Iraqi Journal of Statistical Science (27) 2018
pp[1- 11]

Spatial Prediction for Real Data using Kriging Technique

Shaymaa Riyadh Thanoon*
shaymaaalnuaimi@yahoo.com

Abstract

This study investigates the prediction of unstable random spatial process using the Ordinary Kriging technique, based on the variogram function Y(X,Y) in finding out the predictors, and the universal multivariate technique, which has been applied by computing the Cross Variogram Function. Practically, the data were taken from a geological study conducted in the Department of Geography, Faculty of Education, Mosul University. The study includes the height levels of (40) oil wells in Kirkuk. We discovered out the estimation of the Variogram and Cross the Variogram functions as well as the Mean Squared Error (MSE). Most of the results have been presented in tables and figures and the results were generally good, acceptable and close to reality,Our results showed that the overall Kriging technique or technique that was computed by the Cross Variogram Function was better than the Ordinary Kriging technique or technique based on the variogram function.

Key words:Universal Kriging, Ordinary Kriging, Cross Variogram Function.

الملخص

نركز اهتمامنا في هذه الدراسة على التنبؤ عن العملية العشوائية المكانية غير المستقرة في ايجاد التنبؤات بين $\gamma(x,y)$باستخدام أسلوب الكريكنك الاعتيادي اعتماداً على دالة الفايروكرام أسلوب الكرينك الشامل متعدد المتغيرات الذي تم توفيقه بحساب دالة الفايروكرام المتقاطع cross variogram function.وفي الجانب العملي أُخذت البيانات من دراسة جيولوجية في في قسم الجغرافية لكلية التربية جامعة الموصل تتضمن مستوى ارتفاع(40) بئرا من آبار النفط في مدينة كركوك،أُوجدنا تقدير دالة الفايروكرام وتقدير دالة الفايروكرام المتقاطع و تم كذلك حساب متوسط مربعات الخطأ. وتم توفيق أَغلب النتائج بالجداول والرسم ، إذ كانت النتائج عموما جيدة ومقبولة وقريبة من الواقع الحقيقي، ومن خلال النتائج التي توصلنا إِليها اثبتنا أَن تقنية او أسلوب الكريكنك الشامل الذي تم توفيقه بحساب دالة الفايروكرام المتقاطع كانت افضل من تقنية او أسلوب الكريكنك الاعتيادي اعتمادا على دالة الفايروكرام .

*Assistant lecturer/ College of nursing/ University of Mosul.

Received Date 5/6/2018 **Accept Date 28/6/2018**

Introduction

Most statisticians prefer the term “Prediction” to express the random process estimation in certain locations to distinguish it from the parameter estimation word existing in a certain probable distribution in the spatial statistics.

The word “Estimation” was widely used by geo-statisticians at the beginning of the study of the random spatial process, but generally called the Kriging. The solution for the Estimation issue relies on a continuous model of the random spatial process in unobserved sites relying on measured ones .

Kriging can be divided into : the Simple Kriging, (the least common used one), the Ordinary Kriging, the Fuzzy Kriging, through assuming the fuzzy triangular numbers which can be represented by three values (middle, right, and left) for the random variable, the Indicator Kriging, and finally the Universal Kriging.

Through this study, were used two techniques for Prediction : the Universal Kriging and the Ordinary one Kriging

General Linear Model in spatial:

The spatial linear model of the variable z(x) in the domain or space D can be writ ten in the form of matrices as in the following formula (Gaetan and Guyon,2010):

$$Z = FB + \in \qquad (1)$$

$$z(x) = f^T(x)\beta + e(x) \ldots \ldots . x \in D \subseteq R^p \qquad (2)$$

considering that :

$$z = \left(z(x_1), z(x_2), \cdots, z(x_n)\right)^T$$

$$F = f^T(x) = \left(f(x_1), f(x_2), \cdots, f(x_n)\right)$$

$$B = (B_1, B_2, \cdots, B_n)^T$$

The dependent relationship between the variables of the process is measured by the Variogram function. This function is regarded an alternative measure to correlation since the differences among the observations of the spatial variable values are sometimes large, so the resulted correlation coefficients are small, which makes the results and their explanations inaccurate. Therefore, the Variogram function was suggested by displacement of *h* represented by the following formula:

$$2\gamma(h) = E[z(x) - z(x+h)]^2 \qquad (3)$$

$$2\gamma(h) = \frac{1}{n(h)} \sum_{i=1}^{n(h)} [z(x_i) - z(x_i + h)]^2$$

This function is not dependent on the location of X, but it mainly depends on the distance h. As for y(h), it is called the variogram function, where n(h) represents the number of pairs of observations that are separated from each other by the distance h .

z(x) represents the expected value in the location x

z(x+h) represents the expected value in the location x+h

The variogram function correlates with the covariance if the random processes are stationary due to the fact that for each couple of random spatial variables [Z(x+h) , Z(x)] there is a covariance. This covariance depends on the distance h between Z(x) and Z(x+h) :(Journel,1986).

