

UDK/UDC: 502/504:551.577.5(55)(078.7)

Prejeto/Received: 14.01.2023

Izvirni znanstveni članek – *Original scientific paper*

Sprejeto/Accepted: 02.10.2023

DOI: [10.15292/acta.hydro.2023.09](https://doi.org/10.15292/acta.hydro.2023.09)

Objavljeno na spletu/Published online: 18.04.2024

SPATIAL STATISTICS ANALYSIS OF PRECIPITATION IN THE URMIA LAKE BASIN PROSTORSKA STATISTIČNA ANALIZA PADAVIN V POREČJU JEZERA URMIA

Hossein Aghamohammadi¹, Saeed Behzadi^{2,*}, Fatemeh Moshtaghinejad¹

¹ Department of Remote Sensing and GIS, Faculty of Natural Resources and Environment, Science and Research Branch, Islamic Azad University, Tehran, Iran

² Department of Surveying Engineering, Faculty of Civil Engineering, Shahid Rajaei Teacher Training University, Tehran, Iran

Abstract

Most of the world's population lives in areas facing a severe water crisis. Climatology researchers need precipitation information, pattern analysis, modeling of spatial relationships, and more to cope with these conditions. Therefore, in this paper, a comprehensive approach is developed for describing geographic phenomenon using various geostatistical techniques. Two main methods of interpolation (Inverse Distance Weighting and Kriging) are used and their results are compared. The Urmia Lake Basin in Iran was selected as a case-study area that has faced critical conditions in recent years. Precipitation was initially modeled using both conventional, non-statistical approaches and advanced geo-statistical methods. The result of the comparison shows that ordinary Kriging is the best interpolation method for precipitation, with an RMS of 4.15, and Local Polynomial Interpolation with the exponential kernel function is the worst method, with an RMS of 5.02. Finally, a general regression analysis was conducted on precipitation data to examine its relationship with other variables. The results show that the latitude variable was identified as the dependent variable with the most influence on precipitation, with an impact factor of 81%, and that the slope has the lowest impact on precipitation, at nearly zero percent. The influence of latitude on precipitation appears to be localized, suggesting that it may not be a significant variable for predicting global environmental threats.

Keywords: Precipitation Estimation, Geostatistics, Spatial Relationship Modeling, Kriging interpolation.

Izvleček

Večina svetovnega prebivalstva živi na območjih, ki se soočajo s hudo krizo zaradi pomanjkanja vode. Klimatologi za spopadanje s temi izzivi potrebujejo informacije o padavinah, analize prostorskih vzorcev in modele prostorskih odnosov. V prispevku opisujemo celoviti pristop k opisovanju geografskega pojava z uporabo različnih geostatističnih tehnik. Uporabljeni sta dve glavni metodi interpolacije (metoda inverzne utežene razdalje in Kriging) ter primerjani njuni rezultati. Kot območje študije primera je bilo izbrano porečje jezera Urmia, ki se je v zadnjih letih soočalo s kritičnimi razmerami. Padavine smo najprej modelirali s

* Stik / Correspondence: behzadi.saeed@gmail.com

© Aghamohammadi H. et al.; This is an open-access article distributed under the terms of the [Creative Commons Attribution – NonCommercial – ShareAlike 4.0 Licence](https://creativecommons.org/licenses/by-nc-sa/4.0/).

© Aghamohammadi H. et al.; Vsebinska tega članka se sme uporabljati v skladu s pogoji [licence Creative Commons Priznanje avtorstva – Nekomercialno – Deljenje pod enakimi pogoji 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/).

klasičnimi in geostatističnimi metodami. Rezultati kažejo, da je navadni Kriging najboljša interpolacijska metoda za padavine – z vrednostjo RMS 4,15, metoda z eksponentno jedrno funkcijo pa je najslabša – z vrednostjo RMS 5,02. Na koncu je bila izvedena splošna regresijska analiza padavin. Rezultati kažejo, da je bila spremenljivka širine najvplivnejša odvisna spremenljivka s faktorjem vpliva 81 %, naklon pa ima najmanjši vpliv na padavine s skoraj nič odstotki. Zdi se, da je vpliv zemljepisne širine lokalne narave in morda ne predstavlja pomembne globalne okoljske grožnje.

Ključne besede: ocena padavin, geostatistika, modeliranje prostorskih odnosov, interpolacija Kriging.

1. Introduction

Precipitation plays an essential role in the global water and energy cycle. More than 40% of the world's population lives in areas with a severe water crisis (Bostan et al., 2012). Iran is in one such dry zone, receiving at most one-third of the average global precipitation (Eivazi and Mosaedi, 2012). According to the Food and Agriculture Organization of the United Nations (FAO), the world received an average of 890 millimeters of rain in 2013. In contrast, Iran's average precipitation hovers around 260 millimeters, indicative of a significant water crisis in the country (Maris et al., 2013). Researchers have continually sought accurate rainfall data for precipitation zoning, analyzing rainfall patterns, and estimating precipitation so as to manage diverse environmental conditions (Behzadi and Ali Alesheikh, 2013; Mahjoobi and Behzadi, 2022; Moral, 2010).

More accurate rainfall estimates are essential as the inputs for various analytical models in climate science. Achieving a precise estimate of rainfall is crucial for analyzing rainfall patterns spatially and temporally (Cristiano et al., 2017). In acknowledgement of this issue's importance, barometric stations to measure rainfall were installed in most of the region, alongside synoptic and climatological stations. Nonetheless, researchers often encounter challenges in interpolating and zoning static data, primarily due to the spatial and temporal variability of precipitation (Jalilzadeh and Behzadi, 2019; Moral, 2010).

To address these challenges, researchers have developed various methods for estimating precipitation, with some relying on geostatistical approaches. Significant strides have been made in recent decades, enabling the study and prediction of precipitation's spatial and temporal distribution

(Behzadi and Mousavi, 2019; Benoit and Mariethoz, 2017). Numerous international studies have been conducted on precipitation estimation, often employing Geographic Information System (GIS) applications and ground-based methods to produce comprehensive precipitation estimates and coordinated maps.

For instance, Karayusufoglu et al. (2010) explored and estimated spatial parameters and precipitation distribution in Turkey's Solakli Basin, utilizing interpolation techniques like inverse distance weighting (IDW), Kriging, and others. Their research concluded that the Kriging method demonstrated superior accuracy (Karayusufoglu et al., 2010). Abo-Monasar et al. (2014) analyzed rainfall in China, utilizing general linear regression and Kriging methods with data from 684 meteorological stations. The study opted for the stepwise regression method to select six auxiliary variables: longitude, latitude, elevation, slope, surface roughness, and river density. The Kriging method outperformed other prediction techniques (Abo-Monasar and Al-Zahrani, 2014). Baykal et al. generated predictive maps using GIS to assess climate-induced changes. Employing climate classification and time series methods in select Turkish provinces, the study underscored the suitability of interpolation methods for precipitation estimation (Baykal et al., 2022). Zhong et al. focused on water management strategies like rainwater harvesting, employing LiDAR data to represent catchment and vegetated areas in the Southwestern United States. Their results indicated the feasibility of rainwater harvesting for irrigation for approximately eight months of the year (Zhong et al., 2022). Longo-Minnolo et al. leveraged alternative data sources for estimating precipitation in Italy. Their research demonstrated that auxiliary methods, alongside ground data, could effectively

estimate precipitation (Longo-Minnolo et al., 2022). Zou et al. employed various interpolation methods to estimate precipitation levels (Zou et al., 2021). Jalilzadeh and Behzadi developed a fuzzy algorithm to emphasize the importance of satellite images in flood estimation. The fuzzy logic algorithm identified water levels in the area with an accuracy rate of 87% (Jalilzadeh and Behzadi, 2020). Various studies, including those by Baykal et al., Zhong et al., Longo-Minnolo et al., Zou et al., and Jalilzadeh and Behzadi, collectively highlight the reliability of interpolation methods for estimating precipitation levels and assessing climate-induced changes.

