

SAEED SOLTANI-MOHAMMADI\*, MOHAMMAD SAFA\*

**A SIMULATED ANNEALING BASED OPTIMIZATION ALGORITHM FOR AUTOMATIC VARIOGRAM MODEL FITTING****SYMULACJA ALGORYTMU OPTYMALIZACYJNEGO PROCESU ODPREŻANIA DLA AUTOMATYCZNEGO DOPASOWANIA MODELU WARIOGRAMU**

Fitting a theoretical model to an experimental variogram is an important issue in geostatistical studies because if the variogram model parameters are tainted with uncertainty, the latter will spread in the results of estimations and simulations. Although the most popular fitting method is fitting by eye, in some cases use is made of the automatic fitting method on the basis of putting together the geostatistical principles and optimization techniques to: 1) provide a basic model to improve fitting by eye, 2) fit a model to a large number of experimental variograms in a short time, and 3) incorporate the variogram related uncertainty in the model fitting. Effort has been made in this paper to improve the quality of the fitted model by improving the popular objective function (weighted least squares) in the automatic fitting. Also, since the variogram model function ( $\epsilon$ ) and number of structures ( $m$ ) too affect the model quality, a program has been provided in the MATLAB software that can present optimum nested variogram models using the simulated annealing method. Finally, to select the most desirable model from among the single/multi-structured fitted models, use has been made of the cross-validation method, and the best model has been introduced to the user as the output. In order to check the capability of the proposed objective function and the procedure, 3 case studies have been presented.

**Keywords:** Automatic variogram fitting, Geostatistics, Optimization, simulated annealing

Dopasowanie modelu teoretycznego do eksperymentalnego wariogramu jest kluczowym zagadnieniem w badaniach geostatystycznych ponieważ jeśli parametry modelu wariogramu obciążone są niepewnością, to otrzymamy znaczny rozrzut wyników obliczeń i symulacji. Pomimo, że najpopularniejszą metodą dopasowania jest dopasowanie 'na oko', w niektórych przypadkach wykorzystuje się automatyczne metody dopasowania modelu oparte na zasadach geostatystyki i optymalizacji w celu: 1) dostarczenia podstawowego modelu do dopasowania 'na oko'; 2) dopasowania modelu do większej ilości eksperymentalnych wariogramów w krótkim okresie czasu; 3) uwzględnienia niepewności związanej z wariogramem w dopasowaniu modelu. W pracy podjęto próbę poprawy jakości dopasowania modelu poprzez wprowadzenie zmodyfikowanej popularnej funkcji celu (ważone najmniejsze kwadraty) do au-

\* UNIVERSITY OF KASHAN, DEPARTMENT OF MINING ENGINEERING, KASHAN, IRAN

tomatycznego dopasowania. Ponadto, ponieważ funkcja modelu wariogramu ( $L$ ) i ilość struktur ( $m$ ) ma także wpływ na jakość modelu, opracowano program w środowisku MATLAB który podaje optymalne modele wariogramu w oparciu o metodę symulacji odprężania. W części końcowej wybrano najkorzystniejszy model spośród modeli dopasowania z wykorzystaniem metody walidacji krzyżowej i najlepszy model przedstawiany jest użytkownikowi. W celu zbadania możliwości stosowania proponowanej funkcji celu i przedstawionej procedury, zaprezentowano trzy studia przypadku.

**Słowa kluczowe:** automatyczne dopasowanie wariogramu, geostatystyka, optymalizacja, symulacje procesu odprężania

## 1. Introduction

A key stage in geostatistical studies is the structural analysis of the regionalized variable to describe the spatial variations of the desired variable (grade, thickness, accumulation, etc.) (Chiles & Delfiner, 2012). It is the variogram that characterizes the spatial variability of a regionalized variable (Pardo-Igúzquiza, 1999). Generally, the variogram of a regionalized variable is unknown and the experimental variogram should be calculated using the data gathered from the regionalized variable. Using the moments method, the experimental variogram is calculated as follows (Matheron, 1965):

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{k=1}^{N(h)} [z(x_i) - z(x_j)]^2 \quad (1)$$

where  $\hat{\gamma}(h)$  is the value of the experimental variogram for the vector  $h = |x_i - x_j|$ ,  $N(h)$  is the number of data pairs for  $h$ , and  $z(x_i)$  and  $z(x_j)$  are the observed or experimental value at points  $x_i$  and  $x_j$ . After calculating the experimental variogram, it is necessary that an appropriate model be fitted to it; this is a basic step in geostatistics (Journel & Huijbregts, 1978, Christakos 1984) which largely affects the computation of the krigging weights (Gorsich & Genton, 2000) and, hence, it affects the efficacy of the maps obtained from the krigging estimator and the geostatistical simulation (Genton, 1998a). Although fitting by eye (a procedure in the graphical geostatistical software based on trial and error) is still the most common method of fitting a model to an experimental variogram, with the increasing use of optimization methods in geostatistical studies in the past and recent years, much effort has also been made regarding the automatic fitting of the optimum model to the experimental variogram. Automatic fitting is carried out: 1) to provide a basic model to enhance fitting by eye, 2) when it is necessary that a model be fitted to a large number of experimental variograms in a short time, and 3) when it is necessary that the uncertainty related to the experimental variogram be considered in fitting a model to it (Pardo-Igúzquiza, 1999). The history goes back to formulating the problem of fitting a model to an experimental variogram based on such criteria as least squares (David, 1977; Cressie, 1979), generalized least squares (Cressie & Hawkins, 1980, Taylor & Burugh, 1986, Genton, 1998b), weighted least squares (Starks & Fang, 1982, Cressie, 1985, Jian et al., 1996), and weighted and ordinary least squares (McBratney & Webster, 1986).

