Jordan Journal of Social Sciences, Volume **15**, No. **2**, **2022**

Comparison of Spatial Interpolation Methods for Estimating the Annual Rainfall in the Wadi Al-Mujib Basin in Jordan

Yusra Al-husban *

ABSTRACT

Accurate rainfall data are essential for environmental applications in the actual assessment of the geographical distribution of rainfall. Interpolation methods are usually applied to monitor the spatial distribution of the rainfall data. There are many spatial interpolation methods, but none of them can achieve in all cases the best results. In this study, three different interpolation methods were investigated with regard to their suitability for producing a spatial rainfall distribution. Rainfall data from 14 meteorological stations were spatially interpolated using three common interpolation techniques: inverse distance weighting (IDW), ordinary kriging (OK), and kernel smoothing (KS) were compared and assessed against station rainfall data and modeled rainfall. Cross-validation was applied to evaluate the accuracy of interpolation methods in terms of the root-mean-square error (RMSE). The best results were obtained by the lowest RMSE for interpolating the precipitation (RMSE) = 100.86542, while the inverse distance weighting (IDW) performed the worst, and are least efficient with the largest RMSE=103.43; in addition, the kernel smoothing with the least minimum (-) and maximum (+) error is -92.38 mm and 313.33 mm.

Keywords: interpolation, annual rainfall, IDW, kernel smoothing, ordinary kriging, cross- validation, geostatistical method.

1. Introduction

Recent research on interpolation of climatological and meteorological information with the support of GIS has shown that interpolation has a large development potential within climatology and meteorology (Johnston *et al*., 2001; Burrough and McDonnell, 1998). Rainfall data is an important parameter in environmental studies, such as climate modeling and hydrological modeling (Dobesch *et al.*, 2007; Dryas and Ustrnul, 2007). Continuous precipitation data as input is essential for accurate modeling. In addition, meteorological stations are usually sparse, especially in our study area, where meteorological stations are concentrated in less than half of the area of the basin because of a small population and complex terrain in these areas. Therefore, precipitation in no-monitoring areas must be estimated by numerical analysis. Spatial interpolation is a common method to estimate a new date point, which is in the range of a series of known isolated points (Childs, 2012). In this study, three interpolation techniques in a geographical information system (GIS) are analyzed and compared for estimating the spatial distribution of precipitation in the Wadi Al-Mujib **B**asin. Precipitation events for 14 meteorological stations over the study area were investigated. Each method has advantages and disadvantages, and none can perform best in all cases, (Wagner et al., 2012). Geostatistical interpolation methods such as kriging estimate the value of an unknown point utilizing the statistical properties of neighboring known points and taking the spatial autocorrelation among known points into consideration (Sluiter, 2009; Sunyani, 2004). Accurate rainfall data are the most important factor in connection with climate

* Associate professor, Department of Geography, Faculty of Arts, the University of Jordan, Amman, Jordan.

Received on 25/3/2019 and Accepted for Publication on 16/6/2021.

change. Spatial distribution patterns, which are usually spatially interpolated from point observations, have a significant impact on water resources and ecological environment management (Xiaoke, et al, 2016; Anders, et al., 2004). Rainfall interpolation results are influenced by a combination of interpolation methods, and by the number, spatial distribution of meteorological stations and the general topographics (Al-Bakri, et al., 2016). A wide range of interpolation patterns are available for producing spatial rainfall distributions, ranging from such simple schemes as Thiessen polygons or inverse distance weighting to such complex statistical methods as geostatistical ordinary kriging and co-kriging (Sluiter, 2009; Shu et al., 2005). In geostatistical methods, geostatistical multivariate approaches are widely used to predict the rainfall values over a given area by considering secondary information sampled over the same area to enhance the interpolation quality (Xiaoke, et al, 2016; Wagner et al., 2012).

2. Objectives

The main objectives of this study:

- Three different spatial interpolation methods were employed to produce a spatial rainfall distribution by utilizing rainfall data from 14 meteorological stations in the Wadi Al-Mujib Basin.

- Evaluation of the accuracy of various methods was carried out by means of cross-validation.

3. Data and methodology

3.1 Rainfall data

Rainfall has an unevenly spatial and temporal distribution in the study area. Average annual rainfall is below 250 mm in most areas, and gradually decreases from the northwest toward the southeast. The rainfall is mostly concentrated between November and April, and the rainfall has significant annual variation. In this study, observed rainfall data during the period 1980-2016 was used. The 14 meteorological stations in the study area are highly concentrated. Their distribution is shown in Table 1; they are relatively dense in the northwest and very sparse in the southeast (Figure 1). The rainfall data from these 14 rain gauges was used as basic data for evaluating the quality of the rainfall estimates obtained from the interpolation methods. Data were managed as follows:

- 841 precipitation (rainfall) data across the study area was collected.

