Journal of Basic and Applied Engineering Research
p-ISSN: 2350-0077; e-ISSN: 2350-0255; Volume 3, Issue 1; January-March, 2016, pp. 1-5

http://www.krishisanskriti.org/Publication.html

Rainfall Mapping using Ordinary Kriging Technique: Case Study: Tunisia

Sayan Mukhopadhaya

Department of Geo-Informatics Indian Institute of Remote Sensing, ISRO, Dehradun
E-mail: sayanmukhopadhaya@gmail.com

Abstract—*This paper presents how spatial interpolation techniques can be used to predict the rainfall: using simple kriging with different variogram fitting models. The technique have been illustrated using annual rainfall observation measured at 75 climatic stations in a 165759.24sq.km region of Tunisia. Cross validation has been used to compare the prediction performances of the geostatistical interpolation methods with the kriging method. The geo-statistical method of ordinary kriging is quite useful for understanding the trend and can also be used for prediction.*

The use of ordinary kriging is done instead of other forms of kriging like regression kriging or co-kriging, because, for regression kriging we need another variable to have a regression analysis, but here the dataset only provides the rainfall variable. This work also shows that on getting a simple data-set, we can also provide good results, if our variogram modelling is correct.

The results obtained for prediction errors has been low and as desired. The RMSE (Root Mean Square Error) for 75 samples has been 3.443, while the mean error is 0.038. The geostatistical algorithm results in a much better observation and prediction of rainfall in other regions where samples has not been collected. Methods to remove other difficulties like skewness in data, has also been discussed in the paper. The dataset has been divided into 38 and 37 samples to get a better cross-validation of the variogram model. The RMSE and mean error obtained for the samples were much lower than 75 samples. The analysis of the data, variogram model fitting, and generation of prediction map through ordinary kriging has been accomplished by coding in R software.

1. INTRODUCTION

The total amount of rainfall in an area gives an estimate about crops/vegetation to be planted. It assists agricultural and environmental agencies in planning and monitoring the agricultural production as well as predicting the yields. Rainfall is crucial in ensuring food security, and therefore, rainfall mapping is a very essential for any given area. Spatial interpolation techniques can be used to predict the rainfall of un-sampled locations. These techniques are important when ground data has not been collected all over the region and these techniques can be used on the already available data to predict the amount of rainfall for areas where ground observations were not made. The statistical interpolation techniques along with Geographic Information System forms a robust method for predicting such un-sampled locations. "Based on Tobler's Law of Geography, which stipulates that observations close together in space are more likely to be similar than those further apart, the development of models attempting to represent the way close observations are related can sometime be very problematic. The approaches can be divergent and may therefore lead to very different results." (Dubois, 1997). In this research, ordinary kriging is used to interpolate rainfall in un-sampled locations and also try to Fig. out the best suited variogram model for the given data-set.

2. OBJECTIVE

The following are the objectives of the work:

a) Use of ordinary kriging to interpolate rainfall in un-sampled locations

b) To show the rainfall trend in the region

c) To show the prediction uncertainty of spatial interpolation techniques in rainfall mapping

3. DATA

The data used was annual rainfall of Tunisia. Tunisia is the northernmost African nation, with an estimated land area of 165759.24sq.km. It shares border with Algeria in the west, Libya in the southeast direction and the Mediterranean Sea in the north and east. The following Fig. (Fig. 1.) shows the location of Tunisia with respect to the world, and observed rainfall locations are also shown in Fig. 2.

Data set has total 75 observed locations. The original data is skewed. So, a square-root transformation is done on the data to remove skewness (shown in Fig. s 3a, 3b, 4a and 4b.) in the data. After the data-set is transformed, we find that, the data-set is tri-modal, but, due to lack of much observed points, we ignore the third-mode, and consider the data as bi-modal. The total observed points in the data-set is 75. As, the data-set is bi-modal (considered), we divide the data-set into two parts of 38 observed points and 37 observed points.