In the first program developed for the automatic fitting of the variogram model, use was made of the weighted least squares criterion for the modeling of the objective function and the problem was solved using the quadratic programming method after incorporating such constraints as “monotonicity”, “smoothness” and “convexity” (Shapiro & Botha, 1991). Later, use was

made of the least squares criterion and the simplex optimization method in the VARFIT program (Pardo-Igúzquiza, 1999) developed in the FORTRAN environment. Genton (Gorsich and Genton 2000) made an effort in non-parametric estimation of variogram and its derivatives for the fitting of the appropriate model in the Matlab environment (version 5). Paula & Clayton (2003) enhanced the objective function of the experimental variogram considering lags intervals and the number of pairs at every point (Paula F. Larrondo et al., 2003). Desassis (Desassis & Renard, 2013) developed an algorithm, based on the numerical algorithm, to find the optimum model based on the objective function (5) for single/multi-variable cases. Many of the former methods have been developed to fit a model to a direct experimental variogram, but Emery (Emery, 2010) studied the fitting of a coregionalization model extensively and proposed 3 new algorithms for unknown sill, known sill, and Plurigaussian models. The objective function in his studies was weighted least squares and use was made of the extended Goulard-Voltz, Simulated Annealing, and Levenberg-Marquardt methods respectively to find the optimal solutions.

Effort has been made in this paper to fit the best theoretical nested variogram model to an experimental variogram. The fitted models have been selected from among spherical, exponential, gaussian, and cubic models that can be 1, 2, or 3-structured. Since most simulation algorithms and geostatistical estimations need a variogram model, this research makes possible for the geoscience researchers to select, from among the many possible cases, the most appropriate variogram model with due consideration to the number of structures.

## 2. Materials and methods

### 2.1. Problem formulation

For the random variable  $Z$ , the experimental variograms  $\hat{\gamma}$  are calculable for different vector lags  $h_i; i = 1, \dots, n$ , and the optimum model  $\gamma$  should be fitted to it. An appropriate model is one that its value has the least distance from those of the experimental variogram  $\hat{\gamma}(h_i)$  in every lag  $h_i$  or, in other words, the value of the following objective function should be a minimum (David, 1977; Journel & Huijbregts, 1978; Clark, 1979):

$$\text{Minimize } \sum_i \left( 2\hat{\gamma}(h_i) - 2\gamma(h_i) \right)^2 \quad (2)$$

Since the amount of the experimental variogram uncertainty is different for different lags, the weights of all the lags should not be the same (Webster & Oliver, 2001). The variogram behavior near the origin specifies the spatial continuity of the regionalized variable, and the quality of fitting a model to the primary lags is more important compared to the intermediate and end ones. Therefore, it is necessary that this point be considered in the objective function in the form of adding the value of the theoretical variogram, for the desired lag, in the denominator. The result of this correction is that the more the lag distance increases from the origin, the less its effects will be on the value of the objective function (Fig. 1).

The number of existing pair points for every lag, on the basis of which the experimental variogram has been calculated, is another point to consider; the more the number of pair points, the more the reliability of the value of the experimental variogram. This point can be applied in the objective function using the coefficient of the number of existing pair points for every lag.

In 1985, Cressie (Cressie, 1985) modified the above objective function in the following form to be able to consider the effects of the above 2 peculiarities:

$$\text{Minimize } \sum_{j=1}^k \frac{N_{h(j)}}{\gamma^2(h(j); \lambda)} (\hat{\gamma}(h(j)) - \gamma(h(j); \lambda))^2 \tag{3}$$

where  $N_{h(j)}$  is the number of pairs at every point in calculating the variogram,  $\hat{\gamma}(h(j))$  is the value of the experimental variogram in  $h(j)$ , and  $\gamma(h(j); \lambda)$  is the value of the theoretical variogram for the estimated parameters ( $\lambda$ ). Webster & McBratney 1989 studied the effects of adding these weights and concluded that adding improved the results.

$$\sum_{j=1}^k \frac{N_{h(j)}}{\sqrt{h(j)} \times \gamma^2(h(j); \lambda)} (\hat{\gamma}(h(j)) - \gamma(h(j); \lambda))^2 \tag{4}$$

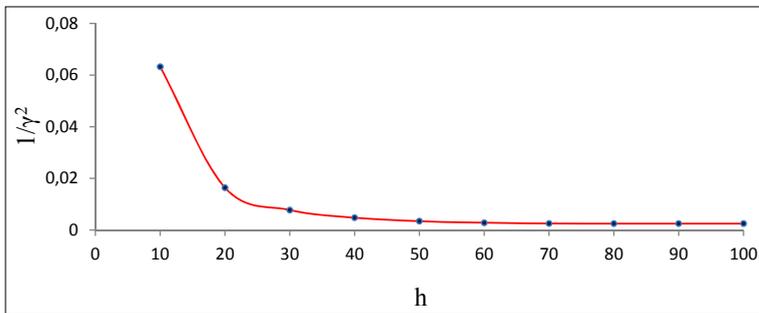


Fig. 1. Trend of the reduction of the value of  $1/\gamma^2$  with respect to the lag distances for the spherical model with a sill and range of 20 and 75 respectively

In fitting the optimum model, besides variogram model parameters (nugget effect, sill and range), the variogram model function  $\mathcal{L}$  and the number of the structures  $m$  are effective too. The rate of the effect of the variogram function increases especially in nested models where every structure can follow one of the models.

In many experimental variograms, it is observed that in end lags (after reaching the value of the sill), the trends are rising, falling or sinusoidal (Fig. 3); therefore, the value of  $(\hat{\gamma}(h(j)) - \gamma(h(j); \lambda))^2$  increases considerably. After range, coefficient  $\frac{1}{\gamma^2(h(j); \lambda)}$  reaches a constant number for spherical, gaussian and cubic models (for the exponential model it tends to a constant value), and  $N_{h(j)}$  does not vary much in the end lag; therefore, coefficient  $\frac{N_{h(j)}}{h(j) \times \gamma^2(h(j); \lambda)}$  cannot compensate for the increase in  $(\hat{\gamma}(h(j)) - \gamma(h(j); \lambda))^2$  and causes the sill of the fitted optimum model to distort. To adjust this increase, it is necessary that a “distance” function be added to the objective function. Now, since by adding  $h$  in the denominator of the objective function the weights allocated to the preliminary points will be illogically very large (compared to other points) (Fig. 2), to solve this problem,  $h$  will be substituted with  $\sqrt{h}$  in the denominator of relation 3 and the objective function will be rewritten as follows:

