JOURNAL OF MECHANICS OF CONTINUA AND MATHEMATICAL SCIENCES
www.journalimcms.org

ISSN (Online) : 2454 -7190 Vol.-15, No.-4, April (2020) pp 105-124 ISSN (Print) 0973-8975

A COMPARISON OF TOPOLOGICAL KRIGING AND AREA TO POINT KRIGING FOR IRREGULAR DISTRICT AREA IN IRAQ

Amera Najem Obaid[1] , Mohammed Jasim Mohammed[2]

[1]GIS Center, Information Technology Office, Central Statistics Organization, Ministry of Planning

[2] Department of Statistics, College of Administration and Economics, University Of Baghdad, ORCID iD: orcid.org/0000-0003-1770-6301

Corresponding Author: Mohamed Jasim Mohammed

E-mail: m.asim@coadec.uobaghdad.edu.iq

https://doi.org/10.26782/jmcms.2020.04.00009

Abstract

The incidences of diseases (morbidities) vary across geographic areas. Spatial statistical analysis concerning spread and direction is useful to study such diseases in the neighborhood. This helps the health provenance for reducing this disease and control spatially it. Many spatial interpolations have employed for predicting the risky diseases based on observed values. In this paper, two methods of the spatial interpolation have studied based on unmeasured values from the same characteristic of spatial data, area-to-point kriging and topological kriging. These methods exploit variogram structure to predict the unmeasured values, then they fit this variogram by one of the parametric variograms. The de-regularization or deconvolution method is iterative and search model of area that reduces the variation between the theoretical semivariogram model and the fitted model for irregular area data. However, it is an approximate method for different regions based on the concept of average distance between irregular areas. Then, area to point kriging method has used using back calculation for approximated irregular areas in topological kriging (top kriging) .The prediction results for top kriging is better than other method. Disease krige map explaining the embedding risk of effective disease from observed frequencies are summarizes and their performances have compared .The goal of this paper is mapping and exploring the spatial variation and hot spots of district- level disease cases in Iraq country

Keywords : Geostatistics , Deconvolution ,Change the support, Interpolatio

I. Introduction

Spatial studies are characterized spatial models of disease risk. Spatial variations of disease risk are caused by grouping diseases, and that may reflect the existence of specific subpopulations that are under higher or lower contracting to the disease. The disease rates are available data of human health. This is often the number of neighbors or contagious patients, went with inside regions that can along wide Size extension, such as census or district units This information comprise of proportion of patient number over population size measured in irregular areas. This is a leaving from the classical disease risk and human health research [XV] .Many of geostatistics methods are particularly appealing, which allow the variable prediction and uncertainty in places where measurements are not available [IX]. The most advantage of methods for geostatistics is that they are the best BLUE, which means that an average quadrated error is small, linear means that the calculation is a weighted mean and objective means that the mean estimated mistake is zero [XIII]. Geostatistical approaches in medical research have been developed and are measured across their plotted Euclidian distance against variogram. From this, the variogram is exploited to estimate the unmeasured variable at the place of interest for a given area size from the point samples. The transform area to point is ignore important points From this, the variogram is estimated which is then used to estimate the unmeasured variable at the source location for specific area size from the point discretizing. The transform area to point is ignoring important points (a) Inaccurately, the local information is usually dropped into its single polygon centre; (b) Prediction accuracy should be equal to that datum (within error free), i.e. the total average of points in any area including an average areal datum(coherence); [XI]. Several authors have suggested kriging to predict point values from region data, an approach point out to as area to point kriging [XI] .In this paper, we ensure an empirical application for interpretation in which Top kriging method is compared with geostatistical ATP kriging method for predicting a smooth disease rate density surface from population data aggregated within district. We found that the two interpolation approaches produce the density surfaces with numerical and discrete error limits, depending of choosing a suitable point variogram within the geostatistical process. We partition this paper into two sections of theoretical and empirical methods of prediction unmeasured data for irregular area

II. M ethods and Materials

Corporation of spatial support in krigimg

Kriging is a spatial prediction forecasting method in which the prediction of a spatial variable unmeasured in areas is predicted as a weighted average in the corresponding areas. By supposing that the process is stationary that is mean the variance just a function in distance. Areal information characterized over various spatial support are interpolated to a discretizing to build map the embedding risk of building up the disease as a density surface. The area to point starts with represent spatial support and a spatial weight work in prediction of area versus area and area versus point.

Kriging of point in Area

In this subsection, suppose every spatial area A_α have equivalent shapes and sizes. These areas can be spatially georeferenced in polygon's center $s_\alpha = (x_\alpha, y_\alpha)$. Z values are determined as a linear blend for the Z-information in K adjacent areas

$$\hat{Z}(s_{\beta}) = \sum_{i=1}^{K} \lambda_i(s_{\beta}) Z(s_i) \tag{1}$$

$\lambda_i(s_{\beta})$ is the weight to random variable $Z(s_i)$ for the prediction at s_{α}. The size K weights are solved by linear kriging method of the next equation

$$\sum_{j}^{n} \lambda_i(s_{\beta}) C(s_i - s_j) + m(s_{\beta}) = C(s_{\beta} - s_i)$$

$$\sum_{j}^{n} \lambda_i(s_{\beta}) = 1 \qquad \text{K=1……, n} \tag{2}$$

If the estimate $\hat{Z}(s_{\beta})$ equal to the Z-data of exact characteristics of kriging, $Z(s_i)$ then β = i.

