Albahit Journal of Applied Sciences ISSN 2708-244X Vol. 1, Issue 1, 2020, 54-59

Evaluation of Geostatistical Interpolation Methods for Rainfall data Estimation in Libya

Lubna S. Ben Taher
Lecturer, Civil Engineering Department, University of Benghazi, Libya.
lubna.bentaher@uob.edu.ly

Abstract- **Rainfall is considered a highly valuable water resource, particularly in arid and semiarid regions like Libya. Moreover, rainfall data has been shown to be fundamental input for accurate distributed hydrologic modeling. Hydrological Modeling using geostatistical interpolation technique is performed to study hydrological process which is essential for water resource management. This study aim to evaluate six geostatistical interpolation techniques such as Empirical Bayesian Kriging (EBK), Radial Basic Function (RBF), Inverse Distance Weighting (IDW), Global Polynomial Interpolation (GPI), Kernel Interpolation With Barriers (KIB), and Diffusion Interpolation With Barriers (DIW), at that point compare their performance in generating spatial distributions of monthly total rainfall data from 63 meteorological stations located across Libya over 40 years period. Different statistical accuracy measurements such as mean absolute percentage error MAPE%, efficiency factor E, and 95% confidence interval CI95% were used to determine the best geostatistical interpolation methods. Results demonstrate that, RBF, and IDW geostatistical interpolation methods, are both better and more reliable than other geostatistical interpolation methods used in this study. Moreover, they provide a maximum, minimum and average predicted values within the range of 95 CI%.**

Keywords- Evaluation, Geostatistical, Interpolation, Techniques, GIS, Libya, Total Monthly Rainfall Distribution.

I. INTRODUCTION

Libya has considerable water related problems such as water drought. So the need to manage, water resources in Libya is how to obtain useful information for decision-making and planning. Thus, it is very important to estimate areal rainfall from points monitoring stations. There are many methods of spatial interpolation regardless of their effectiveness and simplicity of usage variables. Therefore, it can be a challenging task to determine which method produces the results closest to the actual one. Moreover, criteria must be found to determine whether the method chosen is appropriate for a point data set. Geographic information systems GIS are supposed to be powerful tools in the spatial application of interpolation techniques. The geostatistical interpolation is a tool in GIS used to find the values of unknown points. It can be defined as a procedure of estimating the values of properties at un-sampled locations based on the set of observed values at known locations. Numerous interpolation methods have been developed for use with point, line, and area data. No matter which interpolation technique is used, the derived values are only estimates of what the real values should be at a particular location. The quality of any analysis that relies on interpolation of observed data are, therefore, subject to a degree of uncertainty [1].

Many researchers have evaluated various methods for interpolation of rainfall data such as:

Mebrhatu, 2006, in his study, he evaluated different spatial interpolation methods to form a continuous surface based on measured annual rainfall totals. Interpolation methods examined in his study include inverse distance weighted (IDW), Spline, and Kriging. The annual rainfall totals measurements were taken in his study at 22 discrete locations in the highlands of Eritrea. Judging from the statistical his results, the standard implementation of Kriging provides the smallest prediction errors than IDW and Spline. IDW prediction errors were almost similar to Kriging while Spline produced the largest prediction errors [2].

Keblouti, et al, 2012, the spatial interpolation of annual rainfall in 29 years was studied by them using Inverse Distance Weighting, spline, and Ordinary kriging methods. Their study results given that, ordinary kriging and, spline are the less efficient interpolation methods. Further, Inverse Distance Weighting method is the most representative method for characterize rainfall distribution in the Annaba city- Algeria [3].

Firdaus, and Talib, 2015, their study aims to find an optimal interpolation scheme for rainfall data in Selangor and Langat basins in Selangor, Malaysia. Five interpolation methods had been tested by them after exploring data and cross-validation was used as the criterion to evaluate the accuracy of the various methods. According to their results the best method was the kriging method whereas the inverse distance weighting (IDW) perform worst [4].

