International Journal of Advanced Statistics and Probability, 5 (2) (2017) 91-95

International Journal of Advanced Statistics and Probability

Website: www.sciencepubco.com/index.php/IJASP
doi: 10.14419/ijasp.v5i2.7824
Research paper

Estimation method of spatial geostatistical data : Application to rainfall data

HamidReza Erfanian *, Samaneh Barati

Faculty of Basic Sciences and Modern Biological Technologies, University of Science and Culture, Iran
**Corresponding author E-mail: erfanian.hamidreza@gmail.com*

Abstract

Restriction of water resources for agricultural and non-agricultural purposes has caused major difficulties and rainfall is considered as one of the most important water resources. Therefore, predicting rainfall and estimating its rate monthly or annually for each region as one of the most important atmospheric parameters, is of particular importance in optimized usage of water resources.
In this paper in addition to the presenting application of novel statistical methods, prediction of rainfall amount has been performed for the entire map of Iran. In this analysis, data of average rainfall of 108 pluviometry stations in different cities of Iran have been used and zoning of rainfall has been prepared for the country.

Keywords: *Geostatistics; Prediction; Rainfall Data; Spatial Statistics; Universal Kriging.*

1. Introduction

The problem of obtaining values, which are unknown, has drawn attention of many science researchers. For reasons of economy, there will always be only a limited number of sample points, where observations are measured. Hence, one has to predict the unknown values at unsampled locations of interest from the observed data to obtain their values [5]. To develop a reliable, accurate, and continuous surface prediction of values at locations without measurements is an essential task, which we frequently encounter in environmental and health-related studies [3]. For this sake, there exist several prediction methods for deriving accurate predictions from the measured observations. [5]
In this paper, a dataset related to average of rainfall in 108 cities of Iran has been used. This average is from daily rainfall through the September the 23 rd 2015 up to the December the 24 rd 2015 that has been recorded in the pluviometry stations. For each station, the average of rainfall has been measured in millimeter, and its latitude and longitude has been measured in degree. This average is specified just in the pluviometry stations. To predict the rainfall at locations without measurements geostatistical method, which here is Universal kriging, has been proposed. However, there is a wide range of techniques available for spatial interpolation. The advantages and limitations of which are widely discussed in the scientific literature [2]. In principle, these techniques are classified as deterministic (the nearest neighbour and polynomial regression) or stochastic (geostatistical approaches as kriging and cokriging) [3].
The common method for spatial prediction in climatology for about half a century was Thissen Polygon. Until Matheron in the middle of 1960, founded the basics of the geostatistics[6] and it quickly developed to predict environmental variables, particularly in climatology. Geostatistical methods are more accurate than the other methods specially mentioned method, due to the using of spatial correlation structure among the observations, which generally is modeled by variogram function. Furthermore, in kriging methods, the variance of the prediction is presented in each point, which is a particular property of these methods [3].

2. Methodology

For spatial analysis of this data set, exploratory data analysis has been done first.
As John Tukey says, exploratory data analysis is detective in character. It uncovers indications, usually quantitative ones. Exploratory data analysis can never be the whole story, but nothing else can serve as the foundation stone, as the first step."EDA" is an approach to analysis the data and has numerous methods (mostly graphical) so that we can obtain maximum information from data [8]. For spatial analysis, especially, kriging methods, we check the fundamental assumptions such as normality, stationary on average, isotropic and existence of outlier data, because the mentioned analysis has founded based on these assumptions.
To display a variable changes considering the distance, semivariogram function has been used such that, for any distance vector h, the increment $z(x+h)-z(x)$ has zero expectation and finite variance, which is independent of location x .The variance of this increment defines semi-variogram given by:

$$\gamma = \frac{1}{2} Var\left[z(x+h) - z(x) \right] \tag{1}$$

The experimental semi-variogram (by grouping the data pairs according to their distances) from the measured data points can be obtained by:

$$\gamma^*(|h|) = \frac{1}{2N(|h|)} \sum_{i=1}^{N(|h|)} \left[Z(x_i) - Z(x_i + h) \right]^2 \tag{2}$$

[1].Where $\gamma^*(h)$ is the value for semivariogram in distance (h), $Z(x_i), Z(x_i+h)$ are the variable values at locations x_i , x_i+h and $N(h)$ is the number of pairs belongs to each distance (h).