The variance of the prediction which co-operate with the estimate (1) is calculated by the standard kriging equation

$$\sigma^2(s_{\beta}) = C(0) - \sum_{j}^{n} \lambda_i(s_{\beta}) C(s_i - s_{\beta}) - m(s_{\beta}) \tag{3}$$

We can be estimated the point covariance from point variogram under the stationary condition and the equivalently the semivariogram $\gamma(h) = C(0) - C(h)$

$$\hat{\gamma}(h) = \frac{1}{|N(h)|} \sum_{N(h)} \left(Z(s_i) - Z(s_i + h) \right)^2 \tag{4}$$

, then We can computed of kriging system $\lambda_i(s_{\beta})$, the Lagrange multiplier $m(s_{\beta})$, and kriging variance $\sigma^2(s_{\beta})$.

Kriging of Area in Area

Kriging here refers to a situation in which the prediction and measurement domain are a region rather than points for a region of an irregular shape, Within their various polygon centroids, the information cannot be obviously collapsed Kyriakidis (2004) in (1) for the areal value the kriging calculation is

$$\hat{Z}(A_{\beta}) = \sum_{i=1}^{K} \lambda_i(A_{\beta}) Z(A_i) \tag{5}$$

And in(2)for areal value, the kriging calculation is(

$$\sum_{j}^{K}\lambda_i(A_\beta)\bar{C}(A_i,A_j)+m(s_\beta)=\bar{C}(A_\beta,A_i) \quad i=1,.........K$$
$$\sum_{j}^{K}\lambda_i(A_\beta)=1 \tag{6}$$

The primary change is that point kriging and area kriging is covariance point kriging of $\bar{C}(s_i-s_j)$ are changeable by the area covariance term $\bar{C}(A_i-A_j)=Cov(Z(A_i),Z(A_j))$ Similar to customary block kriging, The region covariance calculation C(h) between any two discretizing points is approximated to the point covariance C(h) average;

$$C(A_i,A_j)=\frac{1}{\sum_{a=1}^{p_\alpha}\sum_{a'=1}^{p_\beta}w_{aa'}}\sum_{s=1}^{p_\alpha}\sum_{s'=1}^{p_\beta}w_{aa'}C(s_a,s_{a'}) \tag{7}$$

where p_α and p_β are the number of points s_a and $s_{a'}$ used to discretize the two areas A_i and A_j , respectively. Then, the weight $w_{aa'}$ is computed from the product of area weights population size determined to the discretizing point $n(s_a)$ and $n(s_{a'})$

$$w_{ss'}=n(u_s)\times n(u_{s'}) \text{ with } \sum_{s'=1}^{p_\beta}n(u_s)=n(v_i) \text{ and } \sum_{s'=1}^{p_\beta}n(u_{s'})=n(v_j) \tag{8}$$

The kriging variance for the regional estimator is computed as

$$\sigma^2(A_\beta)=\bar{C}(A_\beta,A_\beta)-\sum_{j}^{n}\lambda_i(A_\beta)\bar{C}(A_i,A_\beta)-m(A_\beta) \tag{9}$$

Where $\bar{C}(A_\beta,A_\beta)$ is the within-area covariance that is approximated using (7) with i = j = β.

Kriging of Area to point

Kyriakidis (2004) proposed an area-to-point prediction approach that is a special case of kriging. This approach is a geostatistical prediction. This may expressly and reliably account for the differences in support between at hand areal data and detection following prediction of point process, yielding coherence (i.e. mass preservation or pycnophylactic) predictions.. One alternative is to make continuous maps of the source data that group in order to disperse or to interpolate area to point. The value of the attribute $Z(s_a)$ can be estimated as the next linear set of regional data at each disaggregate point of the region

$$\hat{Z}(s_a) = \sum_{i=1}^{K} \lambda_i(s_a) Z(A_i) \tag{10}$$

The next system is

$$\begin{aligned} &\sum_{j}^{K} \lambda_i(A_\beta)\bar{C}(A_i, A_j) + m(s_\beta) = \bar{C}(A_\beta, s_a) \quad i = 1, \dots\dots\dots K \\ &\sum_{j}^{K} \lambda_i(A_\beta) = 1 \end{aligned} \tag{11}$$

The difference between two kriging system of area and kriging system area-to- point in (6) and (11) is change of the right side in covariance $\bar{C}(A_i, A_j)$ of area kriging system (6) by the area to point covariance $\bar{C}(A_\beta, s_a)$,that is approximated by a procedure (7) by replacing $P_j = 1$. The coherence is gathering of the P_β point estimates in area A_β must equal the areal value $z(A_\beta)$

$$z(A_\beta) = \frac{1}{n(A_\beta)} \sum_{s=1}^{p_\alpha} n(s_a) \hat{Z}(s_a) \tag{12}$$

This characteristic of the ATP kriging estimator (12) is critical constraint, must be satisfied the same number of regional values used for the estimation of each of the p_β discretizing point us . In this case, the weighted average of the right side covariance terms of the (11) is equal to the right side covariance of the single area kriging system (6).

$$z(A_\beta) = \frac{1}{n(A_\beta)} \sum_{s=1}^{p_\alpha} n(s_a) \bar{C}(A_i, s_a) = \frac{1}{n(A_\beta)} \sum_{s=1}^{p_\alpha} n(s_a) \frac{1}{n(A_i)} \sum_{s=1}^{p_\alpha} n(u_{s'}) C(s_a, s_{a'}) \tag{13}$$ [1]