Babu, 2016, his study aim to understanding the spatial and temporal pattern of rainfall in the Dakshina Kannada district in Karnataka. His study results shown that, Krigging is the best suitable method for the study area [1].

Pandey, et al, 2016, in their study, the spatial interpolation techniques were performed on rainfall data to predict the unspecified values in Bisalpur Catchment-India. Their study results found that, the geostatistical interpolation methods (kriging and, Topo to Raster) gave better results than other mathematical models because they are based on the spatial variability of data [5].

İcaga, and, Tas, 2018, their study objective, compared the interpolation methods to model the spatial distribution of monthly precipitation values in Akarcay Sinanpasa and Suhut sub-basins, Turkey. In their study the inverse distance weighting (IDW), Simple Kriging (SK) and, Co-Kriging (CK) were compared with each other by cross validation technique. According to their results, SK and, CK are clearly showed better performance than IDW method for the application period in the study area with close and less error values [6].

The main objective of this study is to evaluate six geostatistical interpolation techniques such as Empirical Bayesian Kriging (EBK), Radial Basic Function (RBF), Inverse Distance Weighting (IDW), Global Polynomial Interpolation (GPI), Kernel Interpolation With Barriers (KIB), Diffusion Interpolation With Barriers (DIWB), and to compare their performance in generating spatial distributions of monthly total rainfall from 63 monitoring stations located across Libya. The evaluation carryout by compared the coefficient of efficiency (E), mean absolute percentage error (MAPE %), and ninety fifth confidence interval ±CI95% for the six geostatistical interpolation methods. Fig.1, a flow chart showing the study structure.

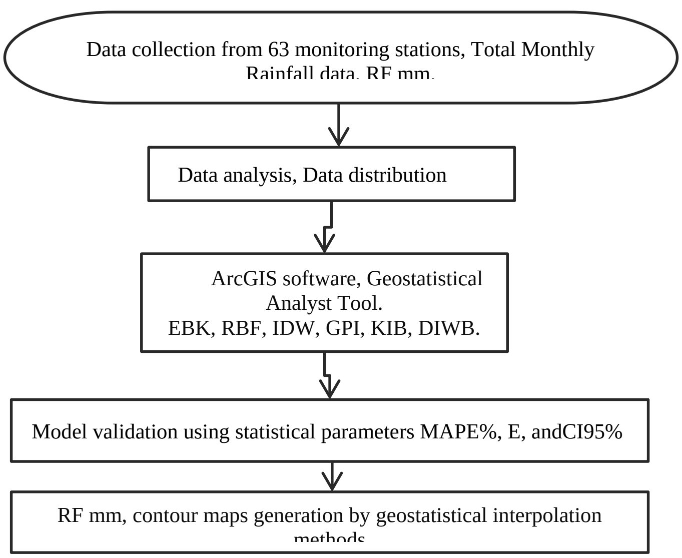

Figure 1.Study methodology process.

II. MATERIALS AND METHODS

1. Area of Study

Libya be located in northern Africa from 20° to 34° N and 10° to 25° E. Total area of Libya is about 1.76 Mkm^2, which roughly 90.8% of the area is hyper-arid, 7.4% arid, 1.5% semi-arid and 0.3% is classified as sub-humid. Topography is generally free of steep terrain, with the exception of two regions in the north-west and north-east where the elevation ranges from 500 to 1000 m above mean sea level. The total population in Libya was about 6.133 million by a growth rate of 1.6 %, nearly 95 % of the population lives in the coastal region in the north, and the rest in widely scattered oases in mid- and south. The mean annual temperature in the coastal region of Libya ranges from 14.2 °C to 21.0 °C and in the interior region between 21.3 °C to 23.4 °C. Libya is one of the driest countries in the world with mean annual rainfall along the coast ranging between 140 to 550 mm and rarely exceeds 50 mm in the interior region. Libya is heavily dependent on the groundwater from five basins in the interior arid zones, with groundwater resources supplying about 80% of the total water consumption, with 97% of the consumption used for agricultural purposes [7].