There are Different theoretical models that have been proposed to fit on an experimental variogram such as, Gaussian, exponential, spherical, linear, Matern and We finish with our preparation of the variogram and can begin with kriging prediction.

Kriging is an optimal or best linear unbiased prediction (BLUP) method. The French mathematician Georges Matheron (1963) named this method kriging, to gratitude efforts of D. G. Krige, south African African mining engineer (1951). There, kriging served to improve the precision of predicting the concentration of gold in ore bodies. However, the object "optimal linear prediction" even appeared earlier in literature, as for instance in Wold (1938) or Kolmogorov (1941a). However, very much of the credit goes to Matheron for formalizing this technique and for extending the theory. The main idea of kriging is that near the sample, points should get more weight in the prediction to improve the estimate [5]. This estimator is defined as a linear combination as bellow:

$$Z^*(x)=\sum_{i=1}^{n}\lambda_i z(x_i)+k \qquad (3)$$

Where $Z^*(x_i)$ is the estimation of the value of the variable Z at location x and λ_i is the allocated weight to the values of Z at location x_i .

Coefficients λ_i, k are chosen so that $Z^*(x)$ be unbiased and has the least square of errors.

If the mean of random field $z(x)$ is constant, known or unknown, to satisfy unbiasedness condition, we must have $k=0$ and $\sum_{i=1}^{n}\lambda_i=1$ .Thus, depend on the mean of random field $z(x)$ be known or unknown and constant or unknown and inconstant, kriging is classified into simple kriging, ordinary kriging and universal kriging. It means that for example when we want to predict the values of locations using universal kriging, the mean of random field must be unknown and inconstant [1].

If the desired variable does not have normal distribution, nonlinear kriging should be used or we should convert variable distribution to the normal distribution using data transformation methods [2].

3. Results

Figure 1 shows geographical locations of the sampled stations with the values of quartiles of the rainfall data.

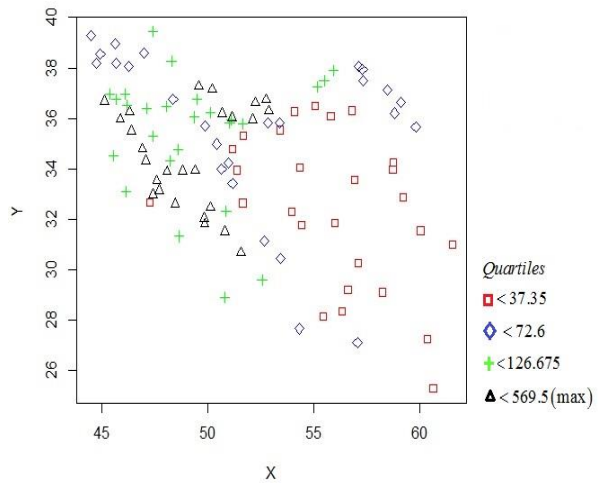

Fig. 1: Distribution Status of Rainfall Observation Stations.

Descriptive statistics of studied variable are given in Table 1.

Table 1: Descriptive Statistics

Statistics	Amount of rainfall(mm)
Minimum	6.8
Mean	106.75
First quartile	37.35
Median	72.6
Third quartile	126.68
Maximum	569.5
Std.Deviation	107.84
Skewness	2.38

After checking normality of data using q-q plot, it was found that the average of rainfall has not been normal distribution. Therefore, we took a natural logarithm from the data, thus we normalize them. We performed Shapiro-Wilk test to ensure that data have normal distribution.

The q-q plot of transformed data and the result of Shapiro-Wilk test have been shown respectively, in Figure 2 and Table 2.