Thus the next relation between area to point kriging and area to area weights exists

$$\lambda_i(A_\beta) = \frac{1}{n(A_\beta)} \sum_{s=1}^{p_\beta} n(s_{a'}) \lambda_i(s_a) \tag{14}$$

kriging variance are computed of area to point estimator as

[1] C(us,us') Green function

$$\sigma^2(A_\beta) = C(0) - \sum_{i=1}^{K} \lambda_i(s_a)\bar{C}(A_i, s_a) - m(s_a) \tag{15}$$

Implication of spatial support deconvolution in seminvariogame

The execution of both area kriging and area-point kriging methods, one has to know the point in area covariance C(h), or identically the point of area semivariogram γ(h). At the beginning, this section presents the analytical expression for the regularization of point domain semivariograms of irregular spatial areas . Then presented new procedure is deconvolution. The common equation of related between point space and regularized semivariogram

$$2\gamma_A(h) = 2\bar{\gamma}(A(s), A(s+h)) - \bar{\gamma}(A(s), A(s)) - \bar{\gamma}(A(s+h), A(s+h)) \tag{16}$$

under the stationarity assumption the above formula becomes

$$\gamma_A(h) = \bar{\gamma}(A, A_h) - \bar{\gamma}(A, A) \tag{17}$$

Where $\bar{\gamma}(A, A_h)$ is area-to-area semivariogram value, it indicate the mean value of the point support semivariogram between an random point in the support A and another in the converted support A_h.

The second term, $\bar{\gamma}(A, A)$, is within area of semivariogram value. In practice, using the next approximation because the length h =|h| is very large compared to the support dimension

$$\gamma_A(h) = \gamma(h) - \bar{\gamma}(A, A) \tag{18}$$

The next mean used for estimating the same area semivariogram value:

$$\bar{\gamma}(A, A) = \frac{1}{p^2} \sum_{a=1}^{p} \sum_{a'=1}^{p} \gamma(s_a, s_{a'}) \tag{19}$$

Where P is the number of discretize points in the area A.

Regularization of Irregular Spatial Areas

In expansion to the preassumption of stationarity, (18) is derived for all the areas are regular in size and shape. Thus, the within-area semivariogram value $\bar{\gamma}(A, A)$ is constant. Another more common expression for the regularization is

$$\gamma_A(h) = \bar{\gamma}(A, A_h) - \bar{\gamma}(A, A) \tag{20}$$

The value of the semivariogram within the region is now different depending on the separating vector h as the field is different in size depending on the distance between them.

If the size and shape of the areas are not equal, the semevariogram is a function in distances within same area. The small areas are paired with the short distances, while the large areas are paired for the smallest area among them .The mean of within-area semivariogram values can be estimated for areas isolated by a given vector h

$$\bar{\gamma}(A,A)=\frac{1}{N(h)}\sum_{\alpha}^{N(h)}[\bar{\gamma}(A_\alpha,A_\alpha)+\bar{\gamma}(A_{\alpha+h},A_{\alpha+h})] \qquad (21)$$

Where $\bar{\gamma}(A_\alpha,A_\alpha)$ and $\bar{\gamma}(A_{\alpha+h},A_{\alpha+h})$ are estimated from (19), and h is calculated from

$$Dist(A_\alpha,A_\beta)=\frac{1}{\sum_{a=1}^{P_\alpha}\sum_{a'=1}^{P_\beta}n(s_a)n(s_{a'})}\sum_{s=1}^{P_\alpha}\sum_{s'=1}^{P_\beta}n(s_a)n(s_{a'})\|s_a-s_{a'}\|$$

If all the areas are the same, then $\bar{\gamma}(A_\alpha,A_\alpha)=\bar{\gamma}(A_{\alpha+h},A_{\alpha+h})$ ∀ h and (21) is equal to the constant $\bar{\gamma}(A,A)$. Similarly, the first term $\bar{\gamma}(A,A_h)$ in (20) is estimated as

$$\bar{\gamma}(A,A_h)=\frac{1}{N(h)}\sum_{\alpha=1}^{N(h)}\bar{\gamma}(A_\alpha,A_{\alpha+h}) \qquad (22)$$

For any two areas, Aα and Aα+h, the value of semivariogram of areas separated by h is computed as

$$\bar{\gamma}(A,A_h)=\frac{1}{P_\alpha P_{\alpha+h}}\sum_{a=1}^{P_\alpha}\sum_{a'=1}^{P_{\alpha+h}}\bar{\gamma}(s_a,s_{a'}) \qquad (23)$$

the number of points used to discretize the two areas Aα and Aα+h are P_α and $P_{\alpha+h}$ are, respectively.

Deregularization of the Regularized Semivariogram

Area support variograms can be extracted by point variograms regularization or quasi-point support variograms. This stands for a kind for convolution of point variogram, As seen in the above paragraph.

The de-regularization is the reverse process of variograms estimated from area support data and is essential to quantify the measurement inherent to the processes fundamental the prevalence being observed. The deduction for point support semivariogram of areal is not direct as the procedures for regularization mentioned above. In several instances, the information in the region in comparison to the area to be calculated are very small. In such instances, a quasi-point support($|A| \approx 0$)is the approximation, and the empirical variogram of the areal data can be regarded to point- support variogram estimator (h) [XI]. The deconvolution is reverse process to resolve inverse problems using an iterative procedure Implemented general method below:

1. The monitoring of the semivariogram of areal data $\hat{\gamma}_A(h) = \frac{1}{|N(h)|}\sum_{N(h)}\left(Z(A_i) - Z(A_i + h)\right)^2$ by identified a support point model and estimate the parameters(sill, range) of the semivariogram model utilizing fundamental rules of deconvolution.