Reviewing these previous studies, two significant research areas stand out: precipitation zoning and the analysis of spatial variations in rainfall. To comprehend the spatial variability of rainfall, it is essential to consider precipitation's intricate and continuous interplay with other elements, both spatially and temporally. This dynamic behavior of precipitation has captured the attention of climatologists and researchers involved in climatology (Papalexiou et al., 2018). One approach to studying precipitation is describing its spatial variation using spatial statistics (Liu et al., 2022). Furthermore, all spatial events inherently possess temporal components (Behzadi and Alesheikh, 2013; Maris et al., 2013).

In contrast to classical statistics, spatial statistics reveal the statistical properties of distributions in space (Abdollahi and Behzadi, 2022; Ghashghaie and Behzadi, 2019). This enables the identification of spatial differences, similarities, specific and unique points, or homogeneous regions. These insights facilitate determining the size or extent of spatial phenomena is possible (Sun et al., 2009). Recognizing the significance and application of spatial analysis, many researchers have delved into identifying climatic elements, particularly rainfall. Precipitation estimation and modeling are fundamental aspects of climate research. As such, considerable efforts have been made by climate scientists, encompassing zoning, estimation, and

analysis of rainfall patterns to strengthen our understanding of various regions and to improve environmental management strategies.

In this paper, we initially employ a range of interpolation methods, including traditional and geostatistical approaches, to identify the optimal model for precipitation in the Urmia Lake Basin. Further, we evaluate the correlations between the obtained precipitation model and various geographic factors, such as geographical location, altitude, and regional topography, utilizing diverse spatial statistics techniques. The remainder of this paper is organized as follows: the next section introduces the theoretical principles. Section 3 outlines the materials and methods employed, followed by a section presenting experimental results and a comparative analysis with other methods. Finally, the discussion is wrapped up and conclusions are drawn in the last section.

2. Theoretical Principles

2.1 Interpolation

Interpolation is estimating quantitative values for unknown points using adjacent and distinct points. This process is conducted when the point data is insufficient for producing co-value maps. Therefore, interpolation means converting point data to zonal data. The general interpolation equation is given by Equation 1; the difference among the various models is only in the weighting factor (Yang et al., 2020):

$$\hat{Z}(s_0) = \sum_{i=1}^N \lambda_i Z(s_i) \quad (1)$$

where $\hat{Z}(s_0)$ is the estimated value in position s_0 , $Z(s_i)$ is the measured value in s_i , i represents the measured points, and N is the number of stations.

There are different methods for interpolation, which are shown in Figure 1 (Bajat et al., 2013; Cheng et al., 2008; Jafarian and Behzadi, 2020).

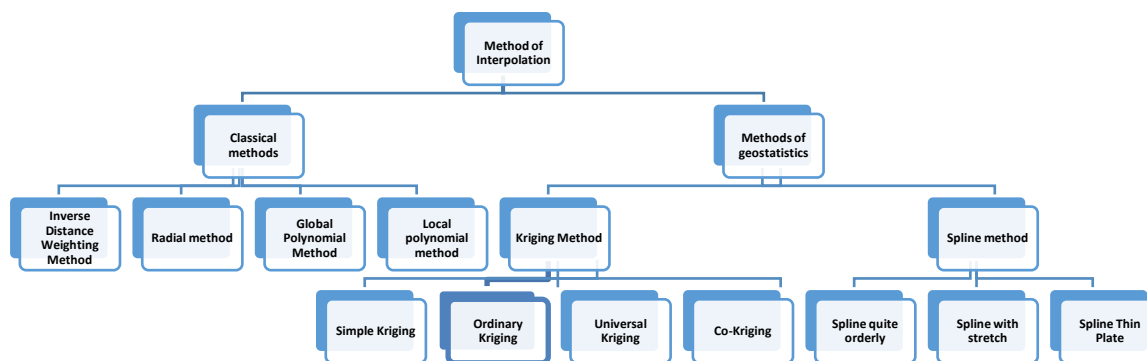


Figure 1: Different types of interpolation methods.

Slika 1: Različne vrste metod interpolacije.

After employing interpolation, such methods should be evaluated (Njeban, 2018). There are several ways of comparing the results obtained from these methods. The cross-validation method is one of the most appropriate and most widely used methods. In this method, the interpolation operation is carried out again by removing each of the measured points (such as the barometric station in this paper). Then the difference between the estimated value and the actual value is calculated as a measure of error. Since the resulting error is obtained for all points, the estimation is done for all points (Kumar et al., 2022). The validation method uses various parameters for comparison. In this paper, the RMS method is used for all interpolation.

2.2 Clustering and analysis of rainfall patterns

In this study, we intended to use interpolation as a prerequisite for other analyses of precipitation patterns and rainfall estimation and modeling. To analyze the precipitation patterns, researchers use a variety of algorithms, including the mean of the nearest neighbor, multi-spatial analysis, k-rupli function, etc.

Using these analyses facilitates studying the cluster behavior of certain phenomena (Triguero et al., 2019). Some of these analyses present the results as a statistical report, and others on the map (Bostan et al., 2012). After analyzing the distribution pattern of phenomena and their spatial distribution as spatial statistics and standard z distributions, researchers must also show cluster and distribution, as well as the pattern of dispersion of complications with their attribute value on a map (Maris et al., 2013). There

are several tools for creating cluster maps, most of which are: 1) Analysis of the cluster with Anselin Local Moran's I index and 2) the Hot Spot Analysis or Getis-Ord G_i^* statistic.

As stated, general Moran statistics only show cluster behavior for annual rainfall and cannot express the type of spatial behavior exhibited by the concerned phenomenon on the map; therefore, to reveal the behavior of annual rainfall in terms of spatial distribution patterns on the map, Moran local statistics are used (Abdollahi and Behzadi, 2022). Hot spot analysis, as another method for mapping clusters, calculates the G_i^* statistic for all data points. The resulting computed statistic, z, indicates where quantities are high or low, representing hot or cold clusters in the study area or dataset (Ghashghaie and Behzadi, 2019).

2.3 Modeling of spatial relationships

One of the most popular methods for modeling the spatial relationships of geographic problems, such as precipitation, is the Ordinary Least Squares Regression (OLS) method (Huang, 2018). In the sense of regression, it is assumed that we look at a variable such as Y over time or among different units and obtain the relevant data, and then its changes are interpreted (Bostan et al., 2012). For this dependent variable, consider the variable or variables that can explain these changes. Suppose (Mirzaei and Sakizadeh, 2016):

$$y_t = f(x_{1,t}, \dots, x_{k,t}) \quad (2)$$

Equation (2) constitutes a mathematical model as it solely represents the mathematical relationship

between the dependent variable (Y) and the independent variables (Xs). If the function F exhibits linearity concerning the variables X_i (as depicted in Equation 3), it is referred to as a linear mathematical model.

$$y_t = \beta_0 + \beta_1 x_{1,t} + \dots + \beta_k x_{k,t} \quad (3)$$

A generalized least squares regression is a form of linear regression for predicting or modeling dependent variables to a set of independent (or explanatory) variables. Recognizing and evaluating the relationship between these two categories of variables helps us better understand the relationship between dependent and independent variables. It also shows everything that is happening in one place. The general squares regression is the most well-known regression technique (Karayusufoglu et al., 2010; Mirzaei and Sakizadeh, 2016). Regression provides an equation to represent this process by providing a global model of variables, and understanding the relationship between variables (Fernández-Delgado et al., 2019).