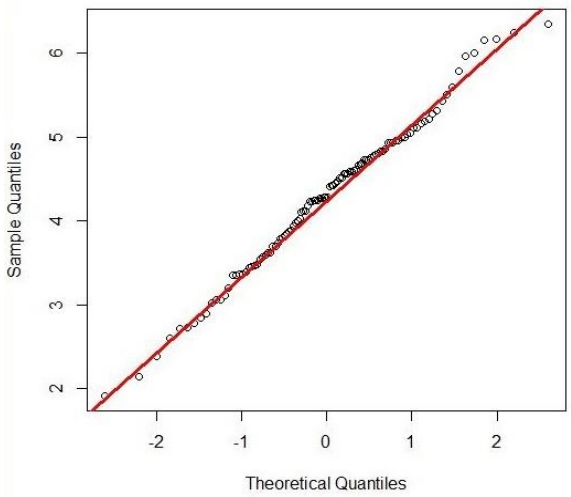

Fig. 2: Normal Q-Q Plot of Transformed Rainfall Data.

Table 2: Shapiro-Wilk Normality Test (for Transformed Rainfall Data)

W	0.9908
$p-value$	0.6795

To consider existence of outlier data in spatial data, two types of outlier data should be considered.

First type is similar to outlier data in classical statistics that can be recognized with the help of box plot or steam and leaf plot. Second type is outlier data in a neighborhood and can be recognized with the variogram cloud or H-scatter plot. According Figures 3 and 4, there is no outlier data in our dataset.

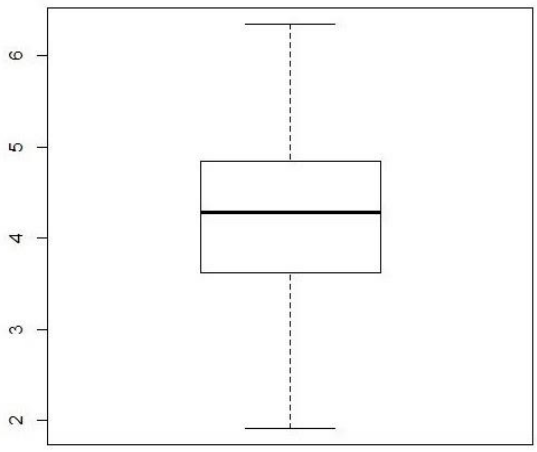

Fig. 3: Boxplot.

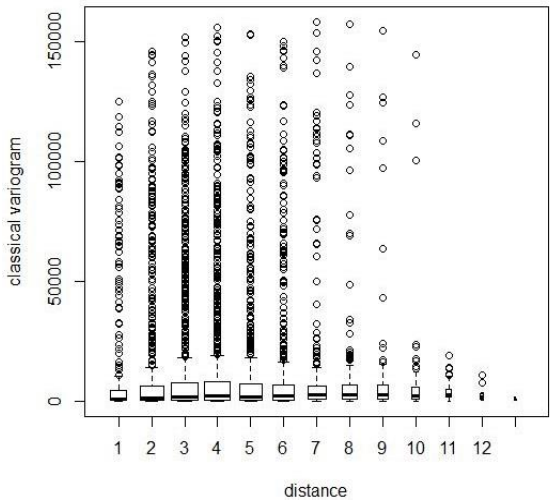

Fig. 4: Variogram Cloud.

To consider if there is a trend in our data or not, we plot the transformed rainfall data in front of latitude and longitude separately. Regarding Figure 5, there is a trend in the observations in direction of x.

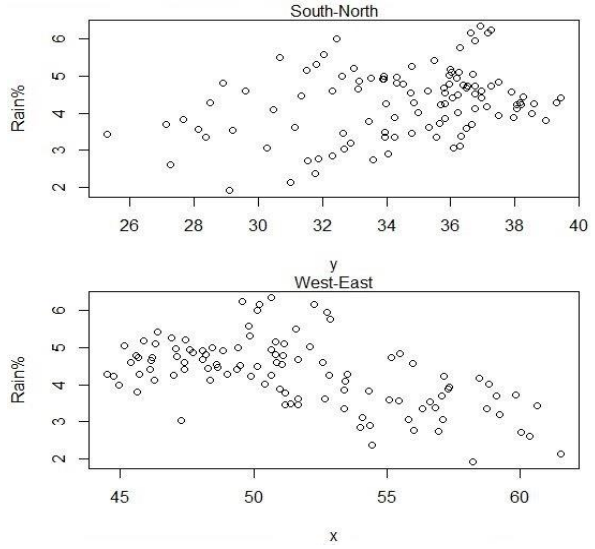

Fig. 5: Rainfall Data along Latitude (Top) and Longitude (Bottom).

