Research Article

Development of a Three-Dimensional Analysis Software for Space-Time Epidemic Expansion Studies: GeoSpread

Eric Bruno Kabe M, Arsene Meying[*]**, Martin Luther Mfenjou**

School of Geology and Mining Engineering, University of Ngaoundere, Meiganga, Cameroon

[*]Corresponding author: Arsene Meying, School of Geology and Mining Engineering, University of Ngaoundere, Meiganga, Cameroon; E-Mail: arsenemeying@yahoo.fr

Abstract

The study of the evolution of epidemics in space and time remains a sensitive subject which increasingly requires digital tools that can be used for space-time modeling of the risks of contagion. The present paper focuses on the development of a software called Geo Spread, whose main function is to produce spatiotemporal prediction models of contagion chances, following a previous field investigation. By the way this paper raises an issue about spatiotemporal covariance functions modeling. The software incorporates algorithms of spatiotemporal indicators modeling in background and was implemented on a synthetic database containing survey results. The surveys have been virtually conducted for 09 days, in some localities in the Adamawa region of Cameroon. Through this practical trial case, this paper presents in detail the features of the software's project in development, as well as the theoretical elements used so far. The results showed that, the use of indicators function is appropriate to assess the chances of contagion in time and space. Especially since through these, it has been possible to assess spatial evolution for each of the 09 days of investigation and to make a prediction after 15 days. Finally, Geo Spread software still observes some limitations, but nevertheless, it can be properly listed as an instrument that can contribute to a better appreciation of response measures to the expansion of a rapidly spreading epidemic.

Keywords: Indicators; Kriging; Spreading Epidemic; Spatiotemporal Covariance Functions.

Introduction

Many geostatistical methods developed in the space-time field have been used recently in epidemiology research, to study the expansion of pandemics while considering the spatial structure of their evolutions. Work such as Chein et al. refers to the use of a Bayesian Maximum Entropy (BME) method to assess SARS contamination in the Guangdong region of China [1]. The separation of the space-time domain into spatial and temporal components, allowed analysis per sub-domain while neglecting the existence of a possible correlation between spatial and temporal patterns. Sharmin et al. used a Bayesian inference to perform Bayesian kriging interpolations to assess the expansion of dengue flavi-virus in Bangladesh. Other works, highlight the adequacy of a spatiotemporal geostatistical consideration to study the evolution and spread of epidemics [2-6].

The use of indicators function as probability estimators has been successful in many studies. It has been the case for Marschallinger et al. where indicator variograms was used to describe the 3-dimensional spatial structure of MS lesion patterns in a standardized manner. It thus allows them to completely automate screening of radiological archives and analyze the spatial behavior of multiple sclerosis. In the field of geosciences, they are used for uncertainty modeling issues [7]. They all need very often to perform variographic or structural analysis before applying kriging methods on data. But in a spatiotemporal domain, there is an important fact they did not consider. Working into a spatiotemporal domain involve the study of the behavior of a random process define in a surface considered defining the space directions, along the time direction. It does not mean to describe that process in the space directions with a summarizing value which should be used after. But in a theoretical sense, it means working with each single point of the surface and analyze they behavior in the time direction.

The present work returns to a concept hitherto rare in the state of the art concerning variographic analysis, which is that of structural functions family of a phenomenon studied in a space-time domain. In that way, one shall consider a set of point which spatially describe the random process, then study each of them in the time direction. It focuses on the development of a tool implementing geostatistical methods to assess the chances of space-time contamination. That tool incorporates within it an automatic variographic analysis program or Automatic Variogram Modeling (AMOVA). Automatic modeling has been chosen to make faster data processing, as the database may be increasingly large depending on the studied locality [8-11].