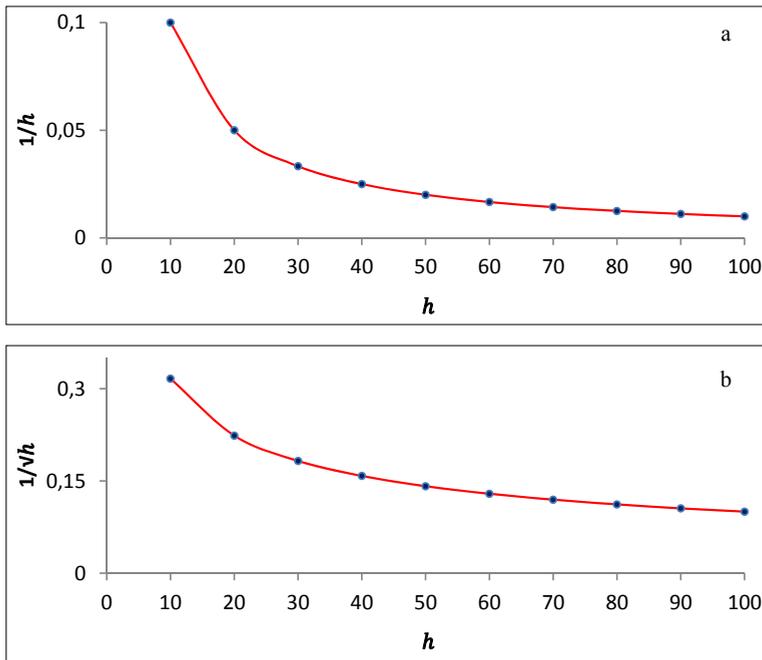


Fig. 2. a) Reduction trend of  $1/h$  with respect to the lags distances, b) Reduction trend of  $1/\sqrt{h}$  (both with respect to lags distances)

Effects of the modified component (applied in the objective function) on drawing near the sill of the fitted variogram model to the real values are shown in Fig. 3. Fig. 3-a is the model fitted based on relation 3 wherein the value of the sill of the fitted variogram model is smaller than the desirable value due to the falling trend in the end lags and Fig. 3-b is related to the model fitted using the modified objective function of relation 4 wherein the value of the sill of the fitted model has increased and looks more desirable.

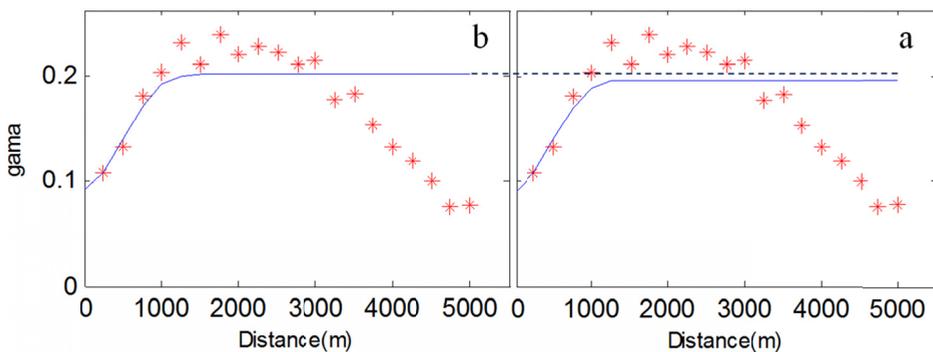


Fig. 3. Models fitted to the experimental variogram: a) with the Cressie weighting relation and, b) with the modified relation

## 2.2. Decision variables

Decision variables in this research are the variogram model parameters including the nugget effect, sill, and range, and are modeled in the form of the linear  $\psi_{((2n+1),1)}$  where  $n$  is the number of the structures of the variogram model. Its first array is the nugget effect and arrays  $2i$  and  $2i+1$  are respectively the sill and the range of the  $i^{\text{th}}$  structure. Another decision variable that affects the output of the algorithm is the variogram model selected for every one of the structures. The possible structures considered in the present study are spherical, gaussian, exponential, and cubic (their specifications are given in Table 1). Since the variogram model type was not definable in the form of a quantitative decision variable, the optimization problem was solved based on different arrangements of the type of model used for the variogram model structures; the better options were then selected based on the value of the objective function.

TABLE 1

Specifications of the common variogram models

Spherical model	$\begin{cases} \gamma(h) = C \left[ \frac{3}{2} \left( \frac{h}{a} \right) - \frac{1}{2} \left( \frac{h}{a} \right)^3 \right] & \text{for } h < a, \\ \gamma(h) = C & \text{for } h \geq a, \end{cases}$
Exponential model	$\gamma(h) = C \left[ 1 - \exp\left(\frac{-3h}{a}\right) \right]$
Gaussian model	$\gamma(h) = C \left[ 1 - \exp\left(-\left(\frac{3h}{a}\right)^2\right) \right]$
Cubic model	$\begin{cases} \gamma(h) = C \left[ 7 \left( \frac{h}{a} \right)^2 - \frac{35}{4} \left( \frac{h}{a} \right)^3 + \frac{7}{2} \left( \frac{h}{a} \right)^5 - \frac{3}{4} \left( \frac{h}{a} \right)^7 \right] & \text{for } h < a, \\ \gamma(h) = C & \text{for } h \geq a, \end{cases}$