$C(h) = E[Z(x+h)\,Z(x)] - \mu^2 \quad \forall\ x، x+h \in D$

$C(h) = Cov[Z(x+h), Z(x)] \quad ، \quad \forall\ x، x+h \in D$

a correlation relationship among the covariance C(h) , the variogram function, and the Contrast σ^2 can be realized as shown in the following formula: (Goovaerts,1997)

Where : $C(0) = \sigma^2$

When h=0, we realize the following:

$c(0) = E[Z^2(x)] - \mu^2 = \sigma^2 = V[Z(x)]$

From the definition of y(h), we notice the following

$\gamma(h) = \frac{1}{2} Var[Z(x) - Z(x+h)]$

$\gamma(h) = \frac{1}{2} [VarZ(x) + VarZ(x+h) - 2Cov(Z(x), Z(x+h))]$

and by substitution, we realize the following:

$\gamma(h) = \frac{1}{2} [C(0) + C(0) - 2C(h)]$

$\gamma(h) = C(0) - C(h)$

See, (Gressie , 1993).

Variogram Function Properties

1- **Range**

$\lim_{|h|\to\infty} \gamma(h) = \gamma(\infty) = \sigma^2$ (4)

The above means that after a certain distance, the semi-variogram function will be constant. The distance σ is defined as the range for variogram function which means that :

$\gamma(h)=\sigma^2 \qquad |h|<a$

$\gamma(h)=\sigma^2 \qquad |h|\geq a$

That is, the variogram function will become stationary after a certain distance and the range gives us a concise meaning for the concept of a mining area that contains minerals, oil wells, or groundwater.

2- **Nugget Effect**

If,

$\lim_{|h|\to\infty} \gamma(h) = \psi_0 \neq 0$ (5)

That is to, say that there is a continuity in the origin point and the value of nugget is ψ_0 which represents a weak spatial continuity.
This means that the variance between the two variable values Z(x) and Z(x+h) taken from two locations close to each other will be high and the value of ψ_0 will be large whenever the distance between these two points is minimized.

3- **The Sill**

If

$\gamma(h)=\sigma^2$ **(6)**

$|h| \geq a$

Then this means that it is an even function realized as $\gamma(h)= \gamma(-h)$ and symmetric $\gamma(y,x)= \gamma(x,y)$ and its utmost is at $h \to \infty$, then, the value of $\gamma(\infty)$ is called the Sill. It simply means the Primary Variance.

Ordinary Kriging:

The first technique for prediction is the Ordinary Prediction . The ordinary prediction sample for the stationary spatial random process, known as the Ordinary Kriging, assumes the following:

$$Z(X) = \mu + \varepsilon(X) \qquad x \in D \tag{7}$$

And when μ is known, we will obtain the prediction as in the following

$$z_0(x) = \sum_{i=1}^{n} \delta_i Z(x_i)$$

here :

$$\sum_{i=1}^{n} \delta_i = 1 \tag{8}$$

And that δ_i represents the weight vector and $\sum_{i=1}^{n} \delta_i = 1$ ensures unbiased prediction , that is :

$$E\left(z_0(x)\right) = E\left(z(x)\right) = \mu$$

It is the estimated value of the weights that makes the square difference between the actual values and the predictive ones as low as possible or which minimizes the Mean Squared Error (MSE) for prediction using the Lagrange technique:

$$\operatorname{var}[z(x)] = E[Z(x) - z_0(x)]^2 \tag{9}$$

We will obtain the following formula,(Kastelec and Kosmmelj,2002) :

$$\sigma^2 = \gamma^{-1}\left[\gamma_\circ + \left(\frac{1-\gamma_0^t \gamma l}{l^t \gamma^{-1} l}\right) l\right] \tag{10}$$

Here:

I=(1,1,…,1)

$\gamma_0 = (\gamma_{01}, \gamma_{02}, \cdots, \gamma_{0n})^T$

$\gamma = \gamma_{\mathbf{ij}} = \gamma\left(x_i - x_j\right)$

either variance of Kriging ,we get it from compensating estimated values and estimated coefficients Lagrange used in the forecast variance formula (9).