2. Compute the theoretically regularized model using approximation (18) and compare it with the experimental curve $\hat{\gamma}_A(h)$
3. Adjust the parameters of the point support model so as to bring the regularized model in line with $\hat{\gamma}_A(h)$

No approval was formulated concerning the way of the parameters of the point support model should be to assess the equivalence between the experimental and theoretically regularized semivariogram models.

III. Topological Kriging

Topological kriging (top kriging) is method for predicting of unmeasured values of irregular area. top_kriging based on two assumption. The intrinsic stationarity assumption is the first process. Namely, the observations variance is only function in distance. The top kriging likes every type of kriging method depends on the quality of selecting the variogram model .The second assumption is that, the observed interest variable can be regarded as the result of a space continuous process. Generally the process cannot be observed at a local (point) scale, but as an integrated or average over a certain wider area interval, called spatial. This process can have the potential to inherent one person with a specific disease in health statistics. In most cases, the observed values can then be seen as the collapsed (linear averages) of the process's local observations over the support.

Change the support (regularization)

The spatially averaged of variable $\bar{z}(v_k)$ of aggregation is supposed to be indicative for a non-trivial Support Area:

$$\bar{z}(A_k) = \frac{1}{|A_k|}\int_{s\in v_k} z(s)ds \tag{24}$$

Where $|A_k|$ point out to the size of A_k and $z(s_i)$ is the variable value at location s_i and under the intrinsic stationary , first we assume constant mean such that

$E(z(A_i)-z(A_j))=0$ .second, if non-zero support A the point variogram in support is a function of separation distance. We should presume that a point variogram $\gamma(h)$ exists, even if the process itself cannot be calculated at a point scale. The describtion of point variogram is the scaling of the variation between areas and function of their spatial domain. Then the variogram of two observations with areas A_i and A_j can then be found through regularization variogram :

$$\gamma(h)=\frac{1}{2}\operatorname{var}\left(Z(A_k)-Z(A_l)\right)=\frac{1}{|A_k||A_l|}\int_{s\in v_k}\int_{s'\in v_k}\gamma_v(|s-s'|)dsds' \\ -\frac{1}{2}\left[\frac{1}{|A_k|^2}\int_{s\in v_k}\int_{s'\in v_k}\gamma_v(|s-s'|)dsds' + \frac{1}{|A_l|^2}\int_{s\in v_k}\int_{s'\in v_k}\gamma_v(|s-s'|)dsds'\right] \tag{25}$$

where and s *and* s' are vectors of position in each integration field. The integration of the semivariogram between the two areas in right term of (25) subtraction. The integration of the semivariogram has decreased the semi-variogram of Z(A) with increased support. The semivariogram is introduced into the kriging scheme and can be solved in the usual way in order to determine the weights λ_i in the formal kriging. The regularized variograms always gives a positive semivariogram.

Reduction for Regularization

The different pairs of distances in irregular area make the computation of semivarigram values by integration is very difficult and expansive .an iterative optimization is a procedure to fit the variogram model for experimental variogram . For each iteration step, all values for semivariogram recomputed among points within the irregular areas . instead of integrating the covariance function Gottschalk et al. (2011) suppose the derivation of the computation for regularization of covariances, by putting the averaged distance, d among all distance in different areas of covariance model ,this more often done by regularizing:

$$\overline{d}_{ij}=\frac{1}{|A_i||A_j|}\int_{A_i}\int_{A_j}\|s_a-s_{a'}\|\ ds_a\ ds_{a'} \tag{26}$$

The regularized semivariance is computing based on the two areas by using d_{ij} for semivariograms in next expression

$$\gamma_{ij}=\gamma_p(d_{ij})-0.5(\gamma_p(d_{ii})-\gamma_p(d_{jj})) \tag{27}$$

where d_{ii} is the average distances within area . The computation of semivariogram values by average distances is faster and mathematically simpler than integrating the semivariogram values for all distances between two areas and it only has to be done once. Though classical regularization indicates that for all candidate variogram models the regularized semivariogram must be recalculated. The impact of the approximation depends on the variogram and the area structure, but in [VIII] it shows that the approximation will more often than not produce a great result for different types of variograms, despite the fact that the finest can be expected for variograms close to linear form, particularly close to the origin.

Binned variogram modeling and estimating

Selecting a random discretization is conceivable, but tests Show that average grid discretion yields numerically better outcomes for the same number of discretionary focus. Since area estimates can differ by a few size orders, we use a flexible grid where the area is scale-up with resolution. This framework will begin with a rough grid trying to cover the source area. This framework will be streamlined for a given region until the minimum number of points required is within the area. This method also means that points used to discrete a large support can again be utilized if smaller supports are discretized in the big one, Sub-district of greater district. After discretizing method we need to model variogram for using in top kriging method. A sample variogram based on the observations supports both empirical variogram and variogram clouds will be calculated using the Variogram function. The empirical variogram is not depend solely on distance but also on support field combinations. It is hence based on three independent variables (distance, smallest support, largest support) rather than a point variogram that is based only on a distance function. This can also be interpreted as containing a group of distance-based variograms for various support area groups. The distance and support distance are generated in the log10-domain with regular spacing. The distance between the center s of mass for the two regions is used for the distance axis. Due to the different support of the results, the theoretical point variogram cannot be construct directly from the sample data. Alternatively, the fitted Variogram optimizes a point variogram whose regularized semivariogram values make the sample variogram values best suited. This could be called a sample variogram back-calculation [XIII]. It is similar to the deconvolution approach defined in [VIII]. Regularized semivariogram values are derived from the variogram model for all combinations of support sizes and pair distances of observations, or from the average distances and areas of the empirical variogram.. If binned observations are used, the support is approximated by squares, allowing partially overlapping of the areas. There are various models of variogram available, as well as different methods of fitting with the least squares. The optimization process is depend on the method of Shuffle Complex Evolution [VI].