3. Materials and methods

3.1 Area of case study

The Urmia Lake Basin is an invaluable aquatic ecosystem situated in the northwest region of Iran, renowned both nationally and globally. This ecosystem serves as a quintessential illustration of a closed basin, wherein all river runoffs within the basin converge. This lake, with an area of 51,876 square kilometers, is one of the main basins in the country, located between the provinces of West Azerbaijan (46%), East Azerbaijan (43%), and Kurdistan (11%). The lake is also the 25th-largest lake in the world in terms of surface area, the largest inland lake in Iran, and the second-largest saltwater lake in the world. The topography of the basin area of Lake Urmia is shown in Figure 2.

Unfortunately, despite this lake's unique importance in Iran, it has been experiencing significant depletion since 2000, and today, it is at risk of drying up completely. Several factors, such as rainfall, river flow, evaporation, and temperature, contribute to this process. Therefore, this basin was selected as the study area to evaluate precipitation's influence on this ongoing drought occurrence.

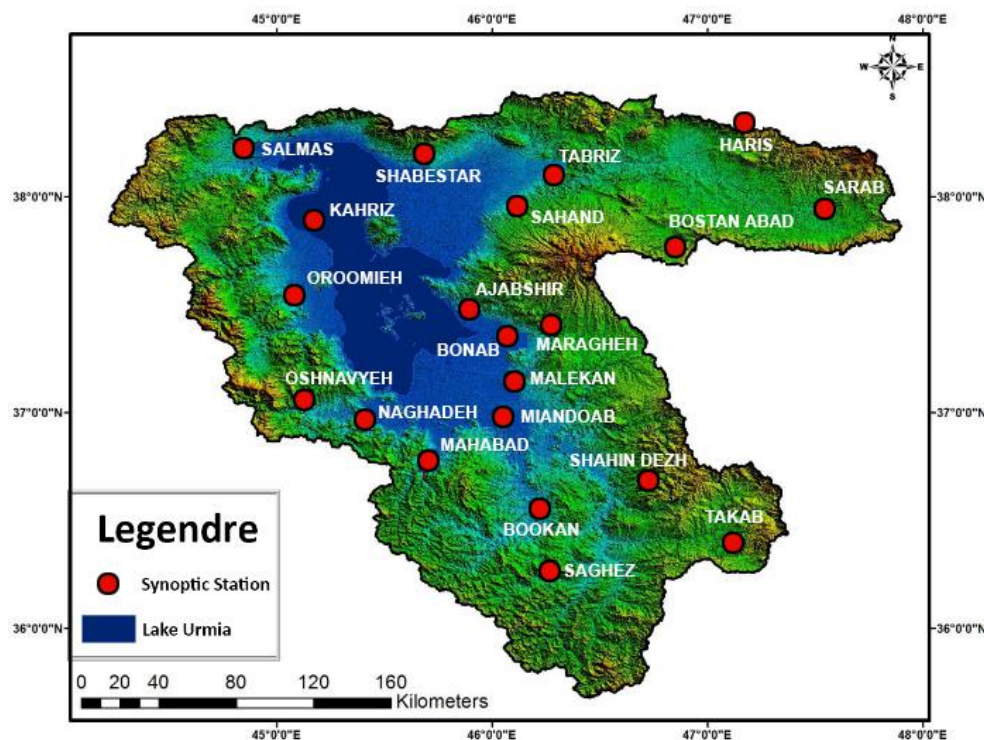


Figure 2: The studied area.

Slika 2: Obravnavano območje.

3.2 Data

For this study, the annual precipitation data of 21 synoptic stations of the Meteorological Organization were used, spanning 63 years of statistics between 1951 and 2014. Then, based on these statistics, the average annual precipitation of the stations was calculated as the dependent variable. In addition, latitude and longitude, along with the height and slope of the terrain at each station, as well as average annual and average annual wind speed, were extracted as independent variables. Figure 2 shows the scope of the study area, and the location of the stations.

3.3 Methods of research implementation

The methods and steps for implementation are summarized in five systematic stages, which serve as a structured approach for analyzing geographic rainfall. This framework is not limited to precipitation alone; it can be adapted for the analysis of various other geographical phenomena.

Within this framework, we begin by interpolating rainfall data using a range of established ground-based statistical and traditional methods. Subsequently, the most suitable interpolation method is chosen based on cross-validation results. In the third stage, we delve into our analysis of rainfall patterns, employing four key indicators. Moving forward, we assess the distribution patterns of these phenomena, employing spatial statistics and standard z distribution. In the fourth stage, we apply two distinct cluster analysis techniques: Anselin Local Moran's I and hot spot analysis. Finally, in the fifth and concluding phase, we model the spatial relationships within the rainfall dataset using the generalized least squares regression method. These relationships are expressed through a linear equation and visualized on a map. It's important to note that these steps collectively offer a comprehensive perspective on addressing the precipitation issue.

4. Results

4.1 Precipitation interpolation

For many interpolation methods, we first need to examine a series of initial statistical tests on input data. These statistical tests include: 1) Examining the normalization of rainfall data at stations and normalizing them if needed; and 2) Examining the data for a trend and removing it if it exists.

4.2 Examining the Normality of Data

We conducted a rigorous examination of rainfall data normalization at weather stations using the Kolmogorov-Smirnov test, a widely recognized statistical method for assessing the data's adherence to normal distribution. The significance level (Sig.), depicted in Figure 3, plays a crucial role in our analysis. This significance level represents the p-value derived from the Kolmogorov-Smirnov test.

In hypothesis testing, a common practice is to set a significance level, often denoted as α , which serves as a critical threshold for determining whether observed data significantly deviate from a normal distribution. In our study, we adopted a standard significance level of 0.05.

The statement "Sig. above 0.05" in Figure 3 signifies that the computed p-value from the Kolmogorov-Smirnov test exceeded this 0.05 threshold. This outcome holds significant implications: it indicates that we lack substantial evidence to reject the null hypothesis. In this context, the null hypothesis posits that the rainfall data at Lake Urmia's weather stations follow normal distribution.

It's crucial to recognize that, despite any visual deviations suggested by the histogram, the statistical rigor provided by the significance level reinforces our conclusion. The data's adherence to a normal distribution is affirmed, as the observed deviations, if present, are not statistically significant.

In light of these findings, we confidently assert that the rainfall data can be considered standard and there is no necessity for data normalization. This conclusion, supported by rigorous statistical analysis, bolsters the reliability of our results.