- Data was divided into separate calibration and validation sets.

-Three maps were created using the geostatistical wizard in ArcMap based on calibration points.

- The root-mean-square error (RMSE) was calculated to quantitatively describe each map’s estimation error when performing spatial interpolation for unmeasured locations.

Table 1. Meteorological stations and their geographic coordinates

Station. N	X	Y	Altitude (m)
1	486311.41	383999.09	692
2	463483.28	385628.06	845
3	508228.47	401358.19	703
4	482807.00	414942.34	750
5	437641.25	375709.72	1247
6	429896.56	390886.28	981
7	496344.13	388163.19	598
8	466463.81	380676.00	857

Station. N	X	Y	Altitude (m)
9	457604.03	409033.28	768
10	461072.84	380088.56	956
11	527657.44	405677.59	860
12	486291.28	397499.41	727
13	493026.47	383108.59	469
14	524798.19	397130.44	782

3.2. Exploring the data

Before calculating the annual mean precipitation over a span of 37 years (1980-2016), data was explored by applying geostatistical analyst tools. The histogram is the most suitable method for checking the Gaussian distribution. If the data conform to a Gaussian distribution, the mean should approach the median, the skewness should be close to zero, and the kurtosis should approximate to a value of three (see Figure 1).

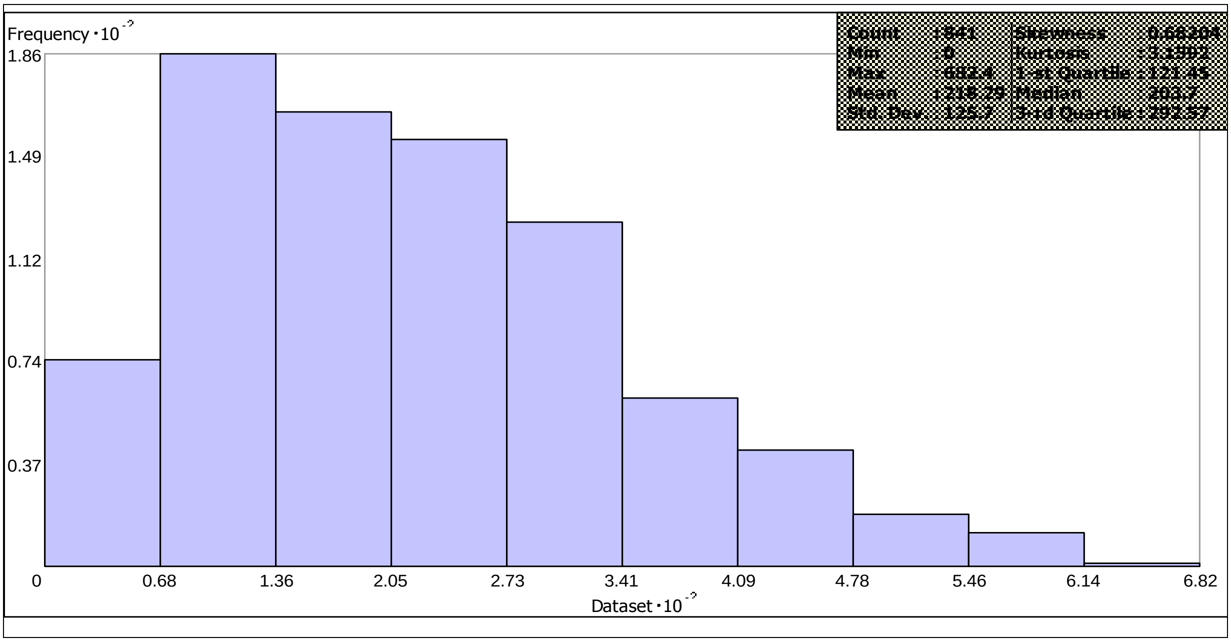

Figure 1: The spatial distribution of annual mean precipitation over 37 years by histogram.