Fig. 1: (Tunisia's Location Map)

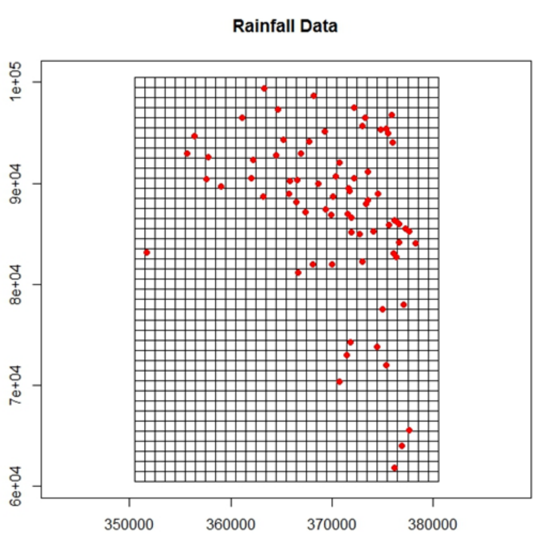

Fig. 2: (Rainfall Map of Tunisia)

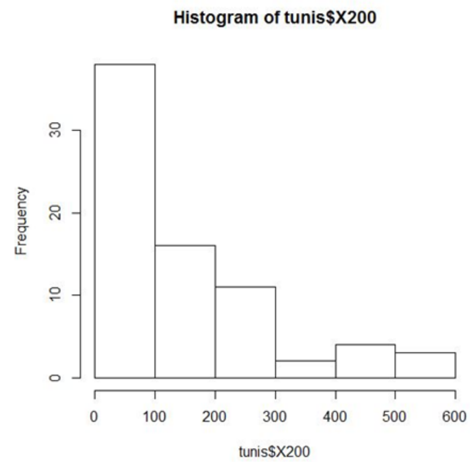

Fig. 3a: (Skewed Data)

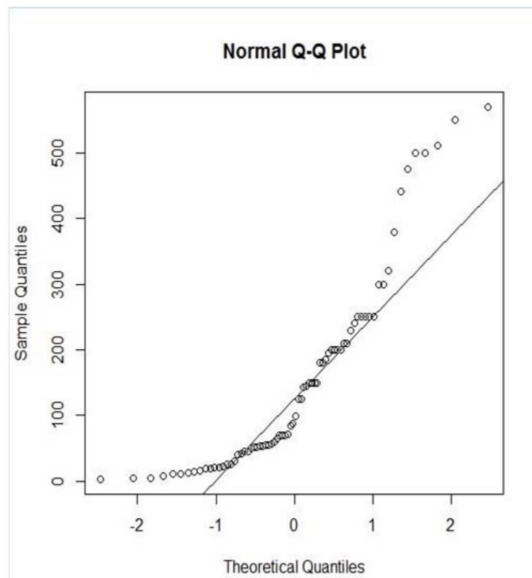

Fig. 3b: (Skewed Data)