As Marschallinger et al. who worked on synthetic data to test the analytical tool developed in their works, in this paper a database of synthetic data is used to evaluate through the software designed, the expansion of an epidemic in a region of Cameroon, this in order to prove the effectiveness of the software. The study will consist of modelling in time and space, of the semi-variograms of indicators representing the spatiotemporal structures of chances to be contaminated inside the studied field. Then space-time predictions models will be obtained by considering a subdivision of the space-time domain into one spatial component and another temporal one.

Model Formulation

Space Component

Consider a locality survey and let $Z_i = Z(x_i)$ be the response to the survey "There are cases of infections at the point x_i". Thus, each point sampled constitutes a node of an oriented graph; moreover, from the measured data it is possible to identify a statistical law associated with the different nodes of the graph. It's a Markov's field type of configuration. Thus, for a V_j given neighborhood one can predict $Z(x_j) = Z_j$ by:

$$Z(x_j) = \sum_{k=1}^{J} Z(x_k)\lambda_k \qquad (1)$$

with $J = Card(V_j)$ and $Z(x_k) = \begin{cases} 1 & ,if\,at\,x_k\,they\,found\,contaminated\,cases \\ 0 & othherwise \end{cases}$

$\lambda_k, k = 1, \ldots, J$ are indicator kriging weights.

The semi-variogram function of $Z(x)$ is given by:

$$g(h) = \frac{1}{2}E\left[\left(Z(x+h) - Z(x)\right)^2\right] \qquad (2)$$

Using the observations made on the studied field, one approaches by gives with the formula:

$$g^{\square}(h_i) = \frac{1}{2N(h_i)}\sum_{i=1}^{N(h_i)}[Z(x_i) - Z(y_i)]^2 \qquad (3)$$

where: $x_i - y_i \vee |h_i|$.

The experimental model obtained must be fitted by an empirical model $\bar{g}(h)$ so that $E\left[|g^{\square} - \bar{g}|\right]$ shall be minimal [12,13,14]. Knowing $\bar{g}$, equation 4 allows calculating the λ_k necessary;

$$\begin{bmatrix} \bar{g}(x_\beta - x_\alpha) & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} \lambda \\ \mu \end{bmatrix} = \begin{bmatrix} \bar{g}(x_j - x_\beta) \\ 1 \end{bmatrix} \qquad (4)$$

obtained by minimizing the variance of the estimation error using the Lagrange multiplier method.

Temporal component

The study along the temporal component of the space-time domain will be conducted, in the same way as according to the spatial component. This mean that it would be necessary to model a structural function over time using the indicator variogram, then predict using indicator kriging. However, there is a small peculiarity for the case of variographic analysis in the temporal sub-domain. Let be the following illustration:
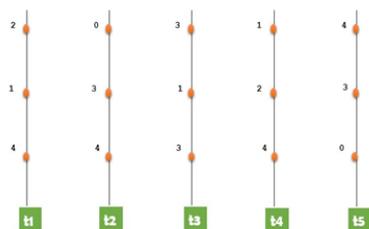

Figure1: Space-time analysis illustration.

Let's suppose for example, that then it is clear that the study among the temporal component will have misinformed about the spatiotemporal evolution of the random process Z. Looking for a covariance function materializing the structure of temporal variation for this case, would be like working with a value for each given . Nevertheless, for each given there are three different. Thus to model the random process into a spatiotemporal domain one shall compute the behavior of each in the time direction then model use each predicted to produce de spatiotemporal model of Z.

In this section must be defined, the encoding used, the type of selection, the crossover and permutation scheme, the evaluation function, and finally the criterion for selecting the best solutions.

Encoding: the encoding used is real-valued encoding. The chromosome is represented in the algorithm by 1-2, 5-4.8-10 for example, where the first element corresponds to a variogram model, the second corresponds to the range, the third to the sill and the last to the nugget effect;

Selection: The selection will be random;

Crossover: The crossover will follow the uniform crossover pattern;

Mutation: The mutation will consist of pulling two positions from the chromosome and transforming the genes involved;

The evaluation function: depending on the type of model selected the program will use the theoretical model formula and the parameters of the solution to evaluate it;

The best solutions: after each iteration the program selects the chromosomes that showed the smallest variance of the gap with the experimental model.