3.3. Interpolation and cross-validation

In this study, three spatial interpolation methods were explored and compared. Cross-validation is usually used to compare the accuracy of interpolation methods. It removes a data point each time from the data sets and predicts values using the other measured points. Then, predicted value at the removed point can be compared with the measured value. In this paper, the interpolation results were validated via cross-validation. To evaluate the goodness-of-fit, the mean absolute error (MAE), the mean relative error (MRE), and the root-mean-square error (RMSE) were calculated (Philips et al., 1992). This procedure was carried out for all the measured points. Interpolation results had to be assessed by statistics that indicate the degree of concordance between the predicted and the actual data observed. Through various statistical calculations, the differences between the predictions obtained by means of the different interpolation methods and the actual data recorded at the rain gauges can be determined. The RMSE and MAE summarize the mean difference in the units of observed and predicted values. The RMSE was used in this study to evaluate the accuracy of interpolation methods, and it can be calculated using Eq. 1.

$$RMSE = \sqrt{(1/N)}\ \Sigma_{(i=1)}^{N} \{z(x_i) - \hat{z}(x_i)\}^{2}$$

where (xi) is the predicted value, z (xi) is the observed value, and N is the number of values.

The interpolation method with smaller values of MAE and MRE is better, and indicators indicate that the predicted value is closer to the observed one (Xiaoke, et al, 2016).

3.3. Interpolation algorithms

3.3.1. Introduction

Many different interpolation methods are available in GIS tools (Geostatistical analyst). These can be employed to interpolate and create a continuous surface by using only the geometric characteristics of point data. In this study, we used IDW as the deterministic method, kriging/co-kriging as a geostatistical method and kernel smoothing as interpolation with barriers. Spatial interpolation is the generation of evaluation values or attributes for unstamped or missing locations within the area covered by existing measurements (Isaaks and Srivastava, 1989).

3.3.2. Inverse distance weighting (IDW)

(IDW) is an advanced nearest-neighbor approach that allows for inclusion of more observations than only the nearest points. The value at a certain grid cell is obtained from a linear combination of the surrounding locations. The weight of each observation is determined by the distance. IDW assumes that every known point has an influence on the predicted point. IDW is widely used in meteorology (Shu et al., 2005; Dobesch, et al., 2007). IDW can be calculated using Eq. 2, (ArcGIS, help; available at http://pro.arcgis.com/en/pro-app/help/analysis/geostatistical-analyst/deterministic-methods-for-spatial-interpolation.htm.

This is calculated employing Eq. 2 as follows:

$$Z* = \frac{\sum_{i=1}^{N}\left(\frac{1}{dP} Z_i\right)}{\sum_{i=1}^{N}\left(\frac{1}{dP}\right)}$$

Where: Z* = estimated value.

Zi = a neighboring data point value

N = the number of the neighboring point

d = the distance between the data point and the point being interpolated

P = a positive-power parameter

3.3.3. Ordinary kriging and co-kriging

Ordinary kriging (OK) is the basic form of kriging. The prediction by ordinary kriging is a linear combination of the measured values. The principle of kriging is to estimate values of a regionalized variable at a selected location (Zk), based on the surrounding existing values (Zi). Selected locations are assigned a relevant weighting coefficient (λi), which represents the influence of particular data on the value of the final estimation at the selected grid node. The condition when assessing the OK technique is that the sum of all weights is equal to 1 (Bohling, 2005). The OK can calculated using Eq. 3.

$$\hat{z}(x_0) = \sum_{i=1}^{n} \lambda_i \,.\, z(x_1)$$

Where: ***Z (xo)*** = the measured value at the *I* the location. λ_i = an unknown weight for the measured value at the *i*th location. s_0 = the prediction location and N = the number of measured values.

3.3.4. Kernel smoothing

Kernel interpolation with barriers is a moving window predictor that uses the shortest distance between points so that points on either side of the line barriers are connected (ArcGIS, help).

4. Study area

Wadi Al-Mujib Basin is a deep canyon in Jordan, which enters the Dead Sea at an elevation of -434 meters (b.m.s.l) and extends to the Kerak and Madaba Mountains to the north and south, reaching 1262 meters (a.m.s.l). The study area is located in the center of the Jordan Rift Valley (JRV), with a total area of 7205.057 km², and extends between 35 º 20′ 0 to 36 º 20 ′ 0″ E and 30 º 20 0″ to 32 º 0 ′ 0″ E .According to the spatial distribution of all 14 meteorological stations in the northern and western parts, the current study area border was determined as where the meteorological stations are available (see Figures 1 and 2). This has an area of ca. 4923.46 km². The topography of the study area is complex, diverse, and heterogeneous. It can be classified into three main topographic zones: the high zone with an elevation > 948 (a.m.s.l); the hilly zone with an elevation ranging between 378-948 (a.m.s.l) and the lowland with an elevation between < 378 (a.m.s.l) - 434 (b.m.s.l), which forms apart of the JRV. The runoff generated by high intensity rainstorm events drains into the Dead Sea. Semi-arid areas such as the study area are characterized by complex rainfall patterns which are significantly influenced by seasonality (Noy-Meir, 1973; Ghanem, 2013). The average annual precipitation is 154 mm and ranges from 300 mm in the northwestern part of the watershed to 50 mm or less in the southeastern corner. The catchments area is sparsely inhabited, with moderate agricultural activity (Zuhair et al., 2013). The study area and distribution of meteorological stations, and the land use/cover are shown in Figures 2 and 3.