To consider isotropy, we plot the variogram, which is a tool to model spatial dependence structure, in four different directions in Figure 6 because four plots are almost conformed to one another, so isotropy is being established.

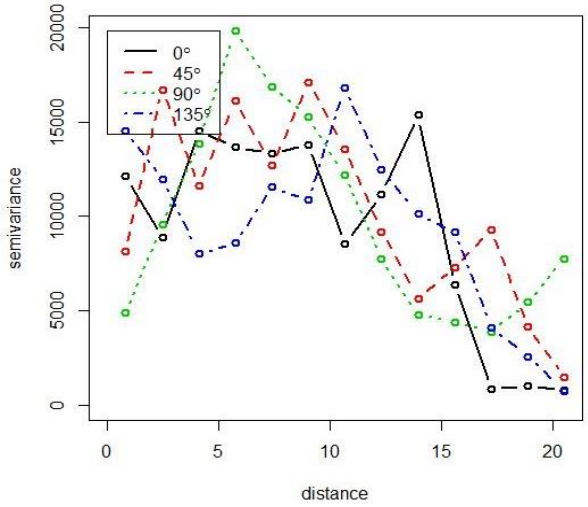

Fig. 6: Directional Variogram.

To remove the trend along longitude, we fit a regression model to our data, as follow:

$$fit <- lm(s \sim x + x\,\hat{}\,2 + x\,\hat{}\,3) \quad (4)$$

Where s is transformed rainfall data which is $\log z$?

Now if we plot the residuals of this regression model in front of longitude, no trend is seen. Figure 7 implies this fact.

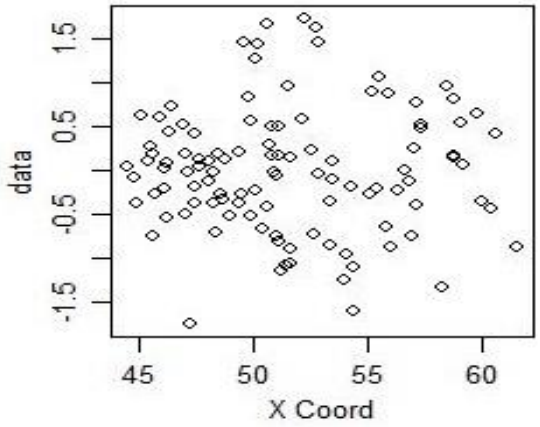

Fig. 7: Residuals of Fitted Regression Model along Longitude.

Figure 8 shows the empirical (semi) variogram of the observations.

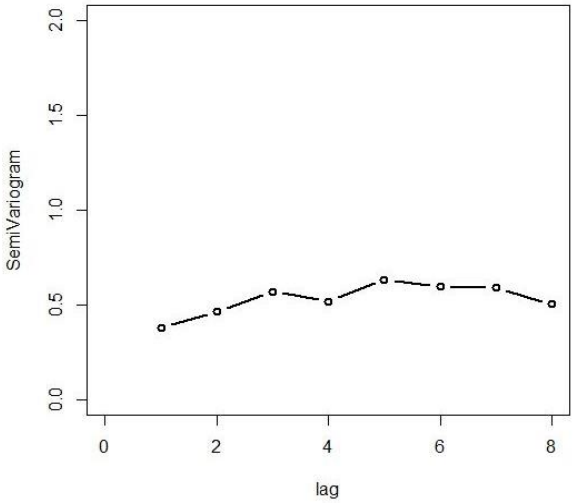

Fig. 8: Empirical Variogram.

Now, we can fit a theoretical variogram to the empirical variogram. The parameters of variogram were estimated by ols method, which is a stand for "ordinary least square" method. According to cross validation between the different theoretical models (included spherical, Matern, exponential and linear models), Matern model was selected as the best model.