To achieve crosses and mutations on selected chromosomes, it is necessary to define a probability of crossover and mutation. The calculation of these values is a subject of discussion regarding to the considerations that shall be observed during the process. However, in the literature there are generally probabilities of crossover close to 70 to 80%, and probabilities of mutation around 40%. It is these values that have been used in the genetic program.

The implementation of the genetic program can be represented by the following diagram Figure 2:
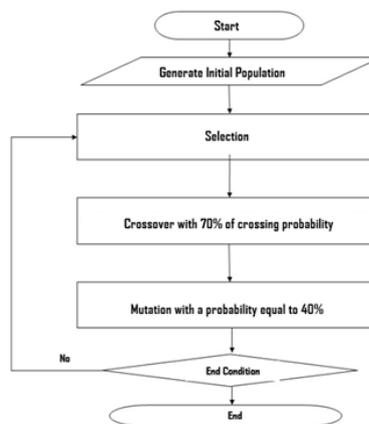

Figure 2: Genetic Program Execution chart.

Implementation and Results

Geospread goals and architecture

GeoSpread is software that has been designed to evaluate the evolution of a random process defined in a space-time domain. In this paper it has been experimented on indicators, but for a wider use, it can very well adapt to real-data variables such as of mortality by locality or the decrease in the concentration of a mineral... It is a laboratory tool developed on the Qt's framework in C++. It is light and simple to handle, plus it is quite user-friendly. This was done in order to allow its use to a wide audience, ranging from the field of epidemiology to the fields of environmental geosciences [15].

GeoSpread was designed to be freely used on Windows (64 bits and 32 bits) and does not depend on any other built-in application. The data manipulated in it must be stored in a '.csv' file (comma separator ','). As mentioned earlier in this paper, GeoSpread incorporates an automatic variographic analysis program, then spatial and space-time prediction. However, his manipulation follows a scheme which is:

Importing data: the limits of the studied area and the survey database ;

Spatial analysis: modeling of the spatial variogram, then spatial prediction;

Temporal analysis: modeling the structural function family, then achieve point-by-point temporal prediction, then variographic (spatial) analysis and finally spatiotemporal prediction.
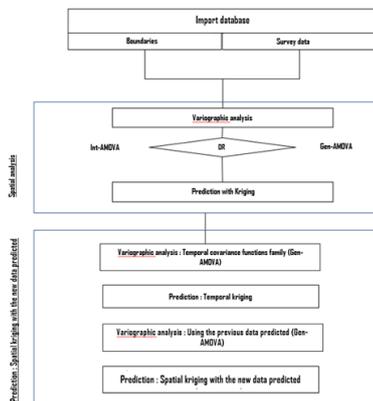
The diagram on figure 3 resumes GeoSpread architecture.

Figure 3: GeoSpread Architecture

Style	Index	Lat	Lon	Proj_x	Proj_y
WAYPOINT	1	6.557239369	11.33731897	6.557239369	11.33731897
WAYPOINT	2	6.464908907	11.56814512	6.464908907	11.56814512
WAYPOINT	3	6.384119753	11.76434735	6.384119753	11.76434735
WAYPOINT	4	6.280247983	11.96054959	6.280247983	11.96054959
WAYPOINT	5	6.187917521	12.12212789	6.187917521	12.12212789

Table 1: Structure of files containing boundaries coordinates of the field studied.

Style	Index	Lat	Lon	Proj_x	Proj_y	Data
WAYPOINT	1	6.557239369	11.33731897	6.557239369	11.33731897	1
WAYPOINT	2	6.464908907	11.56814512	6.464908907	11.56814512	1
WAYPOINT	3	6.384119753	11.76434735	6.384119753	11.76434735	0
WAYPOINT	4	6.280247983	11.96054959	6.280247983	11.96054959	0
WAYPOINT	5	6.187917521	12.12212789	6.187917521	12.12212789	0

Table 2: Structure of files containing data from daily surveys.