2. Data Analysis:

The total rainfall data are available at 63 stations in Libya over the 40 years-period 1970-2010 collected from the Libyan National Meteorological Center. A labels of monitoring stations, and their coordinates are presented in Fig. 2 and Fig. 3. The variable selected for this study was Total Monthly Rainfall (RF mm) and shown to be reliable therefore to be able to distinguish the efficiency of geostatistical interpolation techniques, moreover, enormous of rainfall data give better results than small amounts. The statistical analysis of monthly rainfall data presents at Table I. As shown in Table I, the Libyan rainy season occurs in the months (October, November, December, January, February, and March) with 99% of the total annual rainfall, with the average total monthly rainfall (RF mm) is 194.1mm. As shown in Fig.2, the Shahat station has the highest total monthly rainfall in the eastern coastal region. Whereas Gharyan station, which has the highest total monthly rainfall in the western coastal region. Moreover, total monthly rainfall at inland stations ranges between 0 to 80mm (Garyiat station to Kufrah station).

TABLE I. The Statistical Information f RF mm at 63 Monitoring Stations Used.

Months	**Mean RFmm**	**Minimum RFmm**	**Maximum RFmm**
JAN.	38.67	0.2	133.7
FEB.	25.28	0	90.7
MAR.	21.69	0.05	65.7
APR.	9.774	0	37
MAY	3.36	0	13.8
JUN.	0.582	0	3.3
JUL.	0.0654	0	0.9
AUG.	0.1386	0	1.1
SEP.	7.448	0	21
OCT.	24.87	0.03	60.4
NOV.	25.72	0	71.6
DEC.	36.48	0	125.4
Total Monthly Rainfall(RF mm)	194.1	1.5	583.3

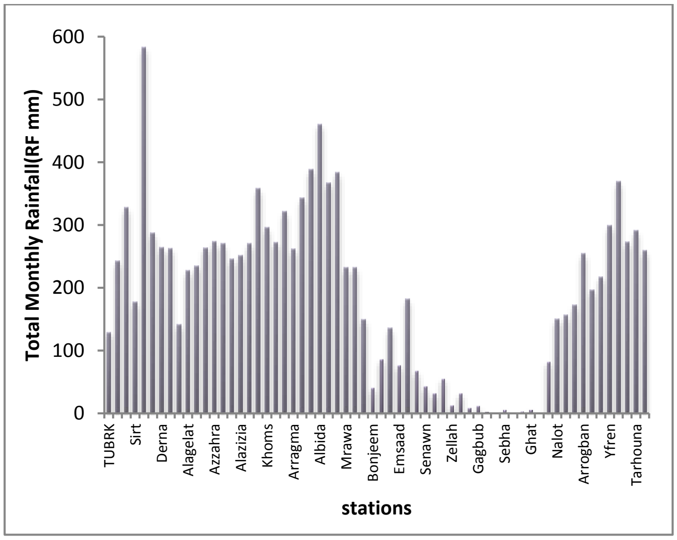

Figure 2. Total monthly rainfall recorded at used stations from Jan. to Dec. of 1970-2010.