Shuffled complex evolution

The technique of probabilistic global search, which aims to mix the impact of simplex search with the concepts of random search managed, competitive evolution, and shuffling of complexes or groups, has been shuffled by [VI]. The algorithm starts by choosing s points randomly in the domain of the parameter such that s = p × m, where; p is the number of complexes each with m points. After evaluation of the function, the points are sorted in order of the increasing value of the objective function and are placed in a matrix that is partitioned into m irregular grid .after that we applied the competitive complex evolution (CCE) method is reprocess independently of each irregular grid (complex) .

IV. Empirical Application (Case Study)

Data and management

The second leading cause of child mortality worldwide is the prevalence of diarrhea among children under the age of five and the related drought induced by the lack of huge quantities of water and mineral salts [XVI]. The data were obtained from the Multiple Indicator Cluster Surveys (MICS) performed in cooperation with the Ministries of Health of the Center and in the area of Kurdistan and with technical expertise and the Central Statistical Organization. This reflects Iraq's population, with a representative sample of 20,520 families in Iraq at the national and Governorate levels. All diarrheal patients who were covered, verified for recurrence and linked to a true spatial location of the census (2009) and residents of each region in Iraq who were accessed from GIS at the Central Statistical organization. Our goal is to predict of this unscaled location in two methods (topological kriging and area to point kriging) and to compare between them.

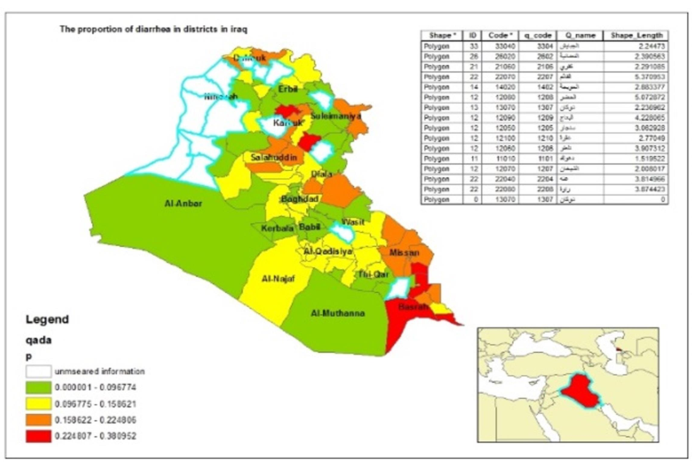

Fig.1. the unmeasured district of disease

V. Application

Applied variograms of topological kriging and area to point kriging

Both methods depends heavily on spatial weighted structure (variogram). This variogram used in solution of kriging equations system for prediction of diarrhea risk disease (average over an area) in the Iraq district. The shape file of administrative boundary of district Iraq are taken from GIS center ,Central Statistical Organization (103 district, 87 is available data and 16 unmeasured it) .

Topological kriging variogram

The sample variogram estimated of Top -kriging an from binned variogram as explained in section 3.2, and it is fitting the variogram model by the parametric fractal model

$$\gamma_p(h) = \alpha h^{\beta}(1 - e^{-(h/\delta)^d}) + C_{0p}$$

α is linked to sill parameter, δ correlation length , β and d are parameters of the long and short distance slope of the variogram [XIII].

Table-1: the estimated parameters of fitted variogram

Parameters	Value
a	4.855870e-05
b	3.032003e+01
c	4.610171e-02
d	6.534774e-01
SSErr	0.279933
criterion	0.06233653

Two important function for improvement in the result includes function convergence for iteration and parameter convergence for estimated parameters of fitted variogram . The optimization procedure has stopped where the two convergences are below .

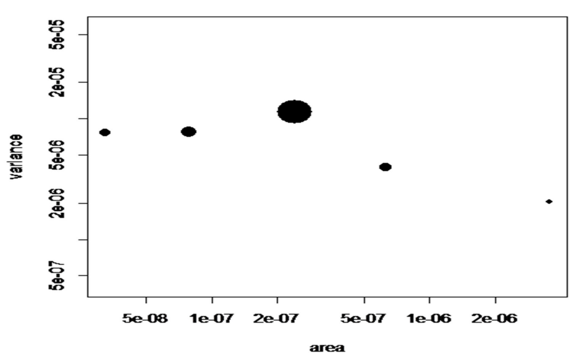

Fig. 2. Relation between observation variances in spatial and region size

In Fig. 2, relation between realization variances in spatial and region size has depicted. The realization have been clustered in groups corresponding to their area, and the variance of the realization computed from every area bin. The implication for the Topological kriging are that the lower realization variances in high area size,

which should be evident from this figure. Circle sizes are rational to the number of observations in every regional category.

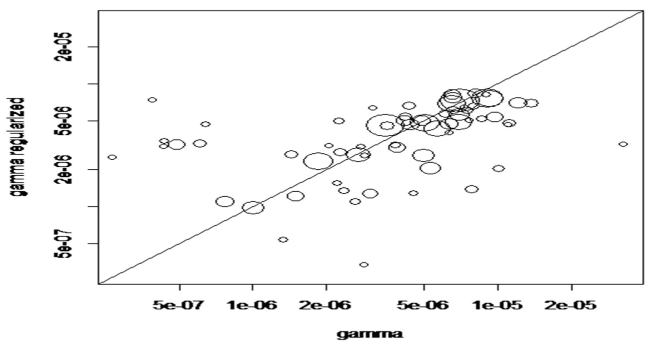

Fig. 3. Comparing the experimental and regularized semivariogram.