GeoSpread works with the 'Lon/Lat' coordinates as well as the 'Easting/Northing' coordinates. However, the results that will follow in this section were obtained using longitude and latitude coordinates.

Space analysis

The analysis on GeoSpread begins with the establishment of spatial distributions associated with each file entered into the database.

Indeed, for this case of study, a survey was carried out at the same study points, inside a test region. Each day corresponds to a file and from this file, it will be necessary to draw a spatial distribution of the chances to have contaminated cases.

A Synthetic Case of Study

A survey was virtually conducted in some localities in the Adamawa region of Cameroon. After that, it has been assumed a question on how to visualize the geographical expansion of the disease observed in these localities and in the extent of the region. To do this we have introduced the collected data in GeoSpread and the results obtained will be presented.

Data input

The data provided as input to the software must have a precise structure. In particular, the software distinguishes files containing only the geographic coordinates of the boundaries of the domains studied from files containing the surveys data.

The first step in this procedure is the structural analysis for each daily data. Indeed, to study the contamination of the inhabitants of a region by a disease such as SARS for example (person by person contamination), it is necessary to define vectors of contamination and then analyze their spatial structure. Here the vector of contamination is of course the displacement of a contaminated case. The structural function that is modeled by the semi-variogram of indicators is the structural variation of the chances for the virus to move inside the domain.

The following variogram models (Figure 4 and Figure 5) were obtained for the survey data conducted in the domain, both from the Int-AMOVA algorithm and from the Gen-AMOVA algorithm:
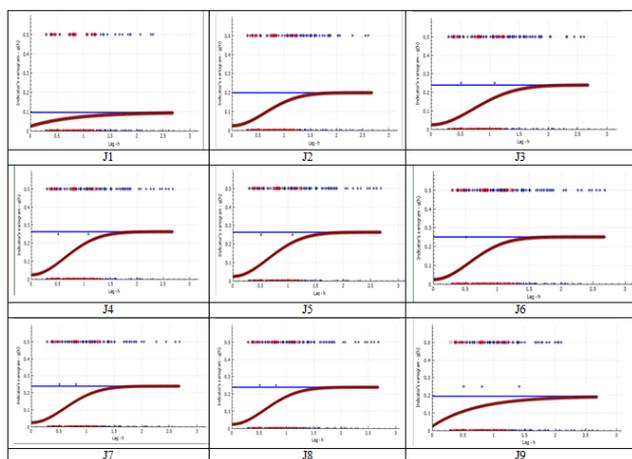
Figure 4: Model of variogram obtained during the 09 days of investigation in Adamawa - with Int-AMOVA
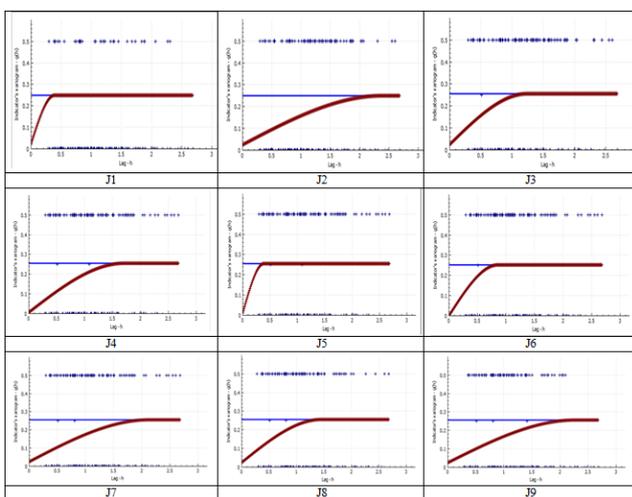