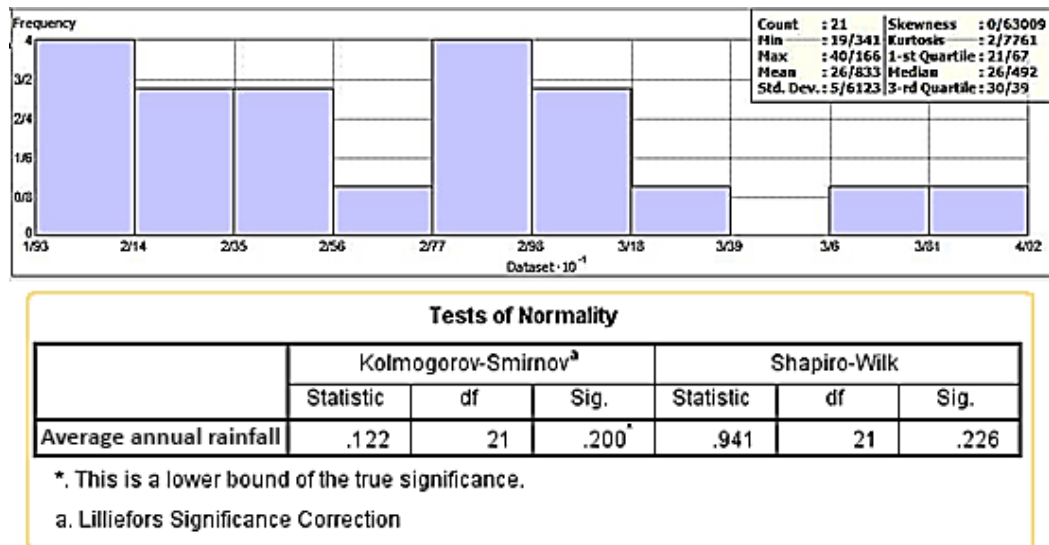


Figure 3: Normality of rainfall data at Lake Urmia.

Slika 3: Test normalne porazdelitve podatkov o padavinah ob jezuru Urmia.

4.3 Examining data trends

Another important consideration is the absence of any discernible trend in the variable data being estimated. In cases where a trend does exist, it should be addressed through appropriate data preprocessing methods. Figure 4 provides a visual representation of this aspect. From a west-to-east perspective, there is no evident trend in the data. However, when examining the data from a north-to-south perspective, a second-order polynomial trend becomes apparent. This trend must be accounted for in certain interpolation methods, such as Kriging and general co-Kriging.

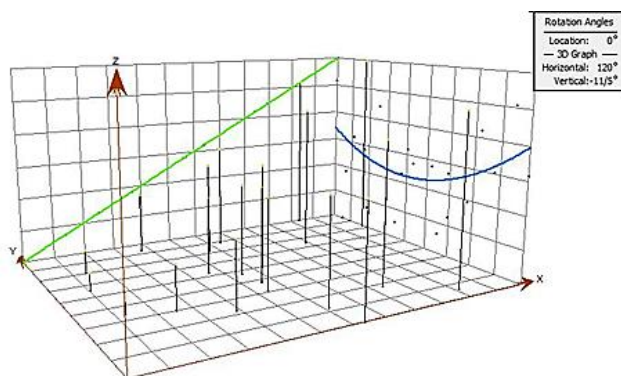


Figure 4: The trend in rainfall data in Lake Urmia.

Slika 4: Trend v podatkih o padavin ob jezuru Urmia.

4.4 Precipitation interpolation in the Urmia Lake Basin

After examining the preliminary statistical tests on the data, various types of traditional interpolation and geostatistical methods were implemented. Then the best rainwater map was drawn using the cross-validation. It should be noted that in the IDW method, the optimal power is 2, and the neighboring type is considered standard. In the Radial Basis Function (RBF) method, various kernel functions are employed, and the kernel function that yields the smallest error is selected as the best estimation. In the context of terrestrial-based techniques, a variety of half-diffraction models were employed, with meticulous attention to minimizing errors. The results of implementing these methods, with the estimation of each method’s accuracy using cross-validation, are presented in Table 1.

According to the results, the Ordinary Kriging method with a half-shift function of the Holl Effect and the lowest RMSE was selected as the best method for precipitation interception in the Urmia Lake Basin. In this way, its interpolation map is shown in Figure 5.

Table 1: Implementation of Various Methods of Rainfall Interpolation for the Urmia Lake Basin.

Preglednica 1: Izvajanje različnih metod interpolacije padavin za porečje jezera Urmia.

Interpolation method	RMSE
IDW (Inverse Distance Weighting)	4.63
RBF (Radial Basis Function) with the Inverse Multi-Quadric kernel function	4.43
GPI (Global Polynomial Interpolation)	4.62
LPI (Local Polynomial Interpolation) with Exponential kernel function	5.02
Ordinary Kriging with Hole Effect Half Effect Function	4.15
Simple Kriging with Stable Shift	4.31
Ordinary Kriging with a Half-Circle Change Function	5.67
Co-Kriging with Exponential Half-Effect Function	4.31
Simple Co-Kriging with Hole Effect Half-Effect Function	4.33
Co-Kriging with a Half-Spherical Change Function	4.72

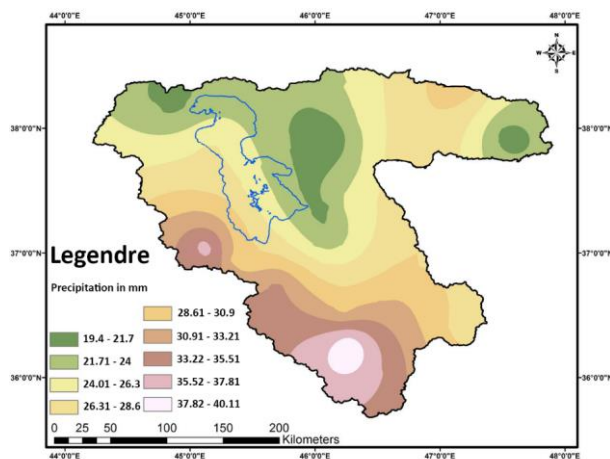


Figure 5: The Urmia Lake Basin reconnaissance map by the ordinary Kriging method.

Slika 5: Karta porečja jezera Urmia z metodo navadnega Kriginga.

4.5 Spatial Data Analysis and Analysis of the geostatistical of Precipitation in the Urmia Lake Basin

In this section, rainfall distribution in the Urmia Lake Basin is studied using geostatistical analysis. To apply these analyses, the data from the Urmia Lake Basin zoning map must be used (Figure 5).

4.6 Spatial Distribution Analysis of Observation Points

The rainfall pattern analysis in the Urmia Basin using the mean of the nearest neighboring method indicates that the annual rainfall pattern of Lake Urmia is a dispersed pattern with a high average of the nearest neighboring point of 1.476. This result is obtained due to the value of $z > 96.1$ and due the magnitude of this value, which is in the sequence of z distribution in the red region of Figure 6 and its value is 187.832, and the P-value statistic is zero. Also, the average value of the observed distance is greater than the mean of the expected distance, which is more evidence of the pattern of precipitation dispersion in the Urmia Lake Basin.

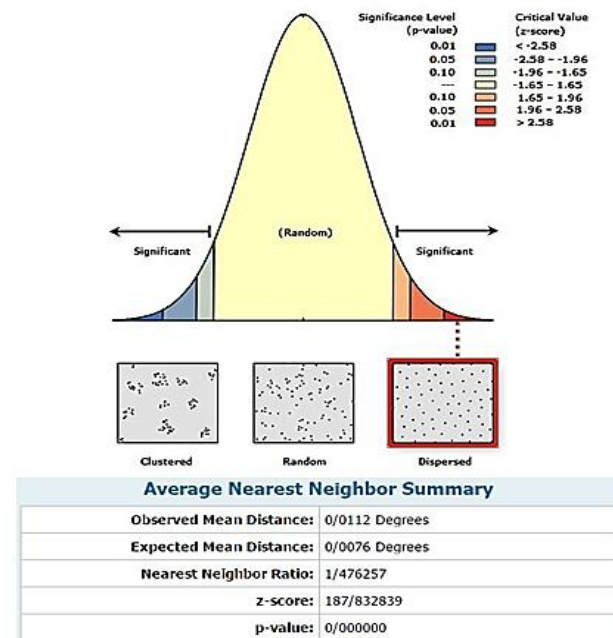


Figure 6: Analysis of the rainfall pattern at Lake Urmia with the Average nearest neighborhood.

