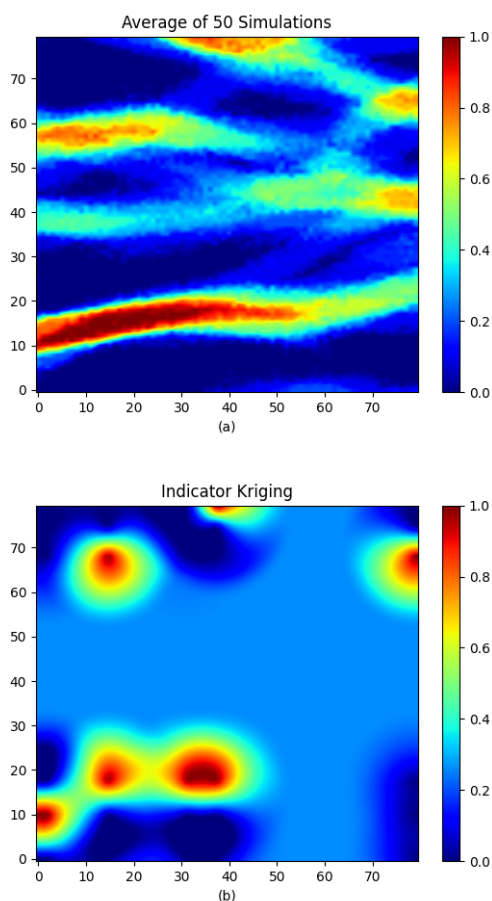


Figure 5. (a) The result of the average of MPS with 50 Simulations and (b) the result of Indicator Kriging

value like MPS result. Indicator kriging cannot reconstruct well the channel's shape (Figure 5b). In kriging, we cannot input the interpretation of geological features, represented by training image. Kriging uses variogram and covariance to estimate the value at unsampled locations. However, the different geological features could give similar variograms if the sampled data configuration is not good enough to represent the geological features. MPS simulation by using a direct sampling algorithm can estimate reservoir property with its geological features. The geological features model is inputted as a training image and act as the guidance to estimate the property. In contrast, indicator kriging uses a variogram, which may be ambiguous with different geological features. This research shows that MPS gives better reconstruction in braided channel features rather than indicator kriging.

#### ACKNOWLEDGMENT

The authors thank Dr. Darharta Dahrin and an anonymous reviewer for careful reading the manuscript.

#### CODE AND MATERIALS

The python code and all materials to re-produce the results in this study are available at <https://github.com/waskitoprano/statistics/tree/main/geostatistics>.

#### References

- Caceres, A., Emery, X. and Riquelme, R. (2010): Truncated gaussian kriging as an alternative to indicator kriging. *The 4th International Conference on Mining Innovation*.
- Journel, A.G. (1983): Nonparametric estimation of spatial distributions. *Journal of the International Association for Mathematical Geology*, ISSN 00205958, doi:10.1007/BF01031292.
- Kelkar, M., Perez, G. and Chopra, A. (2002): Applied geostatistics for reservoir characterization. Society of Petroleum Engineers, ISBN 978-1-55563-095-9.
- Mariethoz, G. (2009): Geological stochastic imaging for aquifer characterization. Ph.D. thesis, Universite of Neuchatel.
- Mariethoz, G. (2018): When should we use multiple-point geostatistics? In: *Handbook of Mathematical Geosciences: Fifty Years of IAMG*. Springer International Publishing, ISBN 9783319789996, doi:https://doi.org/10.1007/978-3-319-78999-6\_31.
- Mariethoz, G., Renard, P. and Straubhaar, J. (2010): The direct sampling method to perform multiple-point geostatistical simulations. *Water Resources Research*, **46**, ISSN 00431397, doi:https://doi.org/10.1029/2008WR007621.
- Switzer, P. (1977): Estimation of spatial distributions from point sources with application to air pollution measurement. technical report no. 9. Tech. rep., Stanford Univ., CA (USA). Dept. of Statistics.