Figure 5: Models of variogram obtained during the 09 days of investigation in Adamawa - with Gen-AMOVA

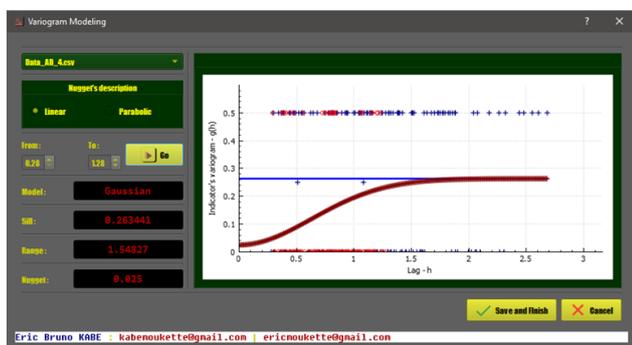In GeoSpread the result looks like this image:

Figure 6: Results of variographic analysis on GeoSpread.

This will allow the user to easily read the model parameters obtained from the automatic modeling of the empirical isotropic variogram (Figure 6).

The results obtained shows that depending on the algorithm chosen, variogram models obtained May (very often this is the case) vary. For

the case of this study for example, in the same region (Adamawa), we get predominantly spherical models using the Gen-AMOVA algorithm and using the Int-AMOVA algorithm the models are much more Gaussians. This difference will not be inconsequential in the future of the space-time modeling process.

Once empirical variogram models are obtained, it is now possible to initiate the calculation of the spatial distribution model of the chances of contamination.
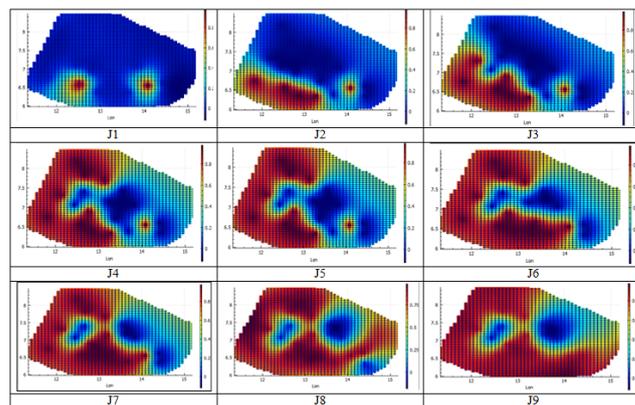

Figure 7: Spatial distribution model of the chances of contamination obtained during the 09 days of investigation in Adamawa - with Int-AMOVA.
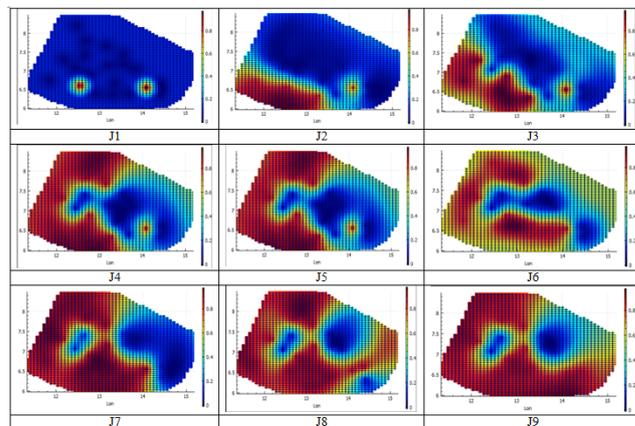

Figure 8: Spatial distribution models of the chances of contamination obtained during the 09 days of investigation in Adamawa - with Gen-AMOVA.

Such models (Figure 7 and Figure 8) allow users to figure out how far the virus is spreading inside the region. They represent the expansion of the chance to detect a contaminated case in a specific locality for each day.

Space-time analysis

At the end of the spatial analysis, it is necessary to save the models made in the form of a 3D grid (.gsgrid), where the third dimension corresponds to the chance of contagion.