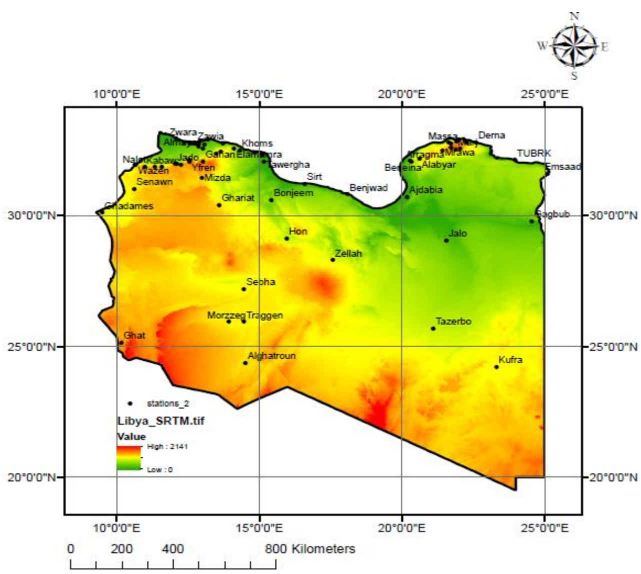

Figure 3. Monitoring stations locations and topography of the study area, Libya.

3. *Geostatistical Interpolation Methods*

The Geostatistical interpolation methods used in the current study were performed by ESRI ArcGIS V.10.1, and the following section briefly introduces the geostatistical interpolation techniques used in this study [8]:

a. *Empirical Bayesian Kriging Interpolation model (EBK):*

Kriging is a prediction method that considers values of a variable in ensamples points as a linear composition of the values of surrounding points. Considering the values of variable **Z** in **n** measured points as following:

$$Z = \left(Z(x_1), Z(x_2), \ldots . Z(x_n)\right) \quad (1)$$

Estimation of Z in point X_0 using Kriging estimation is defined as:

$$Z\ (x_0\ ) = \sum \lambda_i . Z(x_i) \quad (2)$$

Where**:**

$Z\ (x_0) =$ interpolated value for grid node, $Z(x_i) =$ the measured points, λ_i = Kriging statistical weight.

The most important part of Kriging is statistical weighs assigned to λ_i. To avoid bias of estimation, the weights should be determined in a way that summation is equal to one. Further this method can be used to produce an accurate grid of data, or can be custom-fit to a data set by specifying the appropriate variogram model $\gamma(h)$. The experimental variogram measures the average degree of dissimilarity between unknown values and a nearby data value and thus can depict autocorrelation at various distances. The value of the experimental variogram for a separation distance of h (referred to as the lag) is:

$$\gamma(h) = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} [z(x_i) - z(x_i + h)]^2 \quad (3)$$

Where:

n(h) = the number of data pairs within a given class of distance and direction.

b. *Inverse Distance Weighted (IDW) Interpolation Model:*

In interpolation with IDW method, a weight is attributed to the point to be measured. The amount of this weight is depended to the distance of the point to another unknown point. The general formula of inverse distance weighted as follows:

$$Z_j = \frac{\sum_{i=1}^{n} \frac{z_i}{h_{iy}^{\beta}}}{\sum_{i=1}^{n} \frac{1}{h_{iy}^{\beta}}} \quad (4)$$

$$h_y = \sqrt{d_y^2 + \delta^2} \quad (5)$$

Where**:**

h_{iy} = effective separation distance between grid node j and the neighboring point i, $z_{j=}$ Interpolated value for grid node j, z_i =neighboring knowing points, d_y = distance between grid node j and the neighboring point i, $\beta =$ weighting power parameter, $\delta =$ smoothing parameter.

The weighting power parameter determines how quickly weights fall off with distance from the grid node. As the power parameter increases, the generated surface is a "nearest neighbor" interpolator and the resultant surface becomes polygonal. The polygons represent the nearest observation to the interpolated grid node. The smoothing factor parameter is allowed to incorporate an "uncertainty" factor associated with the input data. The larger the smoothing factor parameter, the less overwhelming influence any particular observation has in computing a neighboring grid node.

c. Global polynomial (GPI) Interpolation Model:

Global polynomial interpolation fits a smooth surface that is defined by polynomial mathematical function to the input sample points. The global polynomial surface changes gradually and captures coarse surface-scale pattern in the data. Global polynomial interpolation simply uses multiple regression method for all of the data. A response surface is fitted to the x- and y- coordinates, for an example the third - order trend model is:

$$Z(x_i, y_i) = \beta^0 + \beta^1 x_i + \beta^2 y_i + \beta^3 x_i^2 + \beta^4 y_i^2 + \beta^5 x_i y_i + \beta^6 x_i^3 + \beta^7 y_i^3 + \beta^8 x_i^2 y_i + \beta^9 x_i y_i^2 + \varepsilon(x_i, y_i) \quad (6)$$

Fitting regression models by estimating parameter (β_i) uses ordinary least squares.