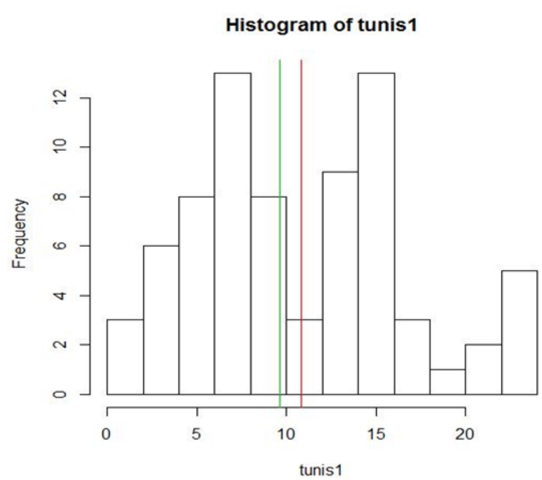

Fig. 4 a.: Histogram of Transformed Data

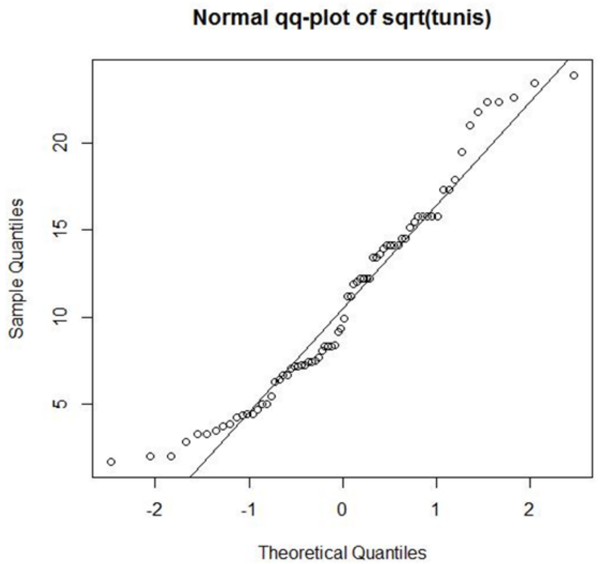

Fig. 4b: Transformed (Square-Root) Data

The data set used (Fig. 5) has a. minimum value of 3.0 and maximum value of 570.0, the median and mean are 93.0 and 151.3 respectively. After the square-root transformation (Fig. 6.), the minimum value, maximum value, median and mean are 1.732, 23.870, 9.639 and 10.820 respectively. The data has been divided into two parts, one with 38 observations and the other with 37 observations so as to obtain two datasets for mapping and validation.

```
summary(tunis$X200)
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  3.0    42.0    93.0   151.3   210.0   570.0
```

Fig. 5: Shows summary of original data

```
summary(tunis1)
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
1.732   6.479   9.639  10.820  14.490  23.870
tunis1= sqrt(tunis$X200)
```

Fig. 6: Shows summary of transformed data

```
> tunis
        X     Y Rainfall SQRT_Rainfall
1  364480 92745      200     14.142136
2  371700 89250      145     12.041595
3  377645 85260       18      4.242641
4  368200 98700       12      3.464102
5  365220 94318      150     12.247449
6  369400 87400      512     22.627417
7  368645 90011      180     13.416408
8  368086 82000      210     14.491377
9  375000 77500      250     15.811388
10 362200 92300      250     15.811388
```

Fig. 7: Shows the Structure of the Data

4. METHODS

a) Variogram

The variogram models were fitted and the variogram model that best fitted the data was selected based on the nugget value (Van Groenigen, 2000). A small nugget and a small SSErr (Sum Squared Error) value means lower uncertainties in the data. The main reason is that there may be insufficient data for deducing accurate nugget values, so the nugget value is presumed to be zero (Western & Blöschl, 1999).

Based on the best fit variogram parameters, ordinary kriging was used on the data-sets.

b) Kriging

Ordinary kriging of a single variable is the most robust and common geo-statistical method and used mostly to predict the variables at unknown location (Webster & Oliver, 2007). It is based on the assumption of unknown mean (Webster & Oliver, 2007) and is said to be unbiased. For this research rainfall amounts at sampled locations were used to predict the rainfall of un-sampled locations as discussed by Webster & Oliver (2007). Kriged prediction/estimation and associated kriging variance were obtained. RMSE (Root Mean Square Error) and ME (Mean Error) were calculated to obtain the uncertainty measure.

5. RESULTS AND ANALYSIS

The exponential model fits the data-set with 75 observation, 38 observations and 37 observations with a cutoff = 15000, but varying bin width of 1500, 2000 and 2500 respectively. The varying bin width is due to the number of points in an observation, as we change the bin-width the number of points per observation reaches below 30, which is undesirable. So, the bin-width is different for each dataset. The exponential model (Fig. s 8, 9, 10) is the best fit for all the variogram for the given data-sets. We get nugget = 0, which follows the theory as discussed by Bohling (2005), for all the datasets, which shows there is high correlation between the observed points. The total sill for the datasets is shown in the below table (Table 1.).