In the space-time analysis process, 50 node indexes are randomly drawn, and the corresponding nodes are identified in the 09 grids of the calculated values and for each t node, in the temporal domain and at a given time (in days) the chances of contagion are estimated. At the

end of this stage we have 50 new values that we use to evaluate the new spatial distribution at the t desired time.

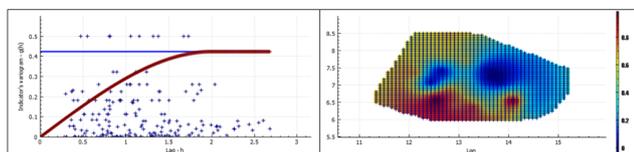The following results (Figure 9) are the variogram model and the spatial distribution obtained at: :

Figure 9: Pattern of contagions predicted to 15th day.

Discussions

The techniques of variographic analysis are at the heart of the quality of this work. However, most of the case study illustrating automatic variogram modeling tends to avoid on the complexity of a structural function on a space-time domain. It come through mind that variographic modeling on a space-time domain shall be consider as the study of a random process in an anisotropic sense, where the third temporal dimension need to be analyzed beside the others from the spatial sub-domain. It should not refer to an analysis which tends to be a multivariate study, because there is just one random process here, which is chance of contagions, no matter if it is deployed on a temporal sub-domain within a spatial domain. In fact, some fancy mind, of course with some effectiveness but also to much hypothesis which can be avoided, as the work of Varouchakis, considers a spatiotemporal covariance function as the semi-variogram of a random process $Z(s, t)$ where s is a vector resulting from the sum of the spatial vectors. But one thing which is now ambiguous is that the temporal dimension cannot be shapeable in the same way as the spatial dimension. For example, if it must be considered a two-dimensional spatial sub-domain, s can be s=x+y. In the time dimension that relation barely has a sense. Plus, it suppose that in the both direction the step scale on space and time subdivisions evolves at the same time in the same way ; this is to much hypothesis which even if it produce some convenient results, may fail to achieve the task ongoing in a realistic sense.

Just as it was the case in the work of Marschallinger et al., Sharmin et al. and Paulo et al. the quality of a prediction (indicator kriging for this case) depends closely on the modeling of structural function, which here is the indicator variogram [2,7,11]. The first approach implemented in GeoSpread to model the empirical indicator variogram is almost rare, if not absent, in the literature relating to this exercise. Yet many manuals allude to the description elements of a variogram model and it is these elements that have been used to develop this automatic modeling algorithm. By this way the process is light to understand, while it keeps a certain theoretical background to prove its effectiveness. It allows a rapid variogram modeling which can produce good results of variogram models.

The idea of using genetic algorithms for modeling empirical variograms, however, is an idea that is already in the literature. Notably the work of Zhanglin et al. who are still puzzled as to the robustness of a variographic modeling process using a genetic program. They have also prescribed more testing on more complex experimental models. In this sense, models of indicator variograms are generally unusual, especially since non-linear geostatistics is a branch that has been very little practiced so far. Thus, this study sees itself as a contribution that would bring mores element on the analysis and appreciation on the use of genetic algorithms for variographic modeling. However, the variographic analysis in GeoSpread software does not consider the study of anisotropy, which can be seems eventually as a limit.

The development of space-time models using indicator functions for estimating the chances of contagion is a very recent idea, if not new. Indeed, even in the most recent work on the subject, the most prominent methods are Bayesian methods and differential with two-component (spatial and temporal). Yet, these methods do not estimate the chances of contagion but rather the number of contaminated cases. The use of indicator functions being appropriate for estimating probability (although not the most effective), thus contributes to the development of planning tools for better control of epidemic developments with prediction of uncertainties. The model used in GeoSpread software considers the displacement of contaminated people in time and space as the only vector of contagion, as is the case in the work of George et al. and what may be a limit to it. Nevertheless, this model of space-time analysis makes it possible to contribute to current work on the subject, considering the spatiotemporal evolution of the chances of contagion by cities and regions.