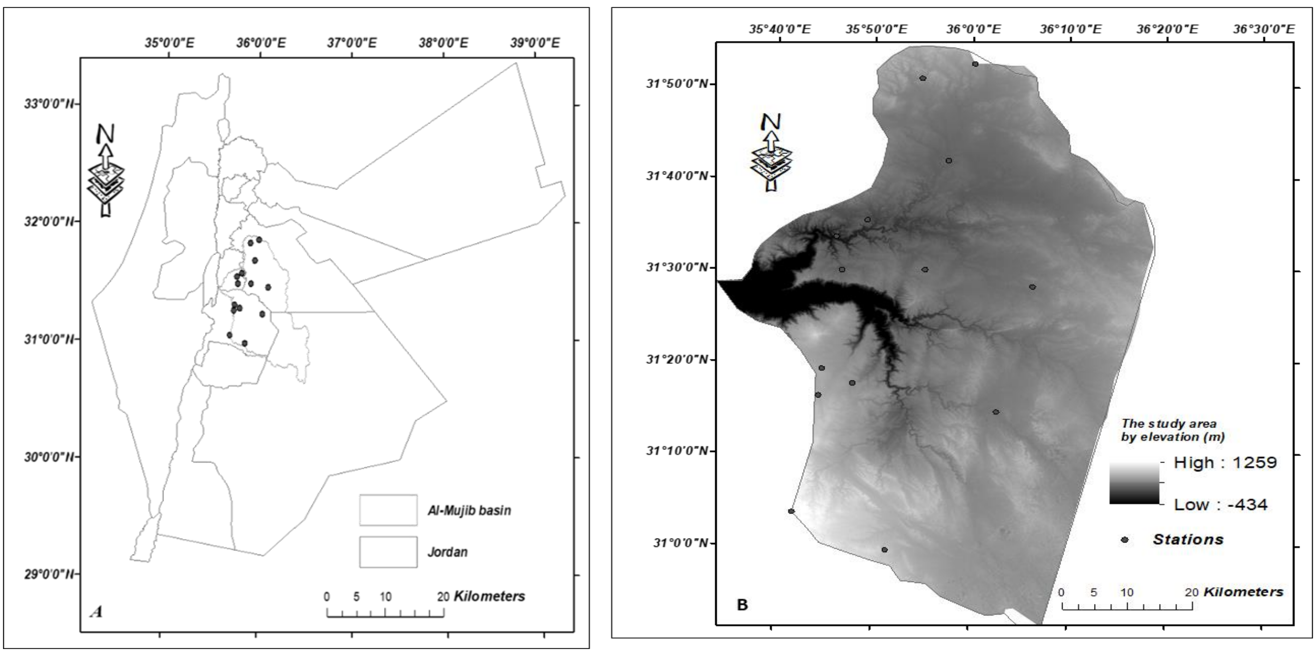

Figure. 2 The location of the study area and the locations of meteorological stations.