Where**:**
$Z(x_i, y_i)$=datum at the location(x_i, y_i), β_j =constant parameter, and $\varepsilon(x_i, y_i)$= random error.

d. Radial Basis Functions (RBF) Interpolation Model:

Radial basis functions (RBF) methods are a series of exact interpolation techniques, that is, the surface must go through each measured sample value. There are five different basis functions: thin-plate splint, splint with tension, completely regularized splint, multiquadric function, and inverse multiquadric splint. Each basis function has different shape and results in different interpolation surface. RBF methods are a form artificial neural network.

e. Kernel Interpolation With Barrier (KIWB) Interpolation Model:

The Kernel Interpolation model uses the shortest distance between points so that points on the sides of the specified nontransparent (absolute) barrier are connected by a series of straight lines. Kernel Interpolation uses the following radically symmetric kernels: Exponential, Gaussian, Quartic, Epanechnikov, Polynomial of order 5, and constant. The bandwidth of the kernel is determined by a rectangle around the observations.

F. Diffusion Interpolation With Barriers (DIWB) Interpolation Model:

Diffusion Interpolation refers to the fundamental solution of the heat equation, which describes how heat or particles diffuse with time in a homogeneous medium. The predictions made using this method gently flow around barriers. The Diffusion Interpolation can use a complex distance metric defined by the cost surface which is a common raster function that calculates the cost of travel from one cell of a raster to the next.

III. Models Performance Verification

Different geostatistical interpolation methods will yield different results, and therefore it is challenging to decide which method is better by just viewing the corresponding contour lines. According to that, the six geostatistical interpolation methods performances used in this study were compared using different types of measures so that errors can be calculated. This evaluation calculates error statistics with the recorded rainfall as the measured data and the interpolated as the predicted values, which are given by the following relations [8]:

A. Mean Absolute Percentage Error (MAPE %):

$$MAPE\% = \frac{100}{n} \sum_{i=1}^{n} \left| \frac{RFobs._i - RFpre._i}{RFobs._i} \right| \quad (7)$$

Where:

n= Number of data.

RF.obs.= Observed value.

RF.pre.= Predicted value.

B. Efficiency Factor (E):

Efficiency factor (E = 0 to 1) is calculated on the relationship between the predicted and observed mean deviations and it can show the correlation between the predicted and observed data:

$$E = 1 - \frac{\sum_{i=1}^{n} (RFobs._i - RFpre._i)^2}{\sum_{i=1}^{n} (RFobs._i - \overline{RFobs._i})^2} \quad (8)$$

Where:

n= Number of data.

RF.obs.= Observed value.

RF.pre.= Predicted value.

$\overline{RF.\text{obs.}}$ = The average of observed data.

C. 95% Confidence Limits (±95CI%):

The 95% confidence intervals are used to reflect the reliability of a statistical estimate and based on observed data. A 95% confidence intervals are most commonly calculated as:

$$\overline{RFobs.} - 1.95\left(\frac{s}{\sqrt{n-1}}\right) < \overline{RFpre.} < \overline{RFobs.} + 1.95\left(\frac{s}{\sqrt{n-1}}\right) \quad (9)$$

Where:

S = Standard deviation of the observed data.

$\overline{RF.obs.}$ = *The average of observed data.*

$\overline{RF.\text{pre.}}$ = The average of predicted data.