In Fig. 3, the diagonal line means an ideal fitting for log-log plot of values for regularized semivariogram against the experimental (observed) variogram values from a fitted variogram model. We can see that the vast majority of the focuses are based on the slanting line. Such deviations can, in any case, be normal, like the distinction between a cloud variogram and a fitted variogram model. The size of the circles is identified with the number of sets in every bin for binned variograms, indicating that the largest anomalies are fundamentally little pairs in bins.

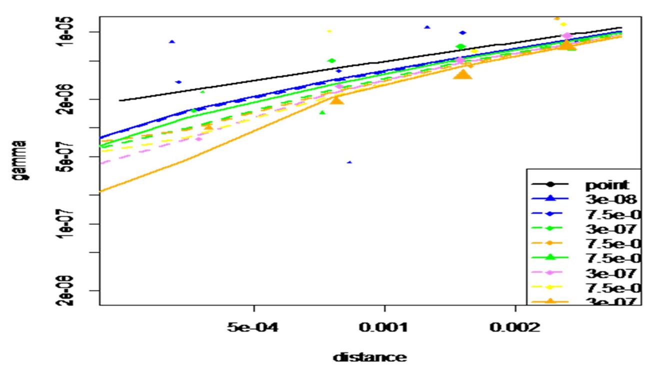

Fig. 4. observed variogram , regularized variogram and fitted model variogram

Fig 4 explains some chosen bins variogram against the regularized semivarigram of these bins. The numbers in the key indicate groups of regions, and means a regularized variogram as a function of Ghosh distance and region. That regulation is not quite capable of reproducing the variance reduction as a function of the region's size, especially for large distances. It has sampled on average 51.4 points from 16 areas creating prediction semi variance matrix.

Area to Point Kriging Variogram

The semivariogram of area to point kriging can be calculated for each area of irregular area. These must be function of distance since the district are unequal in size and shape from 20 and 21 . The small areas are paired with the short distances, while the large areas are paired for the smallest area among them. The mean of within-area semivariogram values can be estimated for areas spaced by a given vector h. The term semivariogram from one area to another equals the sill of the specified varoigram models (spherical or fractal). Therefore, the calculation of the term can be somewhat mitigated. The variation from area- to -point is computed by averaging the covariance values $\overline{\gamma}(A, A_h)$ between the point location and a set of discretizing points of the area [XV].

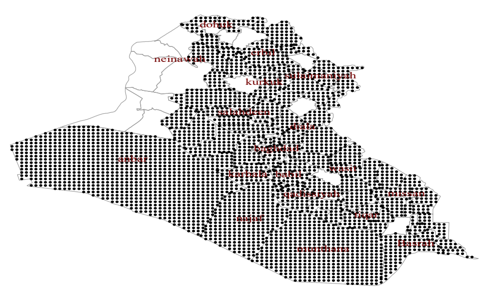

Fig.5. Discretizing points of measured district of Iraq

The deconvolution was conducted discretizing points of irregular district areas including 3619 point (above figure). The discretizing point numbers in Iraq district are uneven and vary from 10 to 933, at 10 kilometers resolution. This optimization method was applied to the regularized models .If the difference statistics between regularized and fitted variogram model are less than one percent as in section 2.2, the optimization method is stopped. After 9 iterations for a region, the results of parameters of fit variogram estimation in Table 2 is depicted as below:

Table-2: the estimated parameters of area and point variogram

Parameters of points	Value	Parameters of area	Value
Sill	3.682128e-05	sill	3.672764e-05
Range	1.65318	range	1.646891
Model	Exp	model	Exp

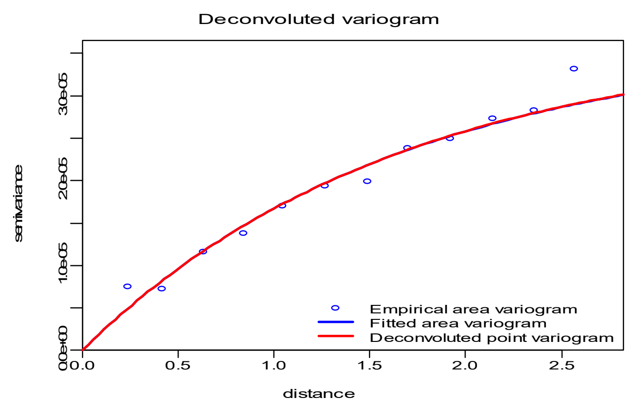

Fig.6. Comparing between fitted and deconvolution variogram

Cross-validation

The results of cross-validation for area to point kriging and Top kriging respectively in three tables below:

Table -3: Cross validation prediction at 10 locations in ATP-kriging methods

Area Id	longitude	latitude	observation	prediction	Var.
1	48.19583	30.09235	0.000255708	0.002080536	2.08673E-06
2	47.98701	30.34886	0.000348784	0.000620087	8.37952E-07
3	47.58872	30.61519	0.000536986	0.002454286	1.20130E-06
4	47.86979	30.77192	0.00066371	0.000398563	1.37906E-06
5	46.52053	30.87212	0.000733174	0.002370908	1.56822E-06
6	47.27059	31.03247	0.000867474	0.001628318	7.75142E-07
7	47.47391	30.97179	0.000980616	0.00080608	4.37074E-07
8	45.37287	30.09589	0.000982795	0.003950199	2.16761E-06
9	45.33639	31.28413	0.001076909	0.001508242	1.12E000-06
10	46.25426	31.11072	0.001100543	0.000992763	3.31175E-07