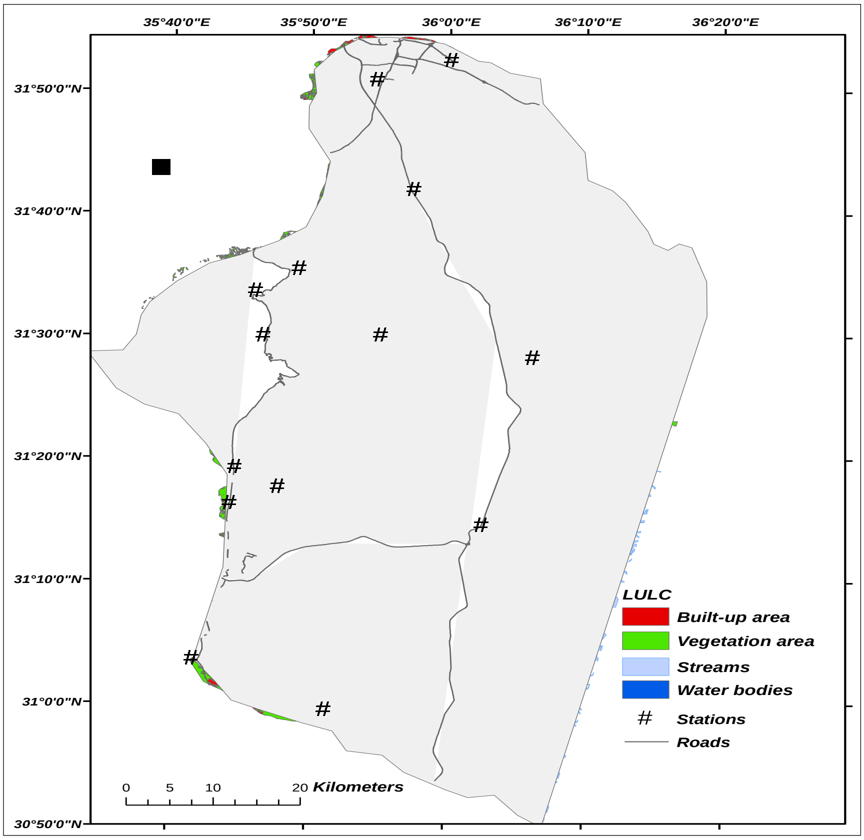

Figure 3. Land use/cover in the study area 2018.

5. Discussion and results

5.1. Discussion

A cross-validation technique was used to evaluate and compare the performance of different interpolation methods. The sample points were arbitrarily divided into two datasets, with one used to train a model and the other used to validate the model. Mean, Root-Mean-Square, Mean Standardized, Root-Mean-Square Standardized, Average Standard Error, and the Root Mean Square Error (RMSE) for error measurement were estimated to evaluate the accuracy of interpolation methods. Data characteristics can be described by statistics, such as mean, median, and skewness. The data characteristic and the distribution after data transformation are shown in Figure 1. Obviously, it is a more approximate Gaussian distribution after transformation. The mean is close to the median in log transformation, and skewness is closer to 0; therefore, log transformation was utilized. IDW, OK and KS were mapped. Cross-validation was carried out for three methods. After a series of test calculations, the mean absolute error (MAE), the mean relative error (MRE), and the root mean square error (RMSE) which are sensitive to this data were chosen as validation criteria for comparing interpolation accuracy. Table 2 shows that kernel smoothing with the minimum of (RMSE) achieved the best results, compared to other methods. KS was found to adjust the results to the topographic variations (Childs, 2004).

Table 2. Cross-validation results

Ordinary Kriging	
Mean	-10.90983813
Root-Mean-Square	109.5139658
Mean Standardized	-0.117898956
Root-Mean-Square Standardized	1.06045281
Average Standard Error	103.2291462
Kernel Smoothing	
Mean	1.710082943
Root-Mean-Square	101.954642
Mean Standardized	0.01676961
Root-Mean-Square Standardized	0.999573831
Average Standard Error	102.0532588
IDW	
Mean	-0.706175188
Root-Mean-Square	103.6049219

The rainfall interpolation maps of the study area showed different spatial distribution patterns depending on which of the different interpolation methods was used (Figure 5). A clear rain gradient extending from the northwest to southeast exists in all the maps, with high rainfall in the northeast of more than 300mm/ year, and low rainfall in the southwest of less than 100 mm/year of the study area. The rainfall distribution reflects the spatial heterogeneity of the natural canyon in the study area. Correlational relationships between observed rainfall data (from 14 meteorological stations) and predicted values from spatial interpolation methods are presented in Figure 4. The CV method adopted in this study is the leave-one-out method, which was applied to all three methods. The prediction errors from cross-validation can help in determining the most accurate prediction model. For an accurate model, the mean error should be close to 0 and the root-mean-square error should be as small as possible (Johnston et al., 2001). The mean absolute error (MAE), the mean relative error (MRE), and the root-mean-square error (RMSE) were calculated for the resulting residuals and compared. The results from the kernel smoothing interpolation method used showed that the best-fit model is where the mean is 1.710082943, the root-mean-square is 101.954642 mm and the root-mean-square error is 100.86542, respectively.