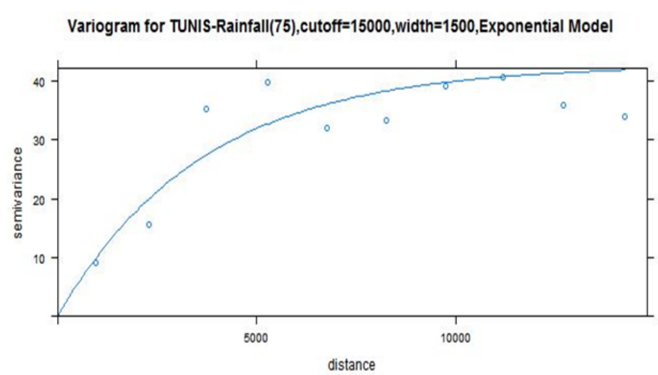

Fig. 8: Sample Variogram

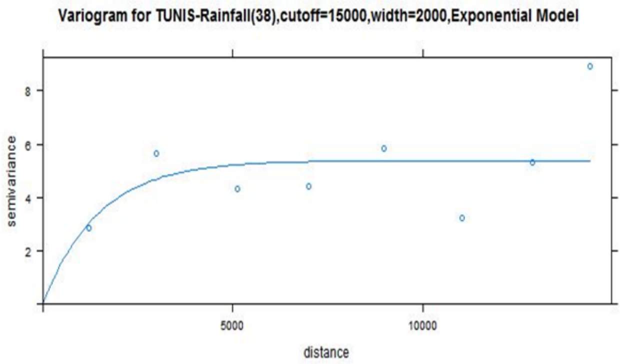

Fig. 9: Variogram of 38 Sample points

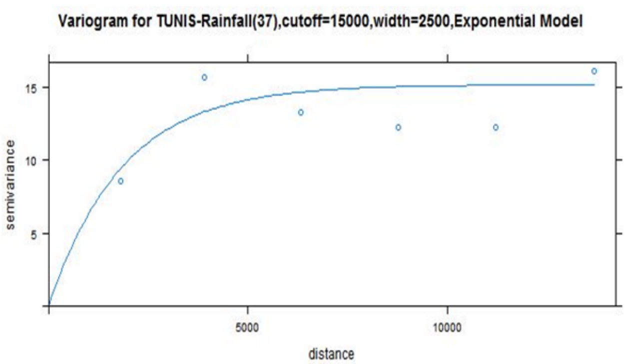

Fig. 10: Variogram of 37 Sample points

Table 1: Values of Total Sill for the data-sets

Dataset with	Total Sill
75 observation points	42.70
38 observation points	5.39
37 observation points	15.20

The sum squared error of the all the three datasets is quite close to zero, this shows that exponential model is the best-fit for the datasets as shown in the variograms below.

Prediction Uncertainty determination

The mean error and root mean square error values (Table 2.) also show that the exponential model is the best-fit for all the datasets.

Table 2: Values of Mean Error and RMSE for the data-sets

Dataset with	Mean Error	Root Mean Square Error
75 observation points	0.038	3.443
38 observation points	0.152	1.756
37 observation points	-0.16	2.631

For, validating the variogram models of both 38 samples and 37 samples, with the sample variogram model, a cross validation was done. These cross validations (Fig. s 11 and 12) resulted in mean error very close to zero and root mean square error quite similar to that of the modelled variogram, which shows that the model fitting to the sample variogram can fit both the smaller datasets of 38 samples and 37 samples, and there is a high correlation between all the sampled data.