The value on the left side of the inequality yields the lower limit, and on the right side yields the upper limit for the mean observed data, known as the confidence levels.

The geostatistical interpolation methods with superior performance; have zero MAPE% and high estimation of E. Thus, the geostatistical interpolation methods if they have a good performance will produce results within the range of 95CI% of the mean observed data. In a specific order, the geostatistical interpolation methods are utilized to create information which save the primary factual attributes of the observed information.

IV. RESULTS AND DISCUSSION:

Six geostatistical interpolation methods ((EBK), (RBF), (IDW), (GPI), (KIB),and (DIW)) were implemented in ESRI's ArcGIS software using Geostatistical Analyst Tool. Each method uses a different approach for determining the output cell values.

The different models are compared with each other through a validation procedure to the most appropriate method which gives greater accuracy in the validation set. Predicted total monthly rainfall obtained were further validated using MAPE%, E and ±CI95% as shown in Table II, and the scatter plots of the true values versus the estimated values are illustrated in Fig. 4. The highest values for E =1and lowest MAPE% =0 respectively with RBF and IDW geostatistical interpolation methods, this indicates that RBF and IDW methods are both better and more reliable than other interpolation methods used in this study. Moreover, they provide a maximum, minimum and average predicted values within the range of 95CI%. While the highest values of MAPE% (280.13, and 230.13) and lowest values of E (0.45, and 0.58) respectively is appear clearly with the application of DIWB and GPI methods. Furthermore, comparing the estimated values to the observed of the aquifer total monthly rainfall was done via a scatter plot, see Fig.4, more the scatterplot is tightened around the predicted line, the better the estimated values by (RBF, and IDW) models.

The thematic maps generated by various methods are shown in Fig. 5 which depicts a continuous total monthly rainfall surface. Red color shows high total monthly rainfall whereas low total monthly rainfall is illustrated by shades of blue color. Clearly as showing at the thematic maps (Fig.5), the total monthly rainfall decreases gradually towards the center of the Libyan coastline and decreases in the south-east of Libya. The range of total monthly rainfall data at the eastern and western stations is (130mm to 583mm), and at the central stations is (35mm to 130mm), with the lowest total monthly rainfall in the Sahara region (0 to 35mm). The highest total monthly rainfall was 583 mm (Shahat station at the eastern) followed by 370mm (Gharyan station at the western), with the lowest precipitation in the coastal region recorded at Nalute station (150mm) followed by Ajdabyia station (142mm). The total monthly rainfall in the northern Sahara region does not exceed 70mm at Mizda, with total monthly rainfall across the southern Sahara region 20mm. However, low variability in total monthly rainfall is identified at stations located at the greatest elevations e.g. (Garyan, Zawia, Khoms, Al-Bayda , Marj, and Shahat), with these stations also receiving the highest maximum total monthly rainfall.

TABLE II.Accuracy Measurements Among The Geostatistical Interpolation Methods.

	Mean RF mm	**Min. RF mm**	**Max. RF mm**	**MAPE %**	**E**	**±CI95 RF mm**
RF Observed mm	194.10	1.50	583.30	--	--	160.61-227.56
RF DIWB mm	208.25	16.52	276.13	280.13	0.45	
RF EBK mm	202.00	2.60	334.70	36.81	0.80	
RF GPI mm	221.18	14.65	286.00	230.13	0.58	
RF IDW mm	194.10	1.50	583.30	0.00	1.00	
RF KIWB mm	194.60	1.60	381.70	25.07	0.80	
RF RBF mm	194.10	1.50	583.30	0.00	1.00	