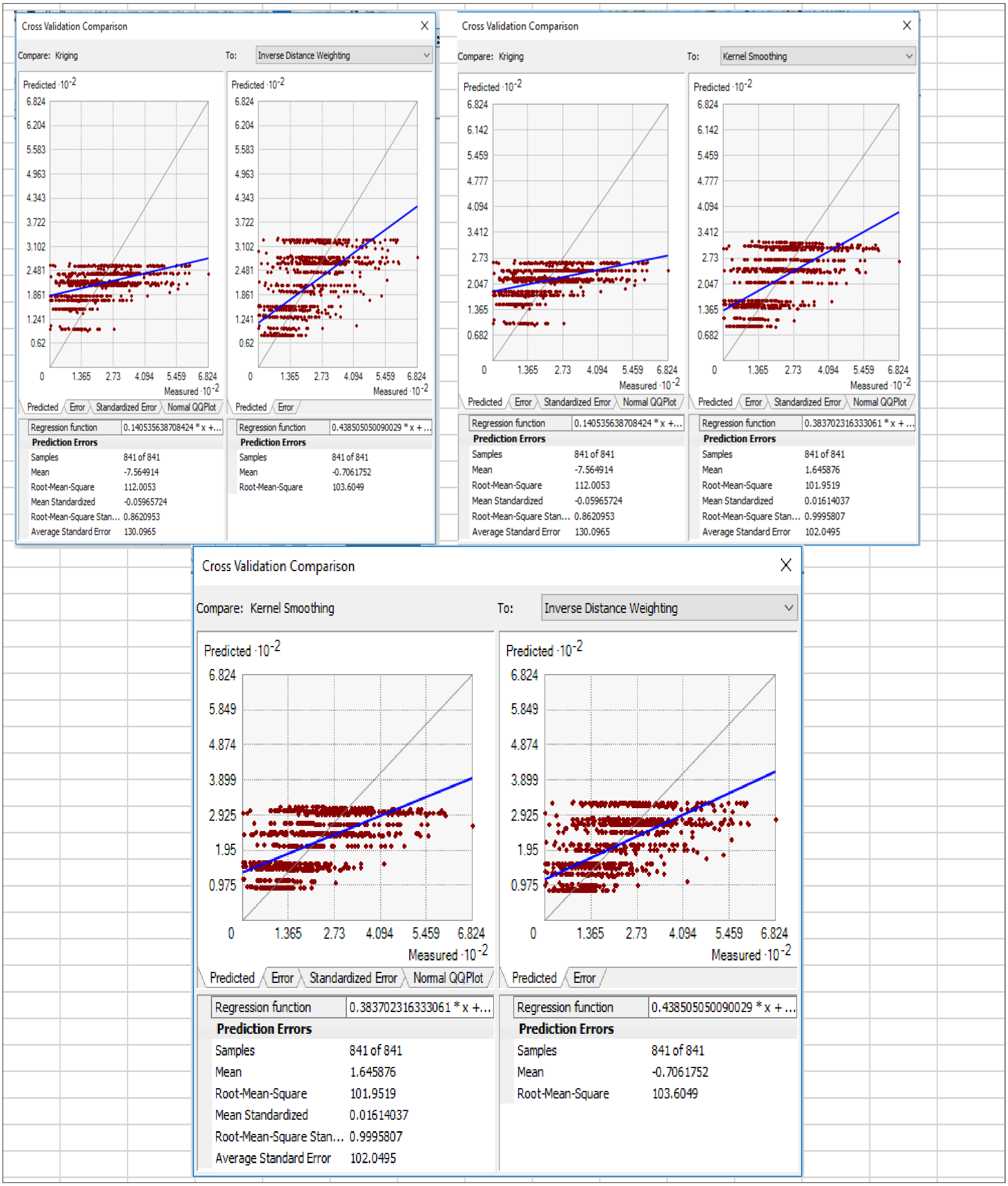

Figure 4. Relationship between observed rainfalls values (14 meteorological stations) and predicted values using the three interpolation methods.