Slika 6: Analiza vzorca padavin ob jezeru Urmia z metodo najbližjih sosedov.

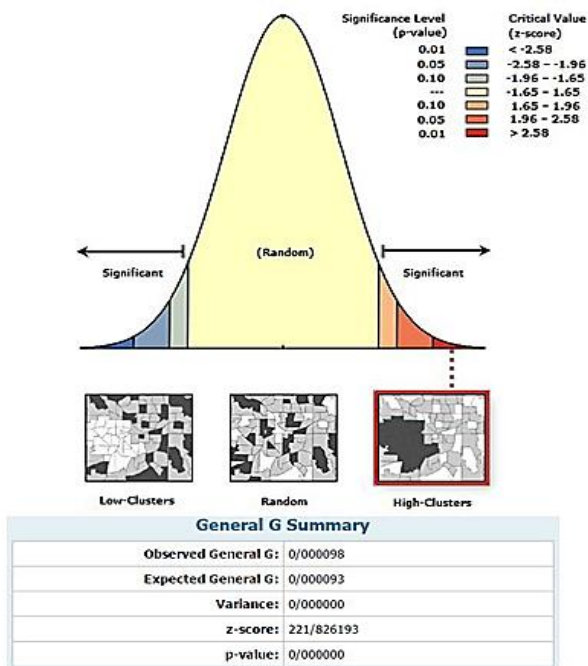


Figure 7: Analysis of Lake Urmia's precipitation pattern using the G_i statistic.

Slika 7: Analiza vzorca padavin na jezeru Urmia z uporabo statistike G_i .

4.7 High/Low Clustering (Getis-Ord General G)

The analysis of precipitation patterns in the Uremia Lake Basin using G_i statistics reveals a high concentration of cluster patterns in its annual precipitation. This result was obtained based on the standard statistic $z > 96.1$, indicating a significance level higher than 96.1% and placing it in the red region of Figure 7. Additionally, the p-value, along with zero variance, signifies a high-focus cluster pattern. The Getis-Ord General G_i^* statistics are also very close to zero, with both the expected and observed values almost reaching zero. If the standard statistic z is situated on the left-hand side of the distribution with a negative value, this pattern exhibits low-cluster behavior. The median of a bell-shaped graph indicates a random or statistically insignificant pattern.

4.8 Multi-Distance Spatial Cluster Analysis Index

The Multi-Distance Spatial Cluster Analysis index shows a cluster pattern in the Lake Urmia

Basin's precipitation behavior. As seen in Figure 8, the red line of observed data is at all distances above the expected blue line.

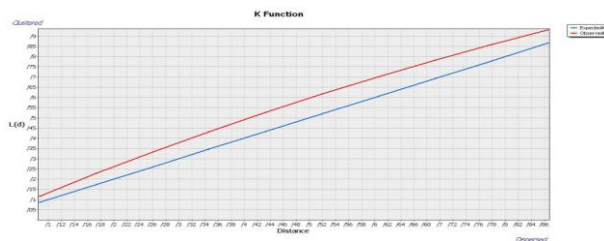


Figure 8: Multi-Distance Spatial Cluster Analysis (Ripley's K) for precipitation data of Lake Urmia Basin.

Slika 8: Analiza prostorskih grozdov z več razdaljami (Ripleyjeva K funkcija) za podatke o padavinah povodja jezera Urmia.

4.9 Spatial autocorrelation analysis of general Moran's I

General Moran spatial autocorrelation analysis shows that the Lake Urmia Basin's annual precipitation has a high concentration of clusters. As seen in Figure 9, the value of the standard statistic $z > 96.1$ is 1% higher than the number 96.1, and it is located in the sequence of z 's distribution in the red region of Figure 9. So, the p-value at zero represents a high-focus cluster pattern, and the universal Moran statistic was 99.9%. The Moran statistic is expressed by a correlation coefficient, and its value varies between -1 and +1. This means, if the amount of Moran statistics moves to +1, it shows a high cluster pattern concentrated in most of the study area. And if the Moran statistic moves to -1, it indicates a dispersed annual precipitation pattern. A value close to zero indicates a random and irrelevant pattern at the desired confidence level.

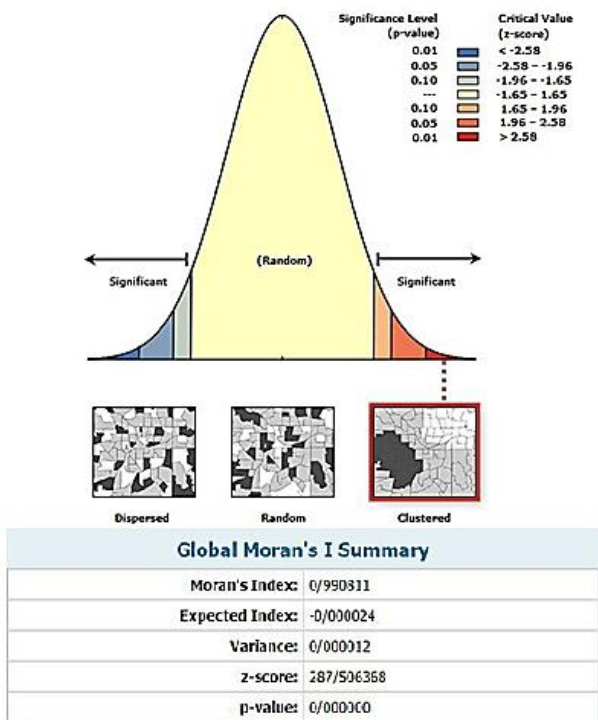


Figure 9: Spatial Autocorrelation Analysis of Precipitation Data of Lake Urmia Basin with general Moran.

Slika 9: Prostorska avtokorelacijska analiza podatkov o padavinah v porečju jezera Urmia s splošno Moranovo statistiko.

4.10 Mapping clusters analyzing

In this section, we utilize two methods of analysis, namely Cluster and Outlier Analysis (Anselin Local Moran's I), and Hot Spot Analysis (Getis-Ord G_i^*), to extract and map clusters. The objective is to shed light on the relevance and insights gained from the cluster analysis of rainfall data in the Urmia Lake Basin.

Cluster and Outlier Analysis conducted with Anselin Local Moran's I statistics (Figure 10) reveals distinctive spatial patterns. High-high clusters, representing areas with the most precipitation, are prominently located in the lake's southern and southwestern regions. Conversely, the central Urmia Lake Basin, as well as two isolated locations near the Sarab and Salmas stations, are identified as low-low clusters, indicating areas with the least precipitation. Additionally, other areas displayed as gray spots on the map exhibit no statistically significant patterns when assessed from

the high-high, low-low, high-low, and low-high perspectives.

This cluster analysis provides valuable insights into the spatial distribution of precipitation patterns within the Urmia Lake Basin, aiding our understanding of the areas with the highest and lowest levels of rainfall.