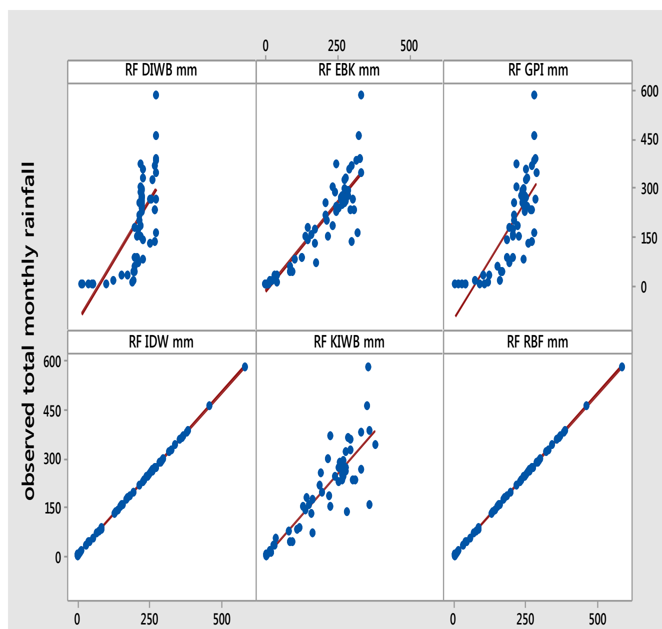

Figure.4. Comparison between the predicted and measured total monthly rainfall data.