The Universal Kriging

The Universal Kriging technique is one of the common techniques for prediction and is used when the expectation is not constant and is associated with the existence of the direction, since most practical applications have a direction in predicting, that is the prediction is not known and depends on the location within the area under study, unlike the ordinary Kriging technique which is used in the practical applications that have no direction. Hence , the expectation is constant (and unknown) all over the sites of the area under study. There are several researches that have dealt with the Universal Kriging technique, (Diggle and Ribeiro,2007) and (Webster an Oliver,2007)

The model is written as in the following formula :

$$z(X) = \mu(x) + \varepsilon(X) \qquad x \in D \qquad (11)$$

as for prediction, it can be obtained from the following formula:

$$Z^{*}(x_0) = \sum_{i=1}^{n} \lambda_i z(x_i) \quad , \text{ where } \sum_{i=1}^{n} \lambda_i = 1 \qquad (12)$$

Considering that:

$$\lambda = \gamma^{-1}[\gamma_0 - x(x^T\gamma^{-1})^{-1}(x^T\gamma^{-1}\gamma_0) - F] \qquad (13)$$

This variance is called Kriging variance and can be represented by the following formula:

$$^{2}_{K} = \gamma^{-1}\gamma_0 - \left[f^T \left(\gamma_0^T\gamma^{-1}X - f^T(x_0)\right)(X^T\gamma^{-1}X)^{-1}\left(X^T\gamma^{-1}\gamma_0 - f(x_0\ )\right)\right] \qquad (14)$$

When computing the Universal Kriging, we need to compute the cross semi-variogram function $\hat{\gamma}_{AB}(h)$ which was presented as the following form (Soares et al,2008):

$$\hat{\gamma}_{AB}(h) = \frac{1}{2}r(h)\sum_{\vartheta=1}^{r(h)}\left(\left(z_A(x_\vartheta) - z_A(x_\vartheta + h)\right)\right)\left(\left(z_B(x_w) - z_B(x_w + h)\right)\right)$$

$$\forall\, A, B = 1,2,\cdots,\eta \quad , \quad \vartheta, \omega = 1,2,\cdots, r(A) \qquad (15)$$

Considering that $A, B, \vartheta. w$ represent the average variances among the spatial observations which are separated from each other by the replacement of h and that r(h) represents number of observations couples $z_B(x_w + h)$,$z_B(x_w)$, and $z_A(x_\vartheta + h)\ z_A(x_\vartheta)$ that the replacement of h separates them. In case that A=B, the variogram computing for the same variable will be as follows : (Chiles and Delfior,1999)

$$\hat{\gamma}_{AA}(h) = \hat{\gamma}_A(h) = \frac{1}{2}r(h)\sum_{\vartheta=1}^{r(h)}\left(\left(z_A(x_\vartheta) - z_A(x_\vartheta + h)\right)\right)^2 \qquad (16)$$

Application

Spatial predictions for the spatial phenomenon were carried out by applying the Ordinary Kriging using the variogram function and the Universal Kriging technique using the cross variogram function using actual data representing the depth of the oil wells in the city of Kirkuk. The data were taken from a study carried out by the Faculty of Education, Department of Geography at the University of Mosul and here is the table of variables in the data .

X	Y	z(x)
364029	422344	340.400
361142	420406	320
361500	421140	320
364045	420715	330
363745	420230	370
364900	420600	393
364446	421829	390
364833	420809	390.600
364200	422200	392.700
363900	420230	345
364350	422320	382
364846	422459	322
364700	420900	362.600
364400	422000	385
361825	425539	360
364818	420448	377.400
364835	420526	386.300
364523	422706	390.100
364710	420840	268.600
361830	425400	380.600
364115	422720	377
363532	423136	340
364500	421500	361
361700	424948	360
361445	424630	351
361828	421257	350
363916	420155	336.600
362200	424200	382
362300	425300	360

363801	420222	340
364543	420357	393.700
364357	421010	397.300
364347	421404	368.200
364142	421648	379.300
364259	422102	350
364500	422500	348.500
364406	421542	350
363542	420942	351
364200	420036	390
363900	422300	360

Table (1) : The actual data with their coordinates

In the present study, we used the above (Table 1) data representing the levels of height of oil wells as an initial variable and depth of the well as a secondary variable. The distance matrix, which is symbolized by h, was calculated. The number of pairs of observations r(h), which are separated from each other by the distance h were also computed. The relationship between h and y(h) were drawn and then we plotted the variogram curve among the points of this relationship . Fig. (1) shows the variogram function. (Haning,2004)