## 2.3. Simulated annealing

The simulated annealing (SA) algorithm is a simple, effective, and meta-heuristic optimization algorithm for the solution of NP optimization problems. Similar to most meta-heuristic algorithms, it has been established by modeling and simulating one of the nature's laws or phenomena. It has been presented based on substituting physical elements in the process of physical annealing (system state, state variation energy, temperature, and freezing state) with the elements of the optimization problem (possible solution, cost, neighborhood solution, controlling parameter, and the heuristic solution) (Kirkpatrick et al., 1983). This algorithm consists of 2 basic mechanisms: 1) producing the substitute, and 2) an acceptance rule (Lee & El-Sharkawi, 2008). To solve an optimization problem, the SA algorithm first starts with a primary solution and then, in an iteration loop, moves to the neighboring solution. If the latter is better than the

current one, the algorithm substitutes it (moves to it); otherwise, it will accept it as the present solution with a probability  $P$  which is found as follows:

$$P = e^{-\frac{\Delta f}{T}} \quad (5)$$

where  $\Delta f$  is the difference between the objective function of the present solution and that of the neighboring one, and  $T$  is the temperature. Many iterations are run at every temperature and then the temperature is reduced gradually. In the preliminary steps, very high temperatures are adjusted

so that worse solutions can be more probable to accept. With the gradual decrease in temperature, there is less probability for the worse solution to be accepted in the final steps; therefore, the algorithm converges on a near-optimal solution. The algorithm flow-chart is given in Fig. 4. The simulated annealing algorithm has been used to find the optimal solution of such different mining problems as optimally locating the additional drillholes (Soltani & Hezarkhani, 2009; Soltani-Mohammadi & Hezarkhani, 2013), calculating the permissible charge weight for blasting operations (Soltani-Mohammadi et al., 2012), geophysics studies (Luo et al., 2013), controlling and guiding exploration and extraction operations (Debba et al., 2009; Xia et al., 2011), and simulating and estimating (Deutsch & Journel, 1992; Lee & Ellis, 1997; Peredo & Ortiz, 2011). Reasons for their high usage are: 1) they do not have to incorporate problem assumptions, and 2) they rapidly find the near optimal global solution. To run the simulated annealing algorithm, it is necessary that, first, such annealing parameters as the annealing function, the temperature updating function, and the initial temperature be specified. The annealing function is either “Boltzmann” or “fast”, and the temperature updating function is selected from among exponential, logarithmic, and linear functions. The initial temperature can be defined both as a number or a function.

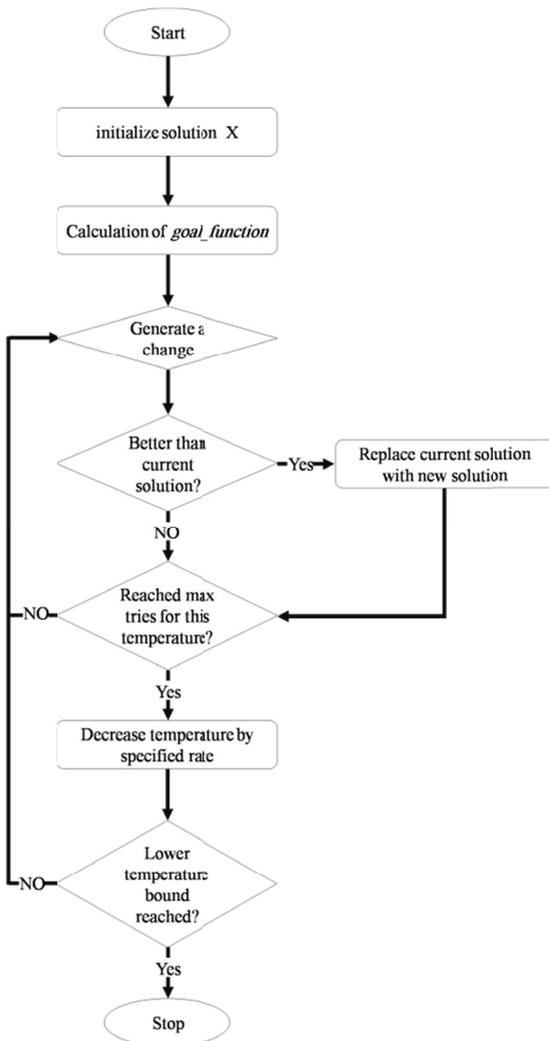


Fig. 4. Simulated annealing algorithm

### 3. Case studies

Three case studies have been presented to assess the capability of the proposed algorithm.

#### 3.1. Case study 1

This case study is related to the automatic fitting of a model suitable to an omni-directional experimental variogram of an  $\text{Al}_2\text{O}_3$  grade dataset with a variance equal to 21.3. This experimental variogram has been calculated in 20 lags with an initial lag of 30 m. Fig. 6 shows this experimental variogram together with the number of pair points in every lag. Running the algorithm necessitates the determination of the preliminary parameters of the annealing; to determine these parameters, different cases were studied. Fig. 5 shows the effects of the selection of the annealing function on the process of optimization. As shown, if the “fast” function is selected as the annealing function: 1) the time needed to reach the near-optimal solution will be reduced (Fig. 5c), and 2) it will be less probable for the solution to be trapped in a local optimal solution (Fig. 5(a and b)). A linear function was selected for temperature updating, and the initial annealing temperature was taken to be 300 degrees.

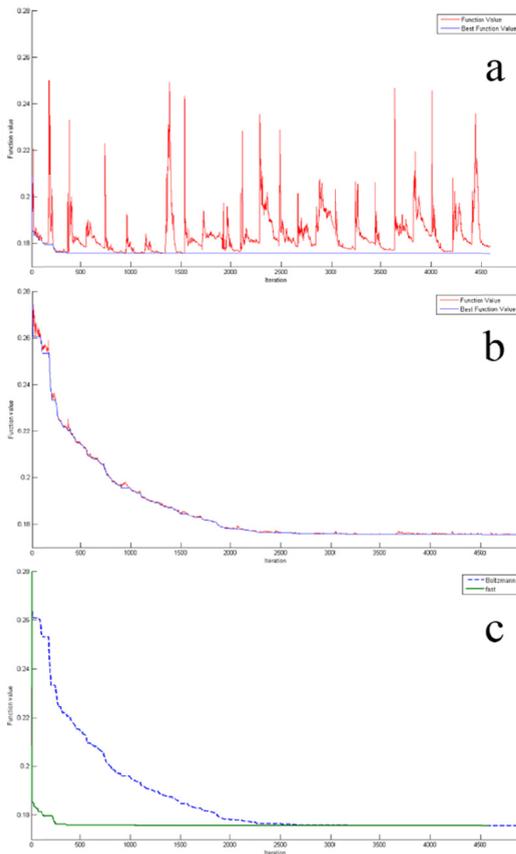


Fig. 5. Results from running the simulated annealing algorithm: a) “fast” annealing function, b) Boltzmann annealing function, and c) comparison of a) and b) regarding the best value of the objective function