Semivariogram model was obtained as follow and its plot has been drawn in Figure 9.

$$\gamma(h) = \sigma^2 \left\{ 1 - \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{|h|}{a} \right)^{\nu} k_{\nu} \left(\frac{|h|}{a} \right) \right\} \quad (5)$$

Where $\sigma^2 > 0$ is the variance, $a > 0$ is range which is here equal to 1.45, $\nu > 0$ is shape parameter which is here equal to 0.5, $\Gamma(.)$ is the gamma function, $k_{\nu}(.)$ is the modified Bessel function of the second kind and order ν and $|h|$ is the norm of vector h [7].

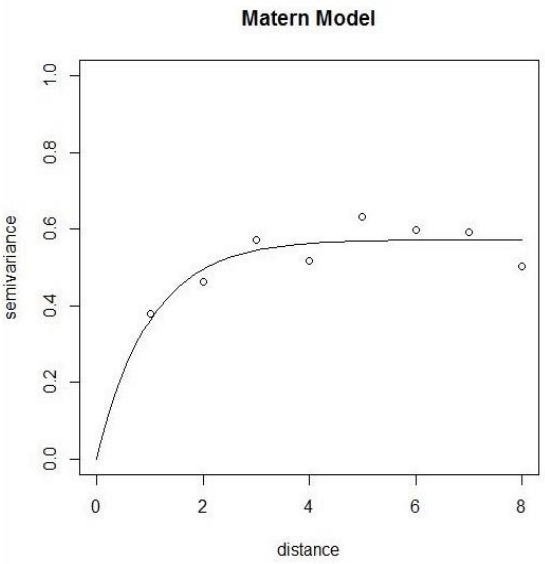

Fig. 9: Theoretical Fitted Variogram.

To predict using universal kriging, final prediction in each location is the sum of estimated trend and prediction of residual of estimated regression model in that location. Figure 10 shows universal kriging of average of rainfall in the country. As can be seen, rainfall amount decreases in central desert areas and east of the country.

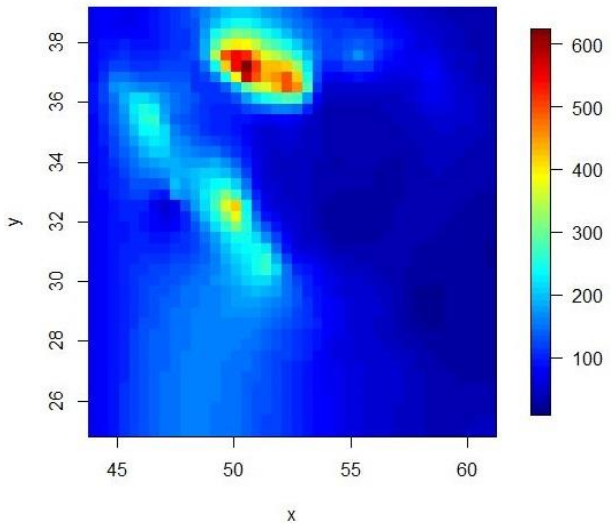

Fig. 10: Universal Kriging Prediction Map of Average of Rainfall Data.

Figure 11 shows variance of universal kriging. Variances of predictions of rainfall amount of the country indicate that predictions in areas where few observations are available have high variance but in the other areas, predictions have high accuracy.

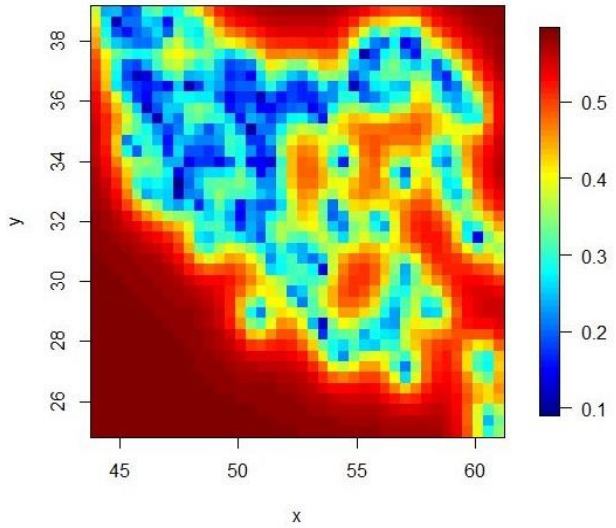

Fig. 11: Map of Predictors Variances.