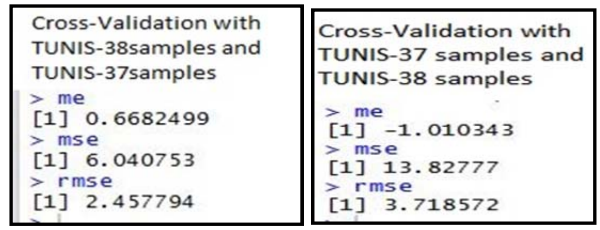

Fig. 11 & 12. .Cross-validation Results

me = mean error; **mse** = mean square error;

rmse = root mean square error.

The predicted value of an un-sampled point such as at (370000, 80000) is 120.98 with a variance of 25.73 and at (355000, 755000) is 176.15 with a variance of 45.12, shows that the sample variogram model can be used to predict the un-sampled points. The following Fig. s (13a and 13b) shows the same.

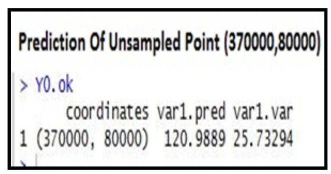

Fig. 13a: Predicted value at (370000, 80000)

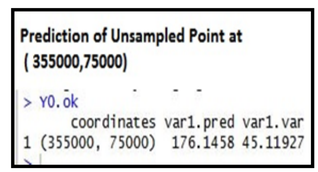

Fig. 13b: Predicted value at (355000, 75000)

Rainfall interpolation

In the rainfall prediction map (Fig. 14), the areas shown in yellow are the areas where higher amount of rainfall is predicted, and the areas shown in blue are the areas with low amount of rainfall.

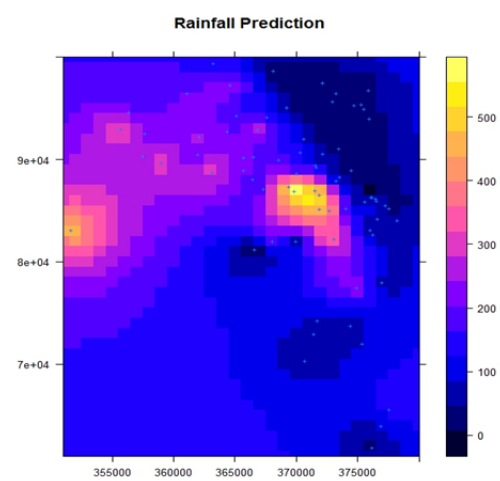

Fig. 14: Rainfall Prediction Map

In the rainfall prediction error map (Fig. 15), the areas shown in yellow are the areas where high error in the prediction of rainfall, and the areas shown in blue are the areas with low error in the prediction of rainfall.

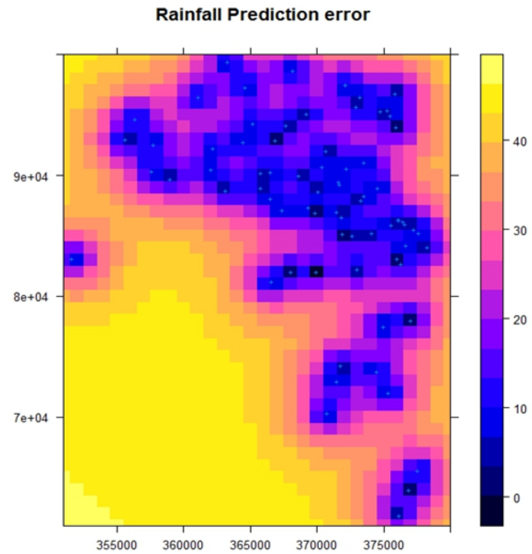

Fig. 15: Rainfall Prediction Error Map

The mapping results shows that the locations that are far from the observed locations have low prediction accuracy, but a high prediction error.

The directional variogram (Fig. 16) shows the distribution of the sampled observations with proper angular direction (0, 45, 90, 135 degree angles).