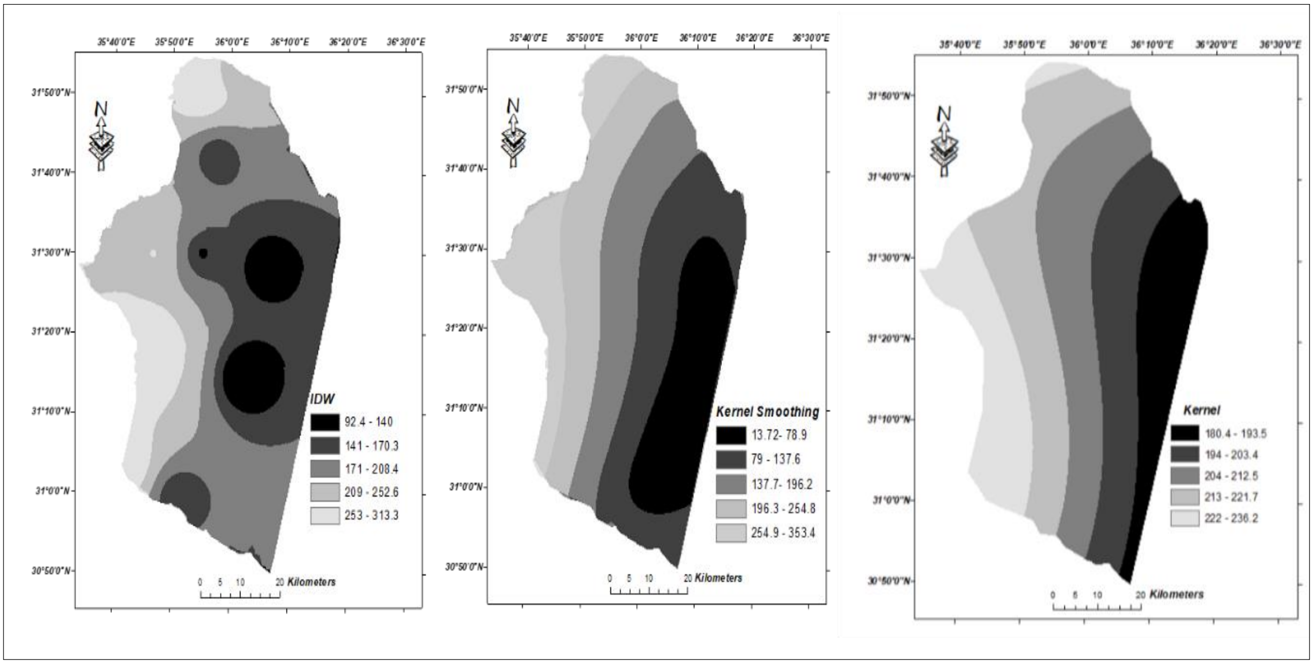

Figure 5. Average annual precipitation surfaces (for the period 1980-2016) as analyzed by different interpolation methods.

5.2. Results

The results, based on cross-validation (which were indicated by the mean absolute error (MAE), the mean relative error (MRE), and the root-mean-square error (RMSE) for the different interpolation methods, are presented in Figure 4 and Table 2. The best results were obtained by the least RMSE for interpolating the precipitation (RMSE) = 100.86542, while the inverse distance weighting (IDW) performed the worst.

5.3. Conclusion

The three interpolation methods of annual mean precipitation over a period of 37 years in the Wadi Al-Mujib Basin were compared. The accuracy of interpolation was determined by cross-validation. Mean, Root-Mean-Square, Mean Standardized, Root-Mean-Square Standardized, Average Standard Error, and the Root-Mean- Square Error (RMSE) were chosen as validation criteria. Visual results are given in Table 2 and Figure 3. The comparison of cross-validation showed that the best results were obtained by KS with the lowest RMSE for interpolating the precipitation (RMSE) = 100.86542.

REFERENCES

Al-Bakri J. T., Sari Shawash, S., Ghanem, A. and Abdelkhaleq, R., 2016, Geospatial Techniques for Improved Water Management in Jordan, ***Water,*** **8**(4), 132; doi:10.3390/w8040132,P1-22.

Amr, Z., Al Azzam, L., Katbeh-Bader, A., and Eid, E., 2013, Odonata of Wadi Al Mujib Catchment with Notes on the Impact of Wadi Al Mujib Dam, Jordan (Insecta: Odonata), ***Jordan Journal of Biological Sciences***, **6**(4), pp 292–299.

Anders, A.M., Roe, G.H., Durran, D.R., Montgomery, D.R., and Hallet, B., 2004, Precipitation and the form of mountain ranges, ***Bulletin of the American***

Meteorological Society, 85, pp 498–499.

Bohling, G., 2005, ***Kriging*** https://www.mailman.columbia.edu/research/population-health-methods/kriging

Burrough, P. A., and McDonnell, R. A., 1998, ***Principles of Geographical Information Systems***, Oxford, Oxford University Press.

Childs, C., 2004, Interpolating Surfaces in ArcGIS Spatial Analyst, ***ESRI Education Services***, http://webapps.fundp.ac.be/geotp/SIG/interpolating.pdf Accessed 12 May 2012.

Dobesch, H., Dumolard, P. and Dyras, I., (Eds), 2007, ***Spatial Interpolation for Climate Data: The Use of GIS in Climatology and Meteorology***, London, ISTE Ltd..

Dryas, I., and Ustrnul, Z., 2007, The Spatial Analysis of the Selected Meteorological Fields in the Example of Poland, in: Dobesch, H., *et al*. (Eds), ***Spatial Interpolation for Climate Data: The Use of GIS in Climatology and Meteorology***, London: ISTE Ltd, pp. 87-96.

Ghanem, A. A., 2013, Case Study: Trends and Early Prediction of Rainfall in Jordan, ***American Journal of Climate Change,*** 2, pp 203-208.