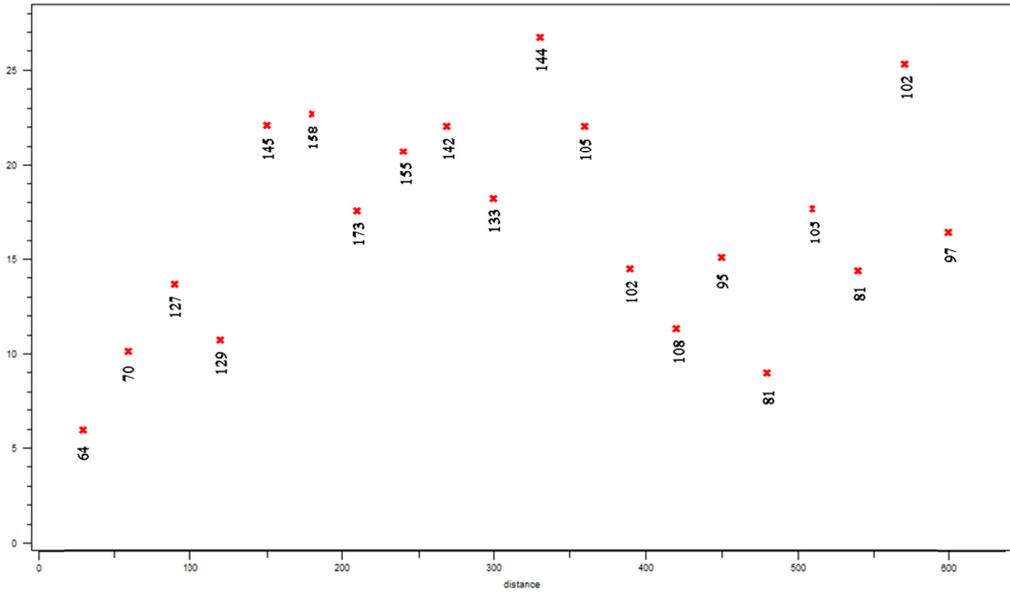


Fig. 6. The exponential variogram of the first dataset

Fig. 7 shows the results from running the algorithm; 4 fitted, single-structured, optimum models can be seen in the first row. As shown, the spherical (7-1), cubic (7-2), and gaussian (7-3) models are fitted acceptably, but the exponential models (7-4) are not so. Four of the best 2-structured and 3-structured models fitted to the experimental variogram are given in rows 2 and 3. As shown, in Figs. (7-5) to (7-11) (except Fig. (7-6)), one structure in all the nested models is cubic. Although using the cubic model has had acceptable results in the fitting of the variogram model, it should be noted that it cannot be used in such business software brands as “Datamine”, “Surpac”, etc and in non-business ones such as “gslib”, “Wingslib”, SGeMS, etc.

Cross-validation was carried out in this research based on the 12 better models (Fig. 7), and the results for Mean Error (ME), Mean Squared Error (MSE), Correlation Coefficient (CC), and Combined Error (CE) are given in Table 2. ME, MSE, and CE values are found as follows:

$$ME = \frac{1}{N} \sum_{i=1}^N \{Z(x_i) - \hat{Z}(x_i)\} \tag{6}$$

$$MSE = \frac{1}{N} \sum_{i=1}^N \{Z(x_i) - \hat{Z}(x_i)\}^2 \tag{7}$$

$$Combined\ Error = (1 - abs(CC)) + MSE + abs(ME) \tag{8}$$

where, for the case being studied,  $Z(x_i)$  is the real value of the component (grade, thickness, etc),  $\hat{Z}(x_i)$  is the estimated value of the component based on the fitted variogram model, and  $N$  is the number of points

The least CE value is related to the single-structured gaussian model (5.1+15.6 gaussian (177)) which is shown in Fig. 7-3 as the selected model and marked with an asterisk (\*) at the left corner. The worst among these 12 models is the sgle-structured exponential model.

TABLE 2

CC, MSE, ME, and CE values for the first dataset

n.	CC	MSE	ME	Combined Error
1	0.35	23.19	0.088	23.94
2	0.21	24.46	-0.119	25.36
3	0.39	19.61	0.031	20.25
4	0.14	29.22	0.038	30.11
5	0.24	20.79	0.097	21.56
6	0.21	21.90	-0.259	22.96
7	0.27	23.81	-0.124	24.66
8	0.04	23.71	-0.118	24.79
9	0.16	22.42	-0.159	23.43
10	0.21	21.41	-0.267	22.46
11	0.21	21.24	-0.153	22.18
12	0.24	21.11	-0.219	22.09

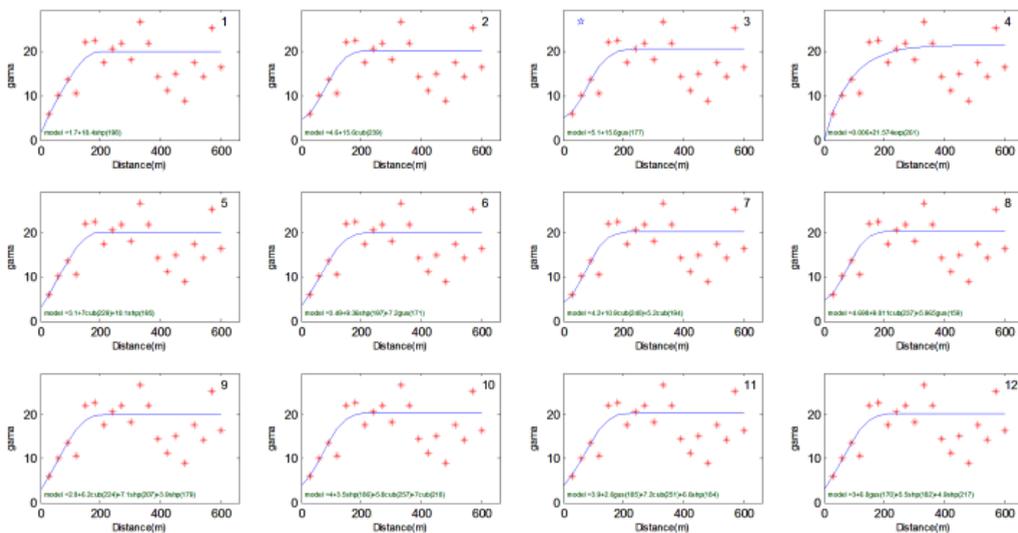


Fig. 7. Models proposed for the basic model of fitting by eye for the first set of data