Table- 4: Cross validation prediction at 10 location in Top-kriging methods

Area ID	longitude	latitude	observation	prediction	Var.
1	48.19583	30.09235	0.000255708	0.002080536	2.08673E-06
2	47.98701	30.34886	0.000348784	0.000620087	8.37952E-07
3	47.58872	30.61519	0.000536986	0.002454286	1.20130E-06
4	47.86979	30.77192	0.00066371	0.000398563	1.37906E-06
5	46.52053	30.87212	0.000733174	0.002370908	1.56822E-06
6	47.27059	31.03247	0.000867474	0.001628318	7.75142E-07
7	47.47391	30.97179	0.000980616	0.00080608	4.37074E-07
8	45.37287	30.09589	0.000982795	0.003950199	2.16761E-06
9	45.33639	31.28413	0.001076909	0.001508242	1.12000E-06
10	46.25426	31.11072	0.001100543	0.000992763	3.31175E-07

1.

Table -5: statistics (R^2 ,mean absolute error and mean error) for two methods

Cross validation statistics	Top kriging	area to point kriging
R^2	0.8434557	0.8304696
ME	3.33E-05	2.06E-05
MAE	0.001494771	0.001563139
RMSE	0.002708977	0.002819277

The two methods of geostatistical prediction can also compute the prediction error in expansion to the prediction itself in the formula of an a kriging standard error.

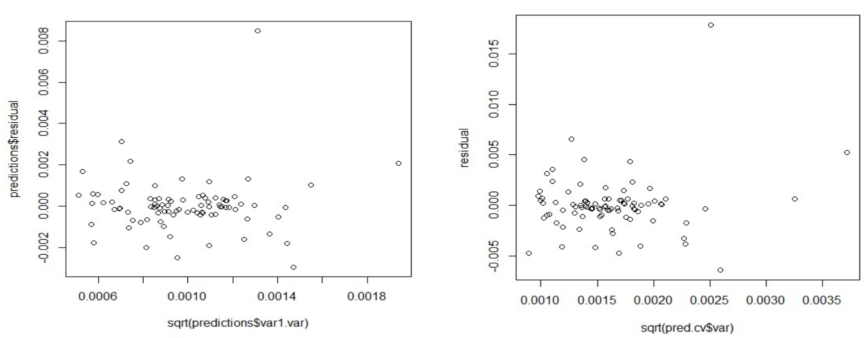

Fig.7. Comparing between cross validation residual and kriging standard error in two methods

Figure 7 displays the relation between the cross validation residuals and the kriging standard error of top kriging. The left explains the histogram of top-kriging. We can see from all district , if the residuals are low values the kriging standard are low values from two methods. The top kriging residuals are in most cases are smaller than ATP kriging standard error .

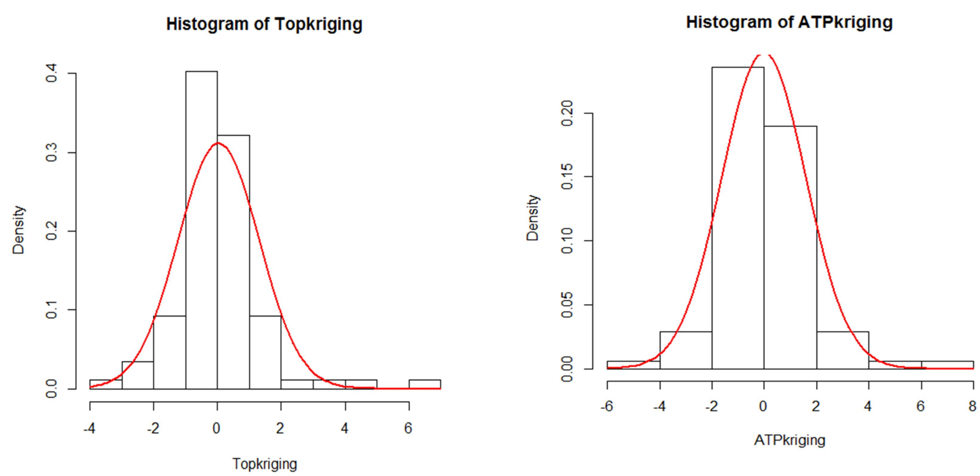

Fig.8. Comparing between zscore histograms and fit in two methods

In Figure 8, the standardized kriging residual should have a customary standard normal error in topological kriging and it deviates less from this ATP kriging . These results are clear from log likelihood of two methods fitted. The top kriging is higher than ATP kriging.