ArcGIS, help; available at http://pro.arcgis.com/en/pro-app/help/analysis/geostatistical-analyst/deterministic-methods-for-spatial-interpolation.htm

Isaaks, E.H, and Srivastava, R.M., 1989, ***An Introduction to Applied Geostatistics***, New York, Oxford University Press.

Johnston, K., Ver Hoef, M. J., Krivoruchko, K. and Lucas, N., 2001, ***Using ArcGIS Geostatistical Analyst,*** ESRI.

Lam K. C., Bryant, R. G., Wainright, J., 2015, Application of Spatial Interpolation Method for Estimating the Spatial Variability of Rainfall in Semiarid New Mexico, USA, ***Mediterranean Journal of Social Sciences***, **6**(4), pp 108-116.

Noy-Meir, I., 1973, Desert ecosystems: environments and producers, ***Ann. Rev. Ecol. Syst.***, 4, pp 398-417.

Philips, D. L., Dolph, J., and Marks, D., 1992, A comparison of geostatistical procedures for spatial analysis of precipitation in mountainous terrain, ***Agricultural and Forest Meteorology***, **58**(1-2), pp 119-141.

Pol, J., Environ, ***Stud***, **25**(3), pp 1339-1345.

Shu, S. J., Yu, Z. F., Wang Y., Ma, M. A., 2005, Statistical model for the spatial distribution of rainfall estimation over the Tibetan complex terrain. ***Chinese J. Geophys***, **48**(3), p 535.

Sluiter, R., 2009, ***Interpolation Methods for Climate Data Literature Review***, KNMI intern rapport; IR 2009-04, De Bilt.

Sunyani, A. M., 2004, Geostatistical Study on Annual and Seasonal Rainfall Patterns in Southwest Saudi Arabia, ***Hydrological Sciences Journal***, **49**(5), pp 803-817.

Wagner, P. D., Fiener, P., Willken, F., Kumar, S., and Schneider, K., 2012, Comparison and evaluation of spatial interpolation schemes for daily rainfall in data scarce regions, ***J. Hydrol***, **464**, p 388.

Zhang, X., Lu, X., Wang, X., 2016, Comparison of Spatial Interpolation Methods Based on Rain Gauges for Annual Precipitation on the Tibetan Plateau, ***Pol. J. Environ. Stud.***, **25**(3), pp 1339 – 1345.

مقارنة بين طرائق الاستيفاء المكاني لتقدير كمية الهطول المطري السنوي في حوض وادي الموجب في الأردن

يسرى الحسبان *

الملخّص

تعتبر دقة البيانات الدقيقة عن كمية الهطول السنوي ضرورية للتطبيقات البيئية في التقييم الفعلي للتوزيع الجغرافي لهطول الأمطار. تستخدم طرائق الاستيفاء لرصد التوزيع المكاني لكمية هطول الأمطار خصوصاً في حال قلة محطات الرصد المطري. تتعدد طرائق الاستيفاء المكاني، ولكن لا يمكن لأية طريقة منها تحقيق أفضل النتائج في جميع الحالات. وفي هذه الدراسة، تم تطبيق ثلاثة طرائق مختلفة من الاستيفاء لتحديد الطريقة الأمثل للتوزيع المكاني لكمية الأمطار. تم استيفاء بيانات الهطول المطري من 14 محطة أرصاد جوية مكانيًا باستخدام ثلاث تقنيات استيفاء شائعة للمقارنة بين دقة نتائجها: وهذه الطرائق هي: معكوس المسافة المرجحة أو الموزونة IDW، والإستفاء الإعتيادي أو العادي Ordinary Kriging، والمنحنيات السلسة للقمم أو الأنوية Kernel smoothing، وتقييم نتائجها باستخدام Cross-validation مقارنة بالبيانات الحقيقة لكمية الهطول المطري من محطات الرصد المطري في منطقة الدراسة، من خلال مقدار متوسط خطأ الجذر التربيعي (RMSE). تم الحصول على أفضل النتائج من خلال أدنى لطريقة الاستيفاء العادي؛ اذ بلغ ماوسط الخطأ 100.86542، بينما كانت نتئج المسافة العكسية الموزونة الأسوأ، والأقل كفاءة مع أكبر نسبة خطأ والبالغة 103.43.

الكلمات الدالّة: الاستيفاء، معدل الهطول السنوي، المسافة العكسية الموزونة، الاستيفاء العادي، المنحنيات السلسلة للقمم، التحقق، طريقة الإحصاء المكاني.

---

* أستاذ دكتور قسم الجغرافيا، الجامعة الأردنية، عمان

تاريخ استلام البحث 2019/3/25، وتاريخ قبوله 2021/6/16.