### 3.2. Case study 2

Effort has been made here to automatically fit a model suitable to an Omni-directional experimental variogram of an S<sub>1</sub>O<sub>2</sub> grade dataset with a variance equal to 8.2. This experimental variogram has been calculated in 18 lags with an initial lag length of 40 m. Fig. 8 shows the experimental variogram together with the number of pair points in every lag.

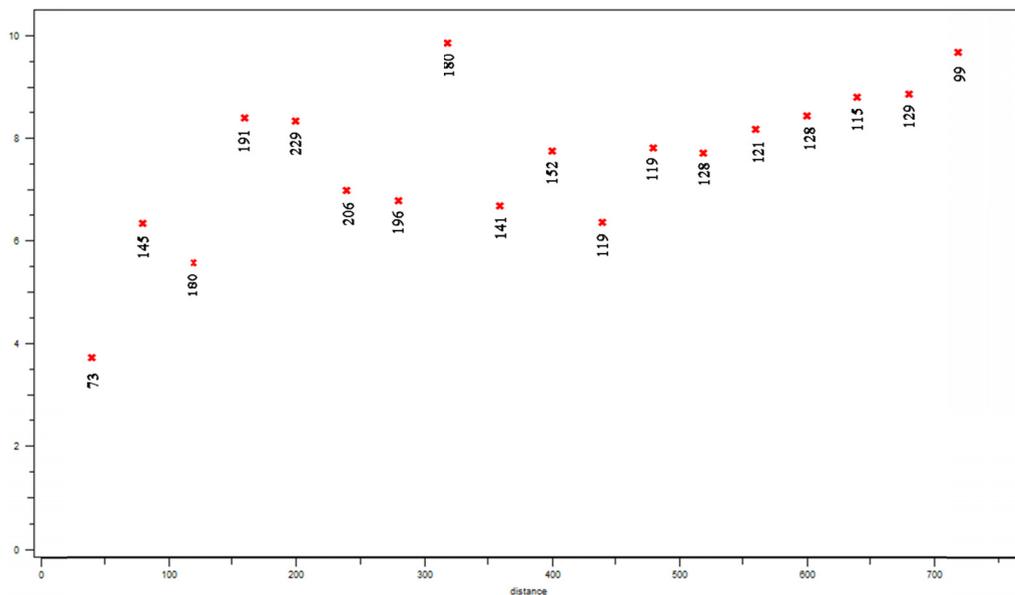


Fig. 8. The experimental variogram of the second set of data

Fig. 9 shows the results of running the algorithm. There are 4 fitted, single-structured, optimum models shown in the first row. As shown, the exponential model (9-1) has been fitted well, but, considering the structure of the experimental variogram, the spherical (9-2), cubic (9-3), and gaussian (9-4) models have not been fitted well. Four best 2-structured and 3-structured models fitted to the data are shown respectively in rows 2 and 3 of the figure. Similar to case study 1, in some cases, with an increase in the number of structures, the quality of the fitted model has somewhat improved, but it is negligible due to the considerable increase in the volume of calculations and the increased time required for the algorithm running and simulation.

Cross-validation of this study too was carried out based on the 12 better models of Fig. 9; ME, MSE, CC, and CE results are given in Table 3. As shown, the least CE is related to the single-structured exponential model (0.14+8 exponential (189)) which is shown in Fig. 9-1 as the selected model and marked with an asterisk (\*) at the left corner.

TABLE 3

CC, MSE, ME, and CE values for the second set of data

<b>n.</b>	<b>CC</b>	<b>MSE</b>	<b>ME</b>	<b>Combined Error</b>
<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
1	0.70	4.09	-0.119	4.52
2	0.48	6.17	0.034	6.73
3	0.43	6.57	0.069	7.21
4	0.56	5.42	0.003	5.86
5	0.45	6.35	-0.007	6.91
6	0.61	5.10	-0.131	5.62

1	2	3	4	5
7	0.45	6.31	-0.085	6.95
8	0.44	6.79	0.039	7.40
9	0.13	8.26	-0.052	9.18
10	0.22	7.71	-0.038	8.52
11	0.57	5.57	-0.039	6.04
12	0.17	8.09	-0.021	8.94

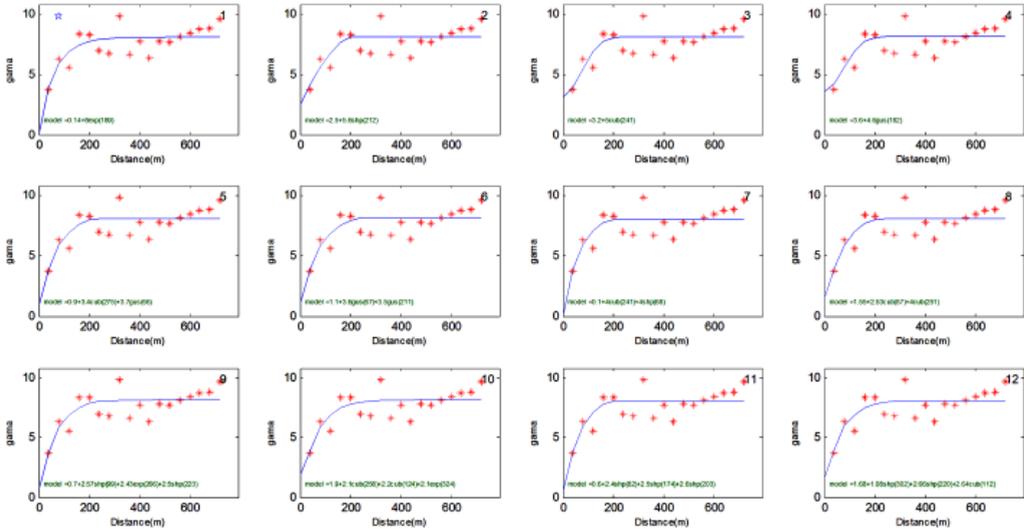


Fig. 9. Models proposed for the basic model of fitting by eye for the second dataset