Conclusions

The study of the spatial evolution of epidemics remains a sensitive exercise, requiring more and more sophisticated tools that can help with the task. In this paper, the software GeoSpread has been presented, whose main function is to produce a space-time model of the chances of contagion as a result of field investigations. Through a synthetic study case, it was implemented to assess its effectiveness and demonstrate its suitability for use on real data. By the end of this study, GeoSpread has appeared adapted to the task involved, and therefore may be of paramount use in epidemiology, and in the fields of geosciences working with similar models.

References

1. Cao C, Chen W, Zheng S, Zhao J, Wang J, et al. (2016) Analysis of spatiotemporal characteristics of pandemic SARS spread in mainland China. Biomed Res Int 2016: 1-13.

2. Sharmin S, Glass K, Viennet E, Harley D (2018) Geostatistical mapping of the seasonal spread of under-reported dengue case in Bangladesh. PLoS Neg Trop Dis 12: e0006947.

3. Tariq B, Zaidi AZ (2019) Geostatistical modeling of dengue disease in Lahore-Pakistan, SN Appl Sci 1: 459.

4. Christakos G, Zhang C, He J (2017) A traveling epidemic model of space-time disease spread. Stoch Environ Res Risk Assess 31: 305-314.

5. Schroder W, Schmidt G (2014) Modeling potential malaria spread in Germany by use of climate change projections: a risk assessment approach coupling epidemiologic and geostatistical measures. (1stedn), Springer Cham.

6. Nkeng MN, Mousseni IBE, Nomo LB, Sache I, Cilas C, et al. (2017) Spatiotemporal dynamics on a plot scale of cocoa black pod rot caused by Phytophthora megakarya in Cameroon. Eur J Plant Pathol 147: 579-590.

7. Marschallinger R, Golaszewski SM, Kunz AB, Kronbichler M, Ladurner G, et al. (2014) Usability and Potential of Geostatistics for Spatial Discrimination of Multiple Sclerosis Lesion Patterns. J Neuroimaging 24: 278-286.

8. Yang Y, Christakos G, Huang W, Lin C, Fu P, et al. (2016) Uncertainty assessment of PM2.5 contamination mapping using spatiotemporal sequential indicator simulation and multi-temporal monitoring data. Sci Rep 6: 24335.

9. Soares A, Nunes R, Azevedo L (2017) Integration of uncertain data in geostatistical modeling. Math Geosci 49: 253-273.

10. Liu Y, Xia Q, Carranza EJM (2019) Integrating sequential indicator simulation and singularity analysis to analyze uncertainty of geochemical anomaly for exploration targeting of tungsten polymetallic mineralization, Nanling belt, South China. J Geochem Explor 197: 143-158.

11. Pereira PEC, Rabelo MN, Ribeiro CC, Diniz HSP (2017) Geological modeling by an indicator kriging approach applied to a limestone deposit in Indiara city-Goias. Int Eng J 70: 331-337.

12. Li Z, Zhang X, Clarke KC, Liu G, Zhu R, et al. (2018) An automatic variogram modeling method with high reliability fitness and estimates. Comput Geosci 120: 48-59.

13. Mohammadi SS, Safa M (2016) A simulated annealing based optimization algorithm for automatic variogram model fitting. Arch Min Sci 61: 635-649.

14. Carvalho PR, Rasera LG, Costa JF, Araujo MG, Varella LE, et al. (2019) Variogram modeling of broadband artifacts of a seafloor map for filtering with factorial Kriging. Appl Geophys 161: 92-104.

15. Ly S, Charles C, Degre A (2011) Geostatistical interpolation of daily rainfall at catchment scale: the use of several variogram models in the our the and ambleve catchments, Belgium. Hydrol Earth Syst Sci 15: 2259-2274.