Table 3 shows that for sampled stations, prediction returns observed values and this indicates that the prediction is precise. The other 5 locations are the new locations that we want to predict amount of rainfall there. These five locations are specified in Figure 12 with red points.

Table 3: Universal Kriging Prediction for A Sampled Location and 5 New Locations and Their Variances

Location	real rainfall amount(mm)	Predicted rainfall amount(mm)	Variance
(45.7,38.17)	72.5	72.5	0
(50,30)	-	126.29	0.508
(55,35)	-	31.84	0.466
(48,34)	-	134.49	0.117
(51,38)	-	316.02	0.499
(47,33)	-	102.02	0.25

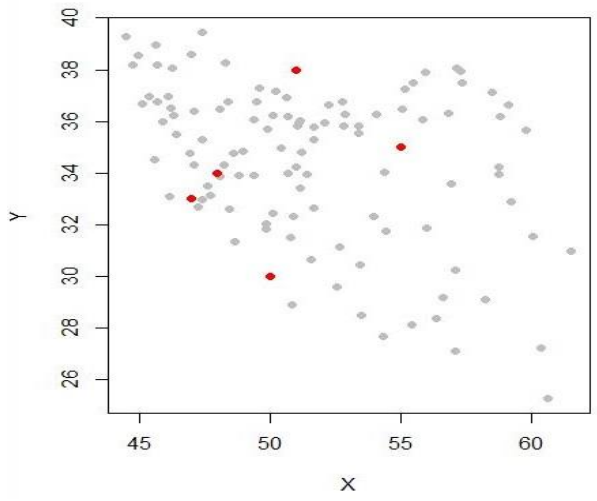

Fig. 12: New Locations to Predict Their Rainfall.

4. Conclusion

This paper discusses about the application of universal kriging to interpolate rainfall amount and the estimation variance for whole country, Iran.

Before performing geostatistical calculations, common statistical surveys, on the 108 observation stations, such as a normality test of data distribution were done. By using Shapiro-Wilk test, it was determined that the average of rainfall data does not follow normal distribution. Of course the high skew ness of this parameter that was presented in Table 1, indicates this too. This dataset became normal using logarithm transformation, which is a special case of Box-Cox normalizing transformations family.

Several polynomial trend and semi-variogram models for the residuals were tried before performing the geostatistical interpolation. Matern model was selected as the best theoretical variogram model.

By using universal kriging, rainfall amounts that cannot be measured can be adequately described by this model, which are required for the water resources planning and management for the country. In addition, the result of interpolation of average of rainfall reveals that there is a decrease in the rainfall amount from west to east and from north to central desert areas. The map of estimation variance for our data concluded that wherever there is significant variation in the estimation of a rainfall amount, it is due to the absence of monitoring stations.

References

[1] B V N P KAMBHAMMETTU, P. Allena and J. King, Application and evaluation of universal kriging for optimal contouring of groundwater levels, Journal of earth system science,vol.120,No. 3,(2011), 413-422. https://doi.org/10.1007/s12040-011-0075-4.

[2] N. Cressie, Statistics for Spatial Data, John Wiley & Sons, New York,NY,USA,1993.

[3] I. Hunova, J. Horalek, M. Schreiberova and M. Zapletal, Ambient Ozone exposure in Czech Forests: A GIS-based Approach to spatial distribution assessment, The Scientific Word Journal, volume 2012,(2012) .

[4] A.G. Journel and C.J. Huijbregts, Mining Geostatistics, Academic Press, London,1978.

[5] A. Lichtenstern, *Kriging methods in spatial statistics*, Bachelor's thesis, Munchen University,2013.

[6] G. Matheron,G, Principles of geostatistics, Economic geology,vol.58,(1963),1246-1266. https://doi.org/10.2113/gsecongeo.58.8.1246.

[7] E. Pardo-Iguzquiza and M. Chica-Olmo, Geostatistics with the Matern semivariogram model: A library of computer programs for inference, kriging and simulation, Computers & Geosciences, vol. 34,(2008), 1073–1079. https://doi.org/10.1016/j.cageo.2007.09.020.

[8] J. Tukey, Exploratory data analysis, Addison-Wesley Publishing Company,1976.