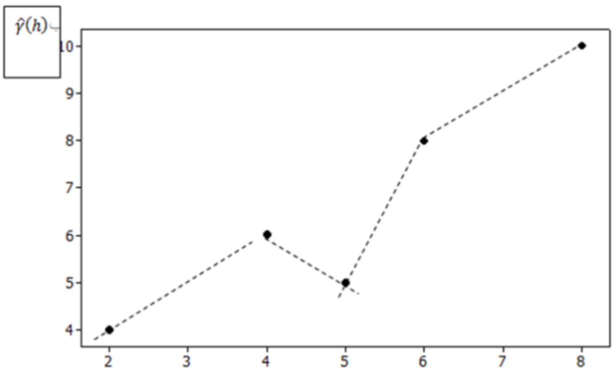

Fig. (1) The Variogram Function

Figure (1) shows the instability of data as the curve continues to rise and be instable. The function was represented according to the following exponential model: $\gamma(h) = C\left[1 - \exp\left(\frac{-h^2}{a}\right)\right]$

considering that

C is a constant

h represents the distance

a represents the range

It is worth mentioning that the exponential model is the most common among the models used to study the level of depth of water and oil wells. From (**4**), (**5**) and (**6**) we could reach the following results:

MSE	Range	Sill	Nugget
0.899	3110	33340	2555

Table (2) : Parameter values of the exponential model and MSE

The direction of the random process can be recognized or not through viewing whether the variogaram curve is stable or unstable.If the curve is stable when reaching the distance of range, this indicates that the random spatial process is not affected by the direction, which means that the prediction is constant at all sites (Boogaart,1999) If the variogram curve is unstable (continuing to rise) when reaching the range distance, the random spatial process is affected by the direction, that is , the prediction varies from one location to another. In this case, the Universal Kriging technique is used.

We used equation (14) to calculate the variogram for both variables, and also used equation (15) to calculate the cross variogram of the two variables.