### 3.3. Case study 3

The experimental indicator variogram has been worked out for a thickness dataset with a variance equal to 0.22. This experimental variogram has been calculated in 20 lags with an initial lag length of 250 m, an azimuth of 135 degrees, and a dip of 0 degree. Fig. 10 shows the shape of this experimental variogram together with the number of pairs in every lag, and Fig. 11 shows the results of running the algorithm. Four fitted, single-structured optimum models are shown in the first row of Fig. 11. As shown, the cubic (11-1) and spherical (11-2) models have been fitted well, but the fitted exponential and gaussian models are not so. It is seen that the quality of the fitted model has not been improved considerably due to the increase in the number of structures. But, this small improvement will lead to the complexity of the variogram model which increases the volume of calculations and the time needed to run the kriging algorithm; therefore, it is negligible, especially when the number of grid points and data are many, or when search volume neighborhood methods are not used.

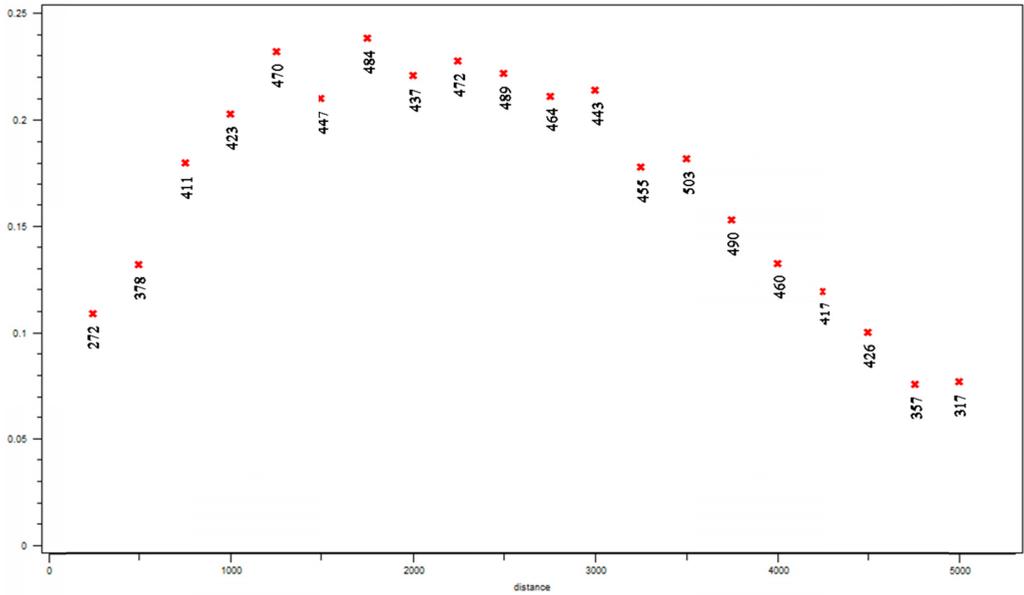


Fig. 10. The experimental variogram of the third dataset

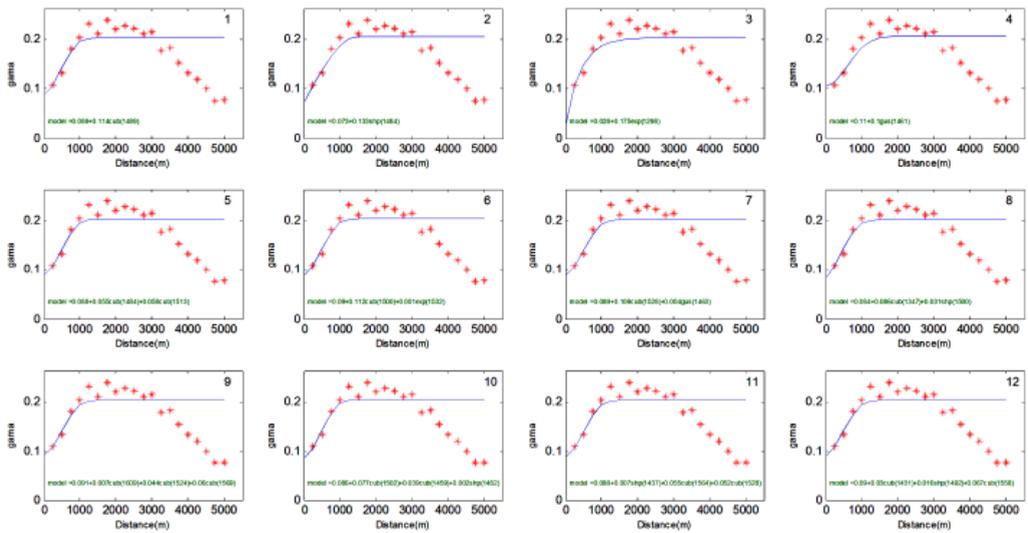


Fig. 11. Models proposed for the basic model of fitting by eye for the third dataset

## 4. Conclusions

In many experimental variograms, we observe rising, falling, or sinusoidal trends in the end lags, and this causes the variogram sill to deviate from the desirable value and go up or down; we added  $\sqrt{h}$  in the denominator of relation 3 to modify this effect. These case studies showed that we were able, by this modification of the objective function, to minimize the effects of these trends and bring the variogram sill closer to the desirable value.