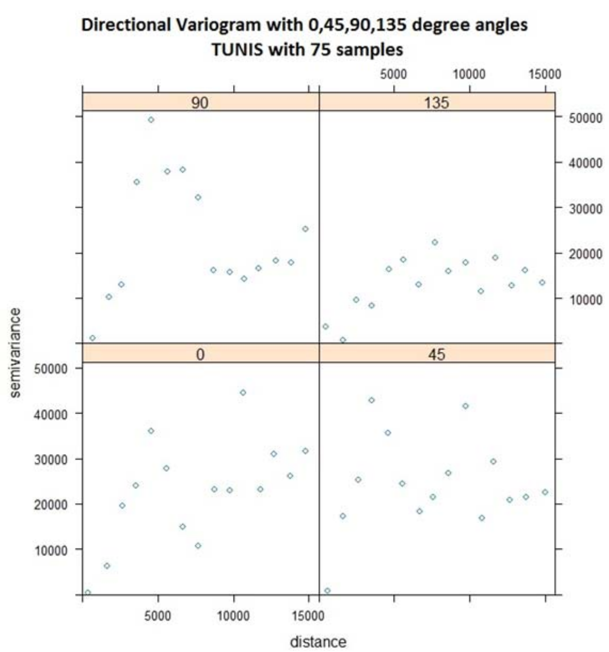

Fig. 16: Directional Variogram Diagram

6. DISCUSSION

In this work, for the given dataset an exponential model variogram can be used to predict the further un-sampled locations. The sample variogram can be used to predict the un-sampled locations that are nearer to the sampled locations as, we know, the far the location from the sampled locations, the higher is the prediction error, due to low correlation between the points. The observations (Fig. 17) show that the rainfall is highly distributed on the north-eastern part of Tunisia, and there is a scarcity of rainfall in the western part, though there are some exceptional locations with high rainfall in the western part of Tunisia. This can be due to the presence of sea in the north-eastern part of Tunisia. The above trend in the rainfall can be used to predict the rainfall in the rest of Tunisia, where there is no sampled locations. The geo-statistical method of ordinary kriging is quite useful for understanding the trend and can also be used for prediction.

The use of ordinary kriging is done instead of other forms of kriging like regression kriging or co-kriging, because, for regression kriging we need another variable to have a regression analysis, but here the dataset only provides the rainfall variable. The co-kriging also requires another variable that may affect the rainfall, which was absent, so further analysis was not possible.

7. CONCLUSION

- Exponential variogram is the best fit for the data as it has the lowest SSErr and a nugget of zero.
- Mean error is close to zero and the RMSE is low as observed in validation and cross validation using ordinary kriging.
- The observations show that the rainfall is highly distributed on the north-eastern part of Tunisia, and there is a scarcity of rainfall in the western part.
- The predicted values obtained from ordinary kriging can be used to locate the areas where the rainfall will be high.

REFERENCES

[1] Bohling, G. (2005). INTRODUCTION TO GEOSTATISTICS And VARIOGRAM ANALYSIS. Earth, (October), 1–20. http://doi.org/10.1162/0162287054769931

[2] Dubois, G. (1997). Spatial Interpolation Comparison 97. Statistics, 97(Sic 97), 1–6.

[3] Van Groenigen, J. W. (2000). The influence of variogram parameters on optimal sampling schemes for mapping by kriging. Geoderma, 97(3-4), 223–236. http://doi.org/10.1016/S0016-7061(00)00040-9

[4] Webster, R., & Oliver, M. A. (2007). Geostatistics for Environmental Scientists. John Wiley & Sons, Ltd (Second, Vol. 1). http://doi.org/10.2136/vzj2002.0321

[5] Western, A. W., & Blöschl, G. (1999). On the spatial scaling of soil moisture. Journal of Hydrology, 217(3-4), 203–224. http://doi.org/10.1016/S0022-1694(98)00232-7