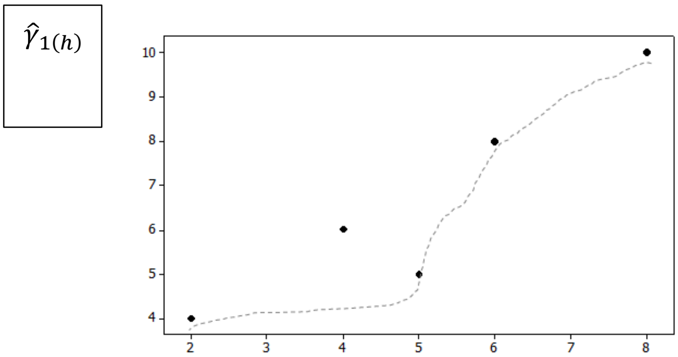

Figure (2) :The curve of VariogramFunction for the primary variable

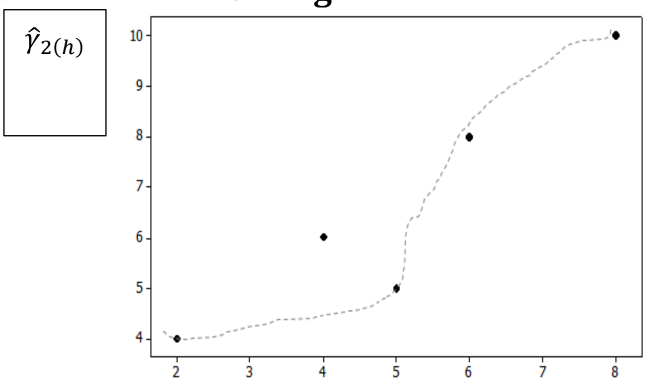

Figure (3): The curve of Variogram Function for the secondary variable

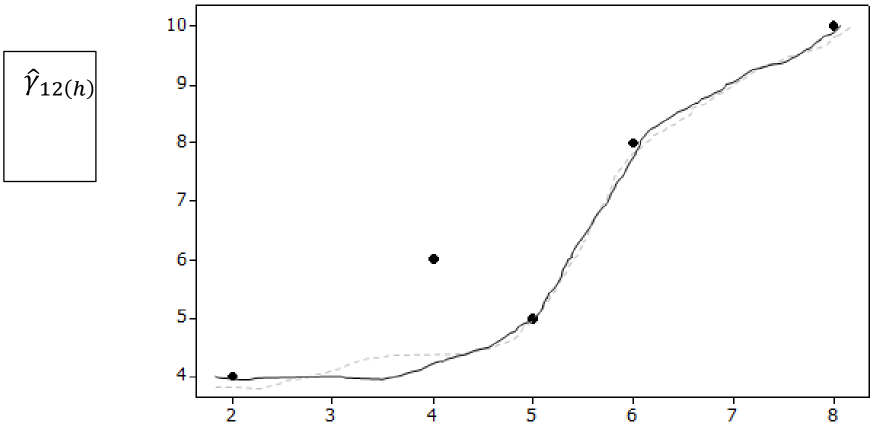

Figure (4): The cross Variogram Function of the two variables

From the above variogram models, it is noticed that variogram values are unstable, thus, the universal Kriging is required. We also used the Gaussian Model in prediction :

Gaussian Model : $\gamma(h)=\psi[1-\exp(-h^2/2a^2)]$

for the primary variogram, the secondary variable, and also for the cross variogram of the two variables

	nugget	sill	range
γ_1	33340	2563	1256
γ_2	0	2110	1879
γ_{12}	0	1276	1143

Table (5) : The values of Gaussian Model parameters

It has been predicted regarding the level of oil height in the real observations sites to compute the prediction values and the variance of the Universal Kriging equation (10). Table (4) below shows a group of the prediction values of the Universal Kriging for the level of oil in the well and its variance. The total of the weights of both variables has also been calculated. The total weights for both variables were also calculated in addition to the calculation of the Mean Squared Error (MSE) and its value is **0.310**.

z(x)	$Z^*(x_0)$	σ^2
340.400	344	0.45400000
320	322	0.45300000
320	318	0.45600000
330	330	0.45400000
370	375	0.45400000
393	390	0.45400000

390	390	0.45400000
390.600	390	0.45300000
392.700	398	0.45400000
345	344	0.45400000

Table (6): shows The real values, prediction values, Kriging variance values

Discussion

In this study, the prediction was applied to actual data for the oil levels in the wells through prediction by the Ordinary Kriging which was represented by the Variogram function and applying the exponential model. In Fig. 1, it was noticed that the curve of the Variogram function continues to rise, which means that it is unstable. In estimation, we adopted the Mean Squared Error (MSE) and the estimation of the model parameters. This was explained in Table (2). The Universal Kriging has been applied to the same data as it comprised the height level of oil in the well as a primary variable and the depth of the well as a secondary variable. We used the Gaussian model for the primary and secondary variables as well as the cross Variogram function. The diagram was illustrated in Figures 1, 2, and 3. In Table (3), the parameters of the Gaussian model were explained. The Universal Kriging variance was also calculated in all the sites. Some of these values were illustrated in Table (4). The results showed that the Universal Kriging variance is very small, which indicates the accuracy in prediction and the convergence of the predictive with the actual values. The Mean Squared Error was compared in both techniques, showing how accurate the prediction is in the Universal Kriging and its representation by the cross variogram function.

References:

1- Boogaart, K.G. (1999): " New Possibility for Modeling Variogram in Complex Geology " , to appear Proceedings of Stat GIS 1999.

2- Chiles, J. P. and Delfiner, P. (1999),"Geostatistics: Modeling Spatial Uncertainty", Wiley and Sons, New York, USA.

3- Cressie,N.(1993) ,"Statistics for Spatial Data", Second Edition John . Wiley ,New York.

4- Diggle, P.J. and Ribeiro,P.J.(2007),"Model-based geostatistics",Springer series in statistics,New York,USA.

5- Gaetan,C. and Guyon,X.(2010) ,"Spatial Statistics and Modeling"

6- Goovaerts,P.,(1997). Geostattistics for natural resource evaluation, Oxford University Press, Oxford.

7- Haining ,R. (2004) , "Spatial Data Analysis,Theory and Practice", 9- Cambridge University Press.London .

8- Journel, A.G., (1986). "Geostatistics Models and Tools for the Earth Sciences", Mathematical Geology, No. 18, PP. (119-140).
9-Kastelc,D.and Kosmelj,k. (2002),Spatial interpolation of mean yearly precipition using universal kriging, in: Mrvar, and Ferligoj, A. (eds) Developments in statistics, FDV, Ljubljana, solvenia, 149-162.
10-Webester,R. and M.A. Oliver (2007),"Geostatistics for Environmental", John wiley& sons,chichester, West Sussex, England .
11- Soares, A., M. J. Pereira and R. Dimitrakopoulos (2008), "Geo ENV VI – Geostatistics for Environmental Applications", Springer, Lisbon, Portugal.