As shown, increasing the number of structures from 2 to 3, does not considerably improve the quality of the fitted model, but it will lead to the complexity of the variogram model which increases the volume of calculations and the time needed to run the krigging algorithm; this will cause problems, especially when grid points and data are many, or when search volume neighborhood methods are not used.

Since, in the simulated annealing, we cannot define the “equal” case when we are defining the constraints, we are not able to define the constraints related to the equality of the sill of the fitted model with the value of the data variance; we can define only the upper and lower constraints of every decision variable ( $c_o$ ,  $c$ ,  $a$ ). To solve this problem, it is suggested that the future researches be focused on such other meta-heuristic algorithms as the genetic algorithm (GA).

## References

- Chiles J.-P., Delfiner P., 2012. *Geostatistics : modeling spatial uncertainty*. Hoboken, N.J., Wiley.
- Christakos G., 1984. *On the Problem of Permissible Covariance and Variogram Models*. Water Resources Research, 20(2): 251-265.
- Clark I., 1979. *Practical Geostatistics*. Applied Science Pub.
- Cressie N.A., 1979. *Straight line fitting and variogram estimation*. Bulletin of the International Statistical Institute, 48: 573-582.
- Cressie N., 1985. *Fitting variogram models by weighted least squares*. Journal of the International Association for Mathematical Geology, 17(5): 563-586.
- Cressie N., Hawkins D., 1980. *Robust estimation of the variogram: I*. Journal of the International Association for Mathematical Geology, 12(2): 115-125.
- David M., 1977. *Geostatistical ore reserve estimation*. Elsevier Scientific Pub. Co.
- Debba P. et al., 2009. *Deriving Optimal Exploration Target Zones on Mineral Prospectivity Maps*. Mathematical Geosciences, 41(4): 421-446.
- Desassis N., Renard D., 2013. *Automatic Variogram Modeling by Iterative Least Squares: Univariate and Multivariate Cases*. Mathematical Geosciences, 45(4): 453-470.
- Deutsch C.V., Journel A.G., 1992. *GSLIB: geostatistical software library and user's guide*. Oxford University Press.
- Emery X., 2010. *Iterative algorithms for fitting a linear model of coregionalization*. Computers & Geosciences, 36(9): 1150-1160.
- Genton M., 1998a. *Highly Robust Variogram Estimation*. Mathematical Geology, 30(2): 213-221.
- Genton M., 1998b. *Variogram Fitting by Generalized Least Squares Using an Explicit Formula for the Covariance Structure*. Mathematical Geology, 30(4): 323-345.
- Gorsich D., Genton M., 2000. *Variogram Model Selection via Nonparametric Derivative Estimation*. Mathematical Geology, 32(3): 249-270.
- Jian X. et al., 1996. *Semivariogram modeling by weighted least squares*. Computers & Geosciences, 22(4): 387-397.
- Journel A.G., Huijbregts C.J., 1978. *Mining geostatistics*, London ; New York, Academic Press.

- Kirkpatrick S. et al., 1983. *Optimization by simulated annealing*. Science, 220(4598): 671-680.
- Lee K.Y., El-Sharkawi M.A., 2008. *Modern heuristic optimization techniques : theory and applications to power systems*. Piscataway N.J., Hoboken N.J., IEEE Press; Wiley-Interscience.
- Lee Y.-M., Ellis J.H., 1997. *Estimation and simulation of lognormal random fields*. Computers & Geosciences, 23(1): 19-31.
- Luo R. et al., 2013. *Optimization of the seismic processing phase-shift plus finite-difference migration operator based on a hybrid genetic and simulated annealing algorithm*. Petroleum Science, 10(2): 190-194.
- Matheron G., 1965. *Les variables régionalisées et leur estimation, une application de la théorie de fonctions aléatoires aux sciences de la nature*. Paris, Masson et Cie.
- McBratney A.B., Webster R., 1986. *Choosing functions for semi-variograms of soil properties and fitting them to sampling estimates*. Journal of Soil Science, 37(4): 617-639.
- Pardo-Igúzquiza E., 1999. *VARFIT: a fortran-77 program for fitting variogram models by weighted least squares*. Computers & Geosciences, 25(3): 251-261.
- Paula F. Larrondo, C.T.N., Clayton V. Deutsch, 2003. *A Program for SemiAutomatic Variogram Modelling*. University of Alberta.
- Peredo O., Ortiz J.M., 2011. *Parallel implementation of simulated annealing to reproduce multiple-point statistics*. Computers & Geosciences, 37(8): 1110-1121.
- Shapiro A., Botha J.D., 1991. *Variogram fitting with a general class of conditionally nonnegative definite functions*. Computational Statistics & Data Analysis, 11(1): 87-96.
- Soltani-Mohammadi S. et al., 2012. *Investigating ground vibration to calculate the permissible charge weight for blasting operations of Gotvand-Olya dam underground structures*. Archives of Mining Sciences, 57(3): 687-697.
- Soltani-Mohammadi S., Hezarkhani A., 2013. *A Simulated Annealing-Based Algorithm to Locate Additional Drillholes for Maximizing the Realistic Value of Information*. Natural Resources Research, 22(3): 229-237.
- Soltani S., Hezarkhani A., 2009. *Additional Exploratory Boreholes Optimization Based on Three-Dimensional Model of Ore Deposit*. Archives of Mining Sciences, 54(3): 495-506
- Starks T.H., Fang J.H., 1982. *The effect of drift on the experimental semivariogram*. Journal of the International Association for Mathematical Geology, 14(4): 309-319.
- Taylor C.C., Burrough P.A., 1986. *Multiscale sources of spatial variation in soil. III. Improved methods for fitting the nested model to one-dimensional semivariograms*. Mathematical Geology, 18(8): 811-821.
- Webster R., Oliver M.A., 2001. *Geostatistics for environmental scientists*. Chichester, England; New York, John Wiley & Sons.
- Xia Y.-M. et al., 2011. *Optimal control of cobalt crust seabed mining parameters based on simulated annealing genetic algorithm*. Journal of Central South University of Technology, 18(3): 650-657.