Finally, we obtain the disease prediction map of Iraq district, where the predicted diarrhea proportion on area from top kriging are on the top panel left. ATP kriging stands for top panel on the right. We can note the variance of ATP kriging higher than Top kriging method

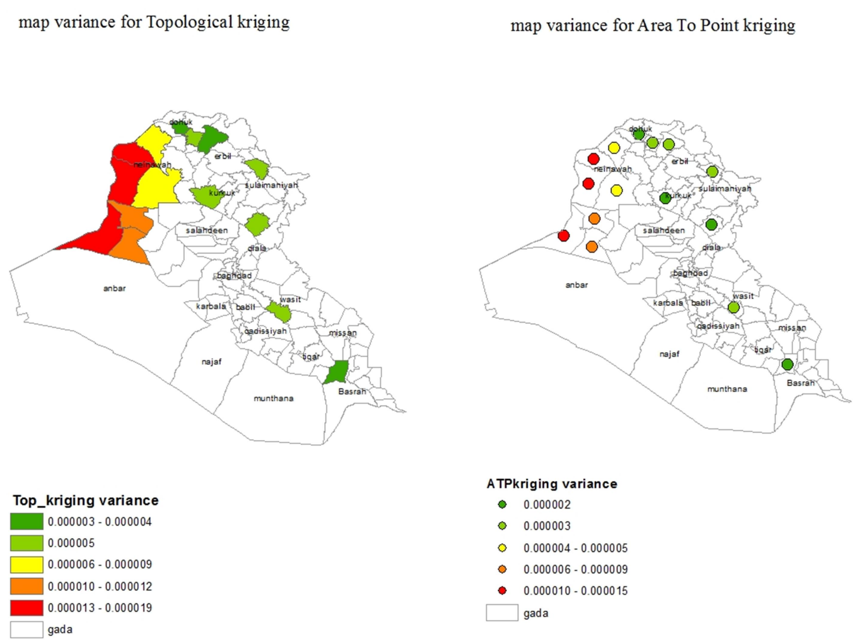

Figure 9. Prediction and variance map of two methods

VI. Conclusions

The geostatistical prediction of an area in different size and shape has seldom been applied in practice. Primarily, it is because of complexity of computation, but in kriging methods, it can measure the dynamic spatial of social and health sciences. The advantage of non-overlapping data (administrative boundary) is relied on the regularity limit of boundary of district. For uncertainty estimator of prediction in all cases, the best method is topological kriging. Uncertainty area to point kriging in sampling for greater support is not taken into account. The coefficients of determination (R^2) of the predictions indicate that top kriging performs more accurately than the other methods. Top kriging supposes linear combination as a linear estimator which ensures the mass conservation over district unit of variables. The method also estimates kriging-standard errors which can be used to evaluate prediction uncertainty with ordinary kriging from this kind of not estimated data properly. A possible reason for the small differences is that districts in Iraq are in different irregular areas, which contributes to the large differences in the observed values of specific extraction . Two kriging approaches applied to estimate the spatial risk of diarrhea distribution in data-limited settings. This calculation will help local public health agencies assess the disease's burden at all locations and define geographic areas. The determination of hotspots will lie in planning in-depth site control strategies rather than in planning the entire region. In addition to the available schemes, dedicated interventions, allocating funds by region, and programs, it will reduce the spread of diarrhea, ensuring better child health. Weight of the infection at all areas, recognize spatial region. The assurance of hotspots will lie in and out location control techniques as opposed to arranging the whole locale. Notwithstanding, the accessible plans devoted intercessions and designating assets by area. This project will decrease the spread of looseness of the disease of diarrhea, ensuring better child health.

References

I. A. Soares et al. (eds.), geo ENV VI – Geostatistics for Environmental Applications, 3-22 , Springer Science& Business Media,vol .15 ,2008.

II. A .Minh, Muhajarine N, Janus M, Brownell M, Guhn M. A review of neighborhood effects and early child development: how, where, and for whom, do neighborhoods matter? Health & place;46:155 -74, 2017.

III. Atkinson, Peter, Jingxiong Zhang, and Michael F. Goodchild. Scale in spatial information and analysis. CRC Press, 2014.

IV. B .Ghosh, Random distances within a rectangle and between two rectangles.Bull. Calcutta Math. Soc. 43, 17- 24, 1951.

V. CA .Gotway, Young LJ , Combining incompatible spatial data. J Am Stat Assoc 97(459):632-648,2002.

VI. Duan, Qingyun, Soroosh Sorooshian, and Vijai Gupta. "Effective and efficient global optimization for conceptual rainfall-runoff models." Water resources research 28.4: 1015-1031, 1992.

VII. Gotway Crawford C.A., Young L.J. , Change of support: an inter-disciplinary challenge. In: Geostatistics for Environmental Applications. Springer, Berlin, Heidelberg, 2005.

VIII. Gottschalk, Lars, Etienne Leblois, and Jon Olav Skøien. "Distance measures for hydrological data having a support." Journal of hydrology 402.3-4 ,415-421,2011.

IX. G. Journel, A.., Huijbregts, C.J., Mining Geostatistics. Academic Press, London, UK, 1978.

X. H .Kupfersberger, Deutsch CV, Journel AG Deriving constraints on small-scale variograms due to variograms of large-scale data. Math Geol 30(7):837-852, 1998.

XI. Kyriakidis, P., A, geostatistical framework for area-to-point spatial interpolation. Geograph. Anal. 36 (3), 259 -289,2004 .

XII. N.Cressie, Statistics for Spatial Data. Wiley, a Wiley-Interscience-Publication, New York, NY, 1991.

XIII. O. Skøien, J., Merz, R., Blöschl, G., Top-kriging - geostatistics on stream network s. Hydrol. Earth Syst. Sci. 10, 277- 287, 2006 .

XIV. O .Skoien, J, Bloschl, G., Laaha, G., Pebesma, E., Parajka, J., Viglione, A., Rtop: An R package for interpolation of data with a variable spatial support, with an example from river networks. Computers & Geosciences, 67, 2014.

XV. P .Goovaerts,, “Kriging and semivariogram deconvolution in the presence of irregular geographical units”. Math. Geosci. 40 (1), 101 - 128 , 2008.

XVI. www.unicef.org