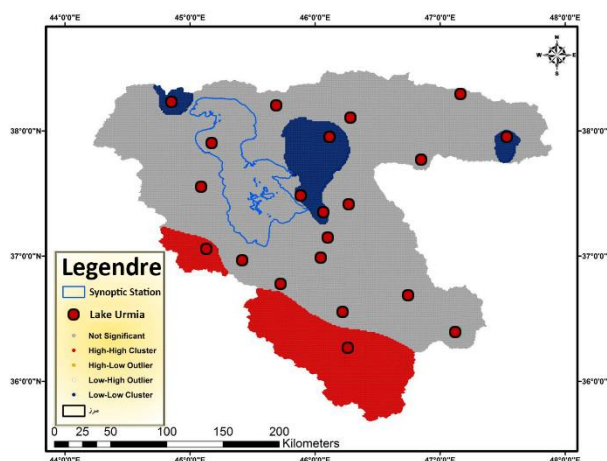


Figure 10: Cluster and Outlier Analysis (Anselin Local Moran's I) for precipitation data of Lake Urmia Basin.

Slika 10: Analiza grozdov in odstopanj (statistika Anselin Local Moran I) za podatke o padavinah v povodju jezera Urmia.

Also, with the aid of Hot Spot Analysis (Getis-Ord General G_i^*), (seen in Figure 11), it can be concluded that hot spots have more areas in the south and southwest of the study area in terms of values with high clusters. This represents high precipitation in terms of spatial autocorrelation analysis. The concentration of clusters with a relatively lower concentration than the higher clusters in the red range of precipitation distribution also appears in some of the areas. Cold spots located in the range of 90 to 99 percent are significant in the north and central parts of the Urmia Basin, and two spots are located in the Sarab and Salmas areas. Cold spots with a low concentration show precipitation and a cluster pattern. Other regions did not follow a specific spatial pattern and are not statistically significant.

4.11 Spatial modeling of precipitation

In the fifth step of our analysis, where we employ the least squares regression method to model the spatial relationships of rainfall in the Urmia Lake Basin, we included several explanatory variables and the dependent variable, rainfall. The selection of these specific variables was not arbitrary but based on prior considerations and a preliminary analysis. Before conducting the regression analysis, we conducted a thorough pre-analysis to determine which variables could potentially influence rainfall patterns in the Urmia Lake Basin. These variables were carefully chosen based on their known or hypothesized impact on rainfall. For instance, latitude and longitude are often considered, as they denote geographic location, which can influence weather patterns. Elevation is another crucial factor, as it affects temperature and atmospheric conditions, which, in turn, influence rainfall (Behzadi and Jalilzadeh, 2020). Variables like gradient, average temperature, and annual wind were selected due to their established associations with precipitation patterns in previous studies. In summary, our variable selection process was informed by existing scientific knowledge and a comprehensive pre-analysis to ensure that we included factors that are likely to influence rainfall within the Urmia Lake Basin.

The value of R2 indicates how much the dependent variable of precipitation can be explained by independent variables. According to the results, the value of R2 is equal to 60%, which is relatively

reasonable. The significance level (sig) in this problem is also less than 0.05, which indicates that the regression model can significantly predict the variation of the dependent variable.

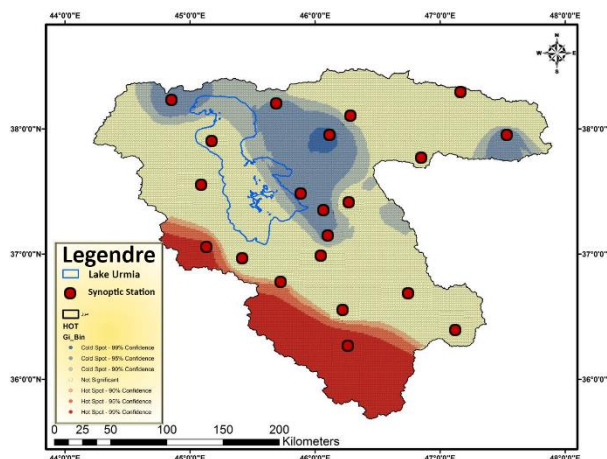


Figure 11: Hot Spot Analysis (Getis-Ord G_i^*) of rainfall data in Lake Urmia.

Slika 11: Analiza vročih točk (Getis-Ord G_i^*) podatkov o padavinah v jezeru Urmia.

As seen in Table 2, the constant value and all independent variables in the model are meaningful with concerning the Sig. value. With the help of the beta column, the relative contribution of each variable is obtained for predicting the dependent variable. Variables with the greatest effect on the dependent variable are determined. The latitude variable has the highest effect, and longitude and wind speed have the most negligible effect on rainfall in the Urmia Basin.

Table 2: Coefficients of regression model for rainfall prediction in the Urmia Lake Basin.

Preglednica 2: Koeficienti regresijskega modela za napovedovanje padavin v porečju jezera Urmia.

Model	Coefficients ^a					Collinearity Statistics	
	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Toleranc e	VIF
	B	Std Error	Beta				
1 (constant)	302.952	3.432		88.277	0.000		
X	-0.545	0.039	-0.072	-13.915	0.000	0.359	2.785
Y	-6.724	0.056	-0.801	-119.851	0.000	0.215	4.658
Elevation	0.002	0.000	0.147	43.564	0.000	0.840	1.191
Slope	-6.27E-8	0.000	-0.018	-5.404	0.000	0.825	1.212
Wind speed	0.674	0.105	0.047	6.447	0.000	0.177	6.642
Average temperature	-0.354	0.008	-0.145	-43.979	0.000	0.878	1.138

a. Dependent Variable: **Precipitation**

Figure 12 also shows an almost normal distribution of residues from the OLS model. This figure shows that the regression model can be applied to this problem.

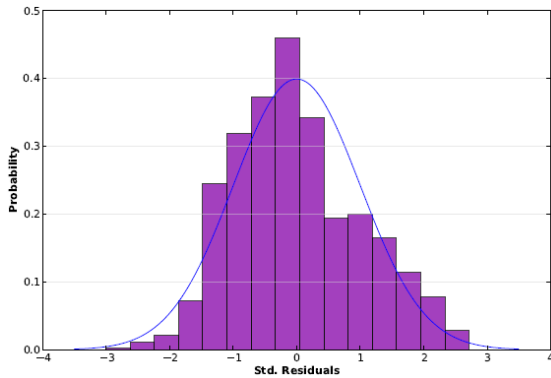


Figure 12: Histogram of residues obtained from the OLS model.

Slika 12: Histogram ostankov, dobljenih z modelom OLS.

In Table 2, column B shows the coefficients of the regression model; as a result, the regression equation for rainfall prediction in the Urmia Basin is:

$$\begin{aligned}
 \text{percipitation} = & 302.952 \\
 & - 0.545 \\
 & \times (\text{Longitude}) - 6.724 \\
 & \times (\text{Latitude}) \\
 & + 0.002 \times (\text{Height}) - 6.278 \\
 & \times 10^{-8} \times (\text{Slope}) \\
 & + 0.674 \times (\text{Wind}) - 0.354 \\
 & \times (\text{Temperature})
 \end{aligned}$$

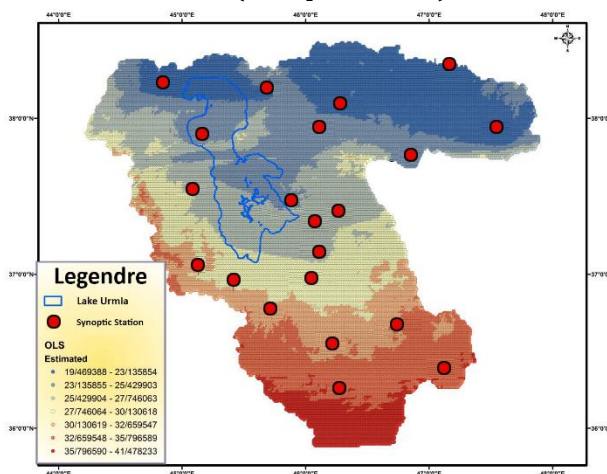


Figure 13: Estimated map with OLS model.