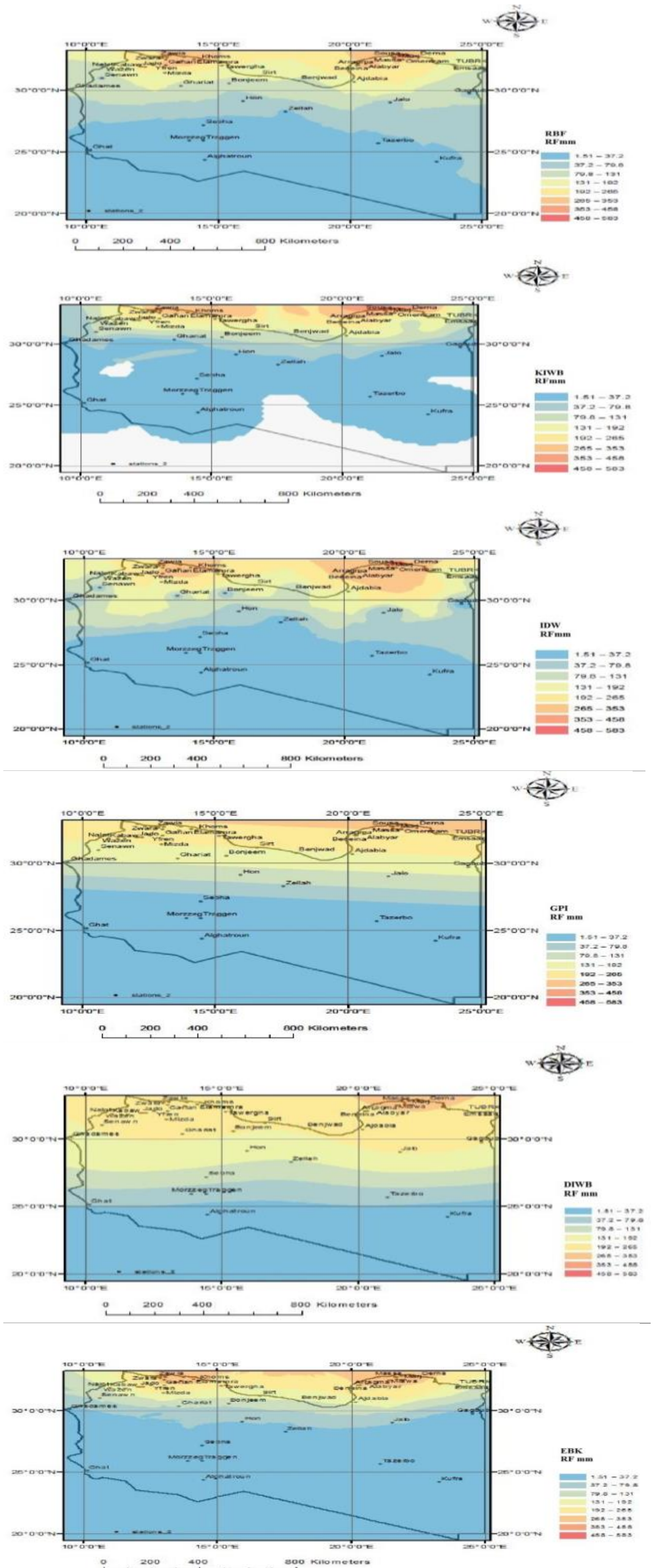

Figure 5. Maps of total monthly rainfall surface in Libya generated by ((EBK), (RBF), (IDW), (GPI), (KIB), and (DIWB)) geostatistical interpolation methods.

V. Conclusion and Recommendations

In this study, the spatial total monthly rainfall in Libya was modeled by six geostatistical interpolation methods((EBK), (RBF), (IDW), (GPI), (KIB), and (DIWB)). The accuracy and precision of these methods were tested by E, MAPE% and CI95% in terms of various performance standards. According to the results, RBF and IDW methods are both better and more reliable than other geostatistical interpolation methods used for the application period in the study area with close and less error values. Since the accuracy of estimated values is varying from each geostatistical interpolation methods based on the topography of the area, concentration and distribution of the measurement stations .

Based on the above results it is suggested that in the future, the climate data resource needs to be regularly updated or improved in detail. It is often needed to construct of new weather stations. Because weather station network using in this study have an irregular spatial distribution, mostly located in populated areas and lower altitudes. Finally obtaining accurate and reliable rainfall maps is an important issue to conduct several studies such as environmental, agricultural, and hydrological.

References

[1] B. S. Babu, "Comparative Study on the Spatial Interpolation Techniques in GIS", International Journal of Scientific & Engineering Research, vol.7, no.2, PP.550-554,February 2016, ISSN 2229-5518.

[2] Y. İçağa1, E. Taş1, "Comparative Analysis of Different Interpolation Methods in Modeling Spatial Distribution of Monthly Precipitation", Journal of Natural Hazards and Environment, vol.4, no.2, PP. 89-104, 2018, DOI: 10.21324.

[3] N. N. Firdaus, S. A. Talib, " Spatial Interpolation of Monthly Precipitation In Selangor, Malaysia – Comparison And Evaluation Of Methods," International Conference on Global Trends in Academic Research GTAR, vol. 1, 2015, PP.346-357.

[4] M. T. Mebrhatu, "Evaluation of Spatial Interpolation Methods for Annual Rainfall on the Highlands of Eritrea", Discov, Innov, vol.18, no.1, PP.15-20, 2006.

[5] M. Kebloutia, L. Ouerdachia , H. Boutaghanea, "Spatial Interpolation of Annual Precipitation in Annaba-Algeria - Comparison and Evaluation of Methods", Energy Procedia ,vol.18 ,PP. 468 – 475, May 2012,DOI: 10.1016.

[6] N. Pandey, K. Panwar, M. Sharma, and M.P. Punia, "Analysis of Spatial Interpolation Techniques for Rainfall Data using Various Methods: A Case Study of Bisalpur Catchment Area", International Journal of Engineering Research & Technology (IJERT), vol.4, no.23, PP. 1:5, 2016, ISSN: 2278-0181.

[7] I. M. AGEENA," Trends and patterns in the climate of Libya (1945-2010)"Ph.D. dissertations, Department of Geography, University of Liverpool,UK,2013.

[8] L. S. Ben Taher." Evaluation of Geostatic Interpolation Methods Based on GIS For Estimation Aquifers Transmissivity (Tazerbo Wellfield – GMMRP, SE Libya) As a Case Study", International Science and Technology Journal, Vol. 17, PP.1-23, April 2019. ISSN: 2519-9838.