Slika 13: Karta ocenjena z modelom OSL.

The estimation map (rainfall estimation) in the Urmia Lake Basin by using the OLS method is shown in Figure 13. The model estimation is more or less than the actual in all regions. However, the estimate lower than the actual value is greater than the estimation higher than actual value.

5. Discussion

In the study, we explored 10 different traditional and geostatistical interpolation models to identify the optimal approach for precipitation model. The result showed that "Ordinary Kriging with Hole Effect kernel function" is the most accurate method for modeling precipitation in the study area. The amount of RMSE for this model was 4.15. Among the models, the "Universal Kriging with Semi-Variogram Circle kernel function" model had the lowest accuracy. The RMSE value for this model was 5.67. However, the accuracy obtained from these models is nearly identical; in other words, the range of RMSE changes is almost one unit. The close proximity among RMSEs shows that the precipitation model is independent of the mathematical model, and each of these is appropriate for producing a precipitation model. Therefore, the "Ordinary Kriging with Hole Effect kernel function" model was selected as the best for precipitation, and it was used to obtain the precipitation model in the study area.

In the second part, the relationship between the precipitation model and spatial phenomena such as longitude, latitude, height, slope, wind and temperature was investigated. The implementation results showed that the height and wind speed are directly related to the precipitation. The higher the height and wind speed, the higher the level of precipitation. On the other hand, the temperature is inversely related to precipitation. The higher the temperature, the less precipitation. Another result was that the slope did not have much effect on precipitation. In other words, the impact of slope on precipitation can be considered zero. More interestingly, latitude and longitude are the variables most influential for precipitation. Latitude was identified as the most significant factor in

precipitation. The dependency between latitude and precipitation is up to 80%.

The high dependency observed indicates that the precipitation model is influenced by various spatial variables. As latitude and longitude change, a multitude of spatial factors also undergo alterations, consequently impacting precipitation patterns.

The results of this study show that geostatistical methods have better accuracy than the traditional methods for modelling the rainfall in the Urmia Lake Basin. On the other hand, the dependent variables and environmental factors considered in the previous articles mainly included the elevation, latitude and longitude variables, while in this paper, factors such as slope, temperature, and wind speed are considered dependent variables along with the previous elements. As observed within the Urmia Lake Basin, our analysis indicates that latitude plays a prominent role in influencing rainfall patterns in this particular region. In addition, this paper introduces an innovative approach to modeling spatial relationships of rainfall data within the Urmia Lake Basin. While our primary emphasis has been on presenting the results of our analysis, we acknowledge that there is room for more in-depth statistical and analytical examination of the presented results to provide a deeper understanding of the patterns and relationships observed in the data.

The results can be summarized as: 1) Interpolation of rainfall data of Urmia Lake Basin and the choice of the ordinary Kriging method as the best interpolation method in this region; 2) Pattern and clustering analysis of precipitation data in Lake Urmia; 3) The analysis of precipitation data in Lake Urmia reveals distinct spatial clusters. Specifically, it highlights the south and southwest regions as high precipitation clusters (hot spots), signifying areas with the most rainfall in the Urmia Lake Basin. Conversely, it identifies the northern and central parts of the Urmia Lake Basin, along with two isolated locations at the Sarab and Salmas stations, as low precipitation clusters (cold spots), representing regions with the least amount of rainfall in this area. 4) Identification of the northern and central parts of the Urmia Lake Basin as well as two isolated spots in the Sarab and Salmas stations

as low-low clusters (cold spots) and least rainfall areas in this region; 5) Modeling the spatial relations of precipitation data in Lake Urmia by using regression; 6) Investigating the correlation between rainfall and explanatory variables such as latitude, longitude, elevation, slope, temperature, and wind speed on rainfall in the lake of Urmia as an independent variable; 7) Identification of "latitude" as the most effective dependent variable on precipitation in the Urmia Lake Basin; 8) Identification of longitude and wind speed as the variables with the least influence on precipitation in Lake Urmia Basin.

Most previous studies for estimating precipitation were based on statistical analyses, while in this study, spatial statistical methods were used for better and more accurate analysis of the subject of research. In this study, we compared our findings with previous research, focusing on a spatial perspective in climatology. Additionally, we aimed to investigate variables that have received less attention in prior studies.

For further evaluation, the results of this study were compared with other related research. Both the present study and Baykal et al. (2022) showed the same result. Although the study area of these two studies was different, Kriging and IDW are promising methods for estimating the amount of precipitation. This shows that the Kriging and IDW methods are independent of the region and can be used for any area and provide acceptable results.

The same issue is also observed in Longo-Minnolo et al. (2022) and in Jalilzadeh and Behzadi (2020). In Longo-Minnolo, interpolation methods were also used to compensate for the lack of ground data. However, Jalilzadeh and Behzadi emphasized that satellite images cannot compensate for this deficiency.

In Zou et al. (Zou et al., 2021), different Kriging interpolation methods were used to estimate the amount of precipitation. The input layers of the current study are almost similar to those of Zou et al., with latitude and longitude introduced as additional variables to the model. This addition highlighted the significance of latitude as a critical criterion in the estimation process, which was not

explored in Zou et al. This finding underscores the importance of considering latitude in precipitation estimation.

6. Conclusion

In this study, the rainfall in the Urmia Lake Basin was comprehensively analyzed using a variety of spatial data mining methods. First, the best interpolation of rainfall was selected among the various types of traditional and geospatial interpolation methods using cross-validation. Then, distribution, cluster, and patterns of rainfall were evaluated in the study area. Next, their spatial relationships were modeled using the least squares regression method. By identifying high and low clusters of rainfall in the study area, managers can make numerous plans for optimal water resource management.

Some key recommendations for future research include: 1) Exploring and incorporating additional dependent variables into the rainfall regression model to improve estimation accuracy; 2) Expanding data collection efforts by including more barometric and synoptic stations across the study area to enhance result precision; 3) Employing advanced mathematical transformations to mitigate data coherence and correlation issues, and considering the use of geographically weighted methods for a more precise modeling of spatial relationships in precipitation data as compared to the conventional OLS model.

References

- Abdollahi, A., Behzadi, S. (2022). Socio-Economic and Demographic Factors Associated with the Spatial Distribution of COVID-19 in Africa. *Journal of Racial and Ethnic Health Disparities*: 1-13.
- Abo-Monasar, A., Al-Zahrani, M. (2014). Estimation of rainfall distribution for the southwestern region of Saudi Arabia. *Hydrological Sciences Journal*, **59**(2): 420-431. <https://doi.org/10.1080/02626667.2013.872788>.
- Bajat, B. et al. (2013). Mapping average annual precipitation in Serbia (1961–1990) by using regression kriging. *Theoretical and applied climatology*, **112**(1-2): 1-13. <https://doi.org/10.1007/s00704-012-0702-2>.
- Baykal, T.M., Colak, H.E., Kılinc, C. (2022). Forecasting future climate boundary maps (2021–2060) using exponential smoothing method and GIS. *Science of the Total Environment*, **848**: 157633. <https://doi.org/10.1016/j.scitotenv.2022.157633>.
- Behzadi, S., Alesheikh, A.A. (2013). Introducing a novel model of belief–desire–intention agent for urban land use planning. *Engineering Applications of Artificial Intelligence*, **26**(9): 2028-2044. <https://doi.org/10.1016/j.engappai.2013.06.015>.
- Behzadi, S., Ali Alesheikh, A. (2013). Introducing AN Agent-Based Object Recognition Operator for Proximity Analysis. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* (3): 91-95. <https://doi.org/10.5194/isprsarchives-XL-1-W3-91-2013>.
- Behzadi, S., Jalilzadeh, A. (2020). Introducing a Novel Digital Elevation Model Using Artificial Neural Network Algorithm. *Civil Engineering Dimension*, **22**(2): 47-51. <https://doi.org/10.9744/ced.22.2.47-51>.
- Behzadi, S., Mousavi, Z. (2019). A novel agent-based model for forest fire prediction. *Earth Observation and Geomatics Engineering*, **3**(2): 51-63. <https://doi.org/10.22059/eoge.2020.283932.1051>.
- Benoit, L., Mariethoz, G. (2017). Generating synthetic rainfall with geostatistical simulations. *Wiley Interdisciplinary Reviews: Water*, **4**(2): e1199. <https://doi.org/10.1002/wat2.1199>.
- Bostan, P., Heuvelink, G.B., Akyurek, S. (2012). Comparison of regression and kriging techniques for mapping the average annual precipitation of Turkey. *International Journal of Applied Earth Observation and Geoinformation*, **19**: 115-126. <https://doi.org/10.1016/j.jag.2012.04.010>.
- Cheng, K.S., Lin, Y.C., Liou, J.J. (2008). Rain-gauge network evaluation and augmentation using geostatistics. *Hydrological Processes*, **22**(14): 2554-2564. <https://doi.org/10.1002/hyp.6851>.
- Cristiano, E., ten Veldhuis, M.-c., Van De Giesen, N. (2017). Spatial and temporal variability of rainfall and their effects on hydrological response in urban areas—a review. *Hydrology and Earth System Sciences*, **21**(7): 3859-3878. <https://doi.org/10.5194/hess-21-3859-2017>.
- Eivazi, M., Mosaedi, A. (2012). An Investigation on Spatial Pattern of Annual Precipitation in Golestan

- Province by Using Deterministic and Geostatistics Models. 106131. <https://doi.org/10.1016/j.atmosres.2022.106131>.
- Fernández-Delgado, M. et al. (2019). An extensive experimental survey of regression methods. *Neural Networks*, **111**: 11-34. <https://doi.org/10.1016/j.neunet.2018.12.010>.
- Ghashghaie, S., Behzadi, S. (2019.) Spatial Statistics Analysis to Identify Hot Spots Using Accidental Event Calls Services. *Journal of Statistical Research of Iran JSRI*, **16(1)**: 121-141.
- Huang, F.L. (2018). Multilevel modeling and ordinary least squares regression: how comparable are they? *The Journal of Experimental Education*, **86(2)**: 265-281. <https://doi.org/10.1080/00220973.2016.1277339>.
- Jafarian, H., Behzadi, S. (2020). Evaluation of PM2. 5 emissions in Tehran by means of remote sensing and regression models. *Pollution*, **6(3)**: 521-529.
- Jalilzadeh, A., Behzadi, S. (2019). Machine learning method for predicting the depth of shallow lakes using multi-band remote sensing images. *Journal of Soft Computing in Civil Engineering*, **3(2)**: 54-64. <https://doi.org/10.22115/SCCE.2019.196533.1119>.
- Jalilzadeh, A., Behzadi, S. (2020). Flood Mapping and Estimation of Flood Water-Level Using Fuzzy Method and Remote Sensing Imagery (Case Study: Golestan Province, Iran), *Forum Geografic. University of Craiova, Department of Geography*, pp. 165.
- Karayusufoglu, S., Eris, E., Coskun, H.G. (2010). Estimation of basin parameters and precipitation distribution of solalki basin, turkey. *Wseas transactions on environment and development* (**5**): 6.
- Kumar, A., Dhakhwa, S., Dikshit, A.K. (2022). Comparative Evaluation of Fitness of Interpolation Techniques of ArcGIS Using Leave-One-Out Scheme for Air Quality Mapping. *Journal of Geovisualization and Spatial Analysis*, **6(1)**: 1-11. <https://doi.org/10.1007/s41651-022-00102-4>.
- Liu, X. et al. (2022). Molecular-level understanding of metal ion retention in clay-rich materials. *Nature Reviews Earth & Environment*, **3(7)**: 461-476. <https://doi.org/10.1038/s43017-022-00301-z>.
- Longo-Minnolo, G., Vanella, D., Consoli, S., Pappalardo, S., Ramírez-Cuesta, J.M. (2022). Assessing the use of ERA5-Land reanalysis and spatial interpolation methods for retrieving precipitation estimates at basin scale. *Atmospheric Research*, **271**: 106131.
- Mahjoobi, M., Behzadi, S. (2022). Solar desalination site selection on the Caspian Sea coast using AHP and fuzzy logic methods. *Modeling Earth Systems and Environment*: 1-9. <https://doi.org/10.1007/s40808-022-01418-2>.
- Maris, F., Kitikidou, K., Angelidis, P., Potouridis, S. (2013). Kriging interpolation method for estimation of continuous spatial distribution of precipitation in Cyprus. *Current Journal of Applied Science and Technology*: 1286-1300.
- Mirzaei, R., Sakizadeh, M. (2016). Comparison of interpolation methods for the estimation of groundwater contamination in Andimeshk-Shush Plain, Southwest of Iran. *Environmental Science and Pollution Research*, **23(3)**: 2758-2769.
- Moral, F.J. (2010). Comparison of different geostatistical approaches to map climate variables: application to precipitation. *International Journal of Climatology*, **30(4)**: 620-631. <https://doi.org/10.1002/joc.1913>
- Njeban, H.S. (2018). Comparison and evaluation of GIS-based spatial interpolation methods for estimation groundwater level in AL-Salman District—Southwest Iraq. *Journal of Geographic Information System*, **10(4)**: 362.
- Papalexioiu, S.M., AghaKouchak, A., Foufoula-Georgiou, E. (2018). A diagnostic framework for understanding climatology of tails of hourly precipitation extremes in the United States. *Water Resources Research*, **54(9)**: 6725-6738. <https://doi.org/10.1029/2018WR022732>.
- Sun, Y., Kang, S., Li, F., Zhang, L. (2009). Comparison of interpolation methods for depth to groundwater and its temporal and spatial variations in the Minqin oasis of northwest China. *Environmental Modelling & Software*, **24(10)**: 1163-1170. <https://doi.org/10.1016/j.envsoft.2009.03.009>.
- Triguero, I. et al. (2019). Transforming big data into smart data: An insight on the use of the k-nearest neighbors algorithm to obtain quality data. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, **9(2)**: e1289. <https://doi.org/10.1002/widm.1289>.
- Yang, W. et al. (2020). Using principal components analysis and IDW interpolation to determine spatial and temporal changes of surface water quality of Xin'anjiang

river in Huangshan, China. *International journal of environmental research and public health*, **17**(8): 2942.

Zhong, Q., Tong, D., Crosson, C., Zhang, Y. (2022). A GIS-based approach to assessing the capacity of rainwater harvesting for addressing outdoor irrigation. *Landscape and Urban Planning*, **223**: 104416.

Zou, W.-y., Yin, S.-q., Wang, W.-t. (2021). Spatial interpolation of the extreme hourly precipitation at different return levels in the Haihe River basin. *Journal of Hydrology*, **598**: 126273.
<https://doi.org/10.1016/j.jhydrol.2021.126273>.