RAINFALL TIME SERIES ANALYSIS IN SEMIARID REGION OF DUHOK GOVERNORATE, IRAQI KURDISTAN REGION

Marwan B. I.Govay[1] Tariq Hama Karim Kaka Hama [2]

Lecturer Prof.

Dept. of Surveying and Geomatics Engin., Faculty of Engin., Tishk Intern. University Erbil, Kurdistan Region

E-mail: marwan.basheer@uod.ac

ABSTRACT

This study was aimed to investigate data homogeneity, detect trend and changes in historical rainfall and forecast future rainfall pattern. Rainfall historical records taken from 22 gauging and non-gauging stations distributed all over Duhok governorate. The dataset includes long time series of monthly and annual rainfall totals (mm) from 1998 to 2020. To detect data variability, four different homogeneity tests were used. To analyze trends and sequential shifts in historical data, parametric and non-parametric tests were applied. Theil-Sen's slope estimator test was applied to calculate the magnitude of change over time. Rainfall forecasting was based on the Box-Jenkins methodology. The homogeneity test results revealed that the majority of the monthly and annual rainfall series were labeled as useful. Furthermore, the annual rainfall at most of the study stations presented positive trend with sen's slope ranged between 0.545 -43.03 mm yr^{-1}. At five stations tended to beyond the upper limit of the CUSUM chart during 2019 and 2020. Conversely, it tended to be below the lower limit at 3 stations during the period from 2000 t0 2002. ARIMA (0, 1,1) was the best model for predicting yearly rainfall for 75% of the stations, while the model (1,1,1)(1, 0, 1)$_{12}$ was the best suited model for predicting and forecasting monthly rainfall for more than 83% of the selected stations.

Keywords: test of data homogeneity, trend analysis, ARIMA Model, rainfall forecasting, drought

تحليل السلاسل الزمنية لأمطار في المنطقة شبه الجافة بمحافظة دهوك ، إقليم كردستان العراق

مروان بشير اسماعيل[1] طارق حمه كريم كاكه حمه[2]

مدرس أستاذ

قسم هندسة المساحة والجيوماتكس ،جامعة تيشك الدولية ،أربيل ،إقليم كوردستان.

المستخلص

هدف الدراسة اختبار تجانس البيانات, اكتشاف الاتجاه والتغيرات في بيانات الأمطار المسجلة سابقا والتنبؤ بنمط هطول الأمطار في المستقبل. المعدلات المسجلة لهطول الأمطارالتي تنتمي إلى 22 محطة قياسية وغير قياسية موزعة في كافة أنحاء محافظة دهوك، إقليم كردستان العراق. تتضمن مجموعة البيانات سلسلة زمنية طويلة لمجموع الأمطار الشهرية والسنوية (مم) من 1998 إلى 2020. للكشف عن تباين البيانات، تم استعمال أربعة اختبارات تجانس مختلفة. لتحليل الاتجاهات والتحولات المتسلسلة في البيانات المسجلة سابقا، تم تطبيق الاختبارات المعلمية (البارامتري) واللامعلمية (اللابارامتري). تم تطبيق اختبار Theil-Sen'sلحساب حجم التغيير مع الزمن. استند التنبؤ بهطول الأمطار على منهجية Box-Jenkins .أظهرت نتائج اختبار التجانس أن غالبية سلاسل هطول الأمطار الشهرية والسنوية بأنها مفيدة. علاوة على ذلك، أظهر هطول الأمطار السنوية في معظم محطات الدراسة اتجاهاً إيجابياً مع الانحدار وتراوح بين 0.545 -43.03 ملم في العام. خلال عامي 2019 و2020 كانت هناك خمس محطات تميل إلى تجاوز الحد الأعلى لمخطط CUSUM وعلى العكس من ذلك، وجد بان هناك 3 محطات تميل إلى أقل من الحد الأدنى خلال الفترة من 2000 إلى 2002. وفقًا لنتائج هذه الدراسة، وجد بان أفضل نموذج للتنبؤ بهطول الأمطار السنوية لـ 75٪ من المحطات هو(0,1,1) ARIMA، بينما كان النموذج (1,1,1) (1,0,1) $_{12}$ هو أنسب نموذج للتنبؤ بهطول الأمطار الشهرية لأكثر من 83٪ من المحطات المختارة.

كلمات المفتاحية: اختبار تجانس البيانات، تحليل الاتجاه ،نموذج ARIMA، التنبؤ بهطول الأمطار، جفاف

Received:15 /7/2022, Accepted:27/11/2022

INTRODUCTION

Climate change has the ability to affect all natural systems., thus becoming a threat to human development and survival economically, socially and politically (27). Droughts are caused by high temperatures and insufficient precipitation, posing major concerns to food security (21). Precipitation regimes that change with global warming and climate changes affect the countries in environmental, economic, and social dimensions (10). It has been shown that Iraq has been experiencing rainfall deficits for more than a decade, with a significant decline in the amount of rainfall recorded, and rainfall prediction has been the focus of multiple studies (5, 35). Besides limited water resources, the water quality is deteriorating due to urban expansion. As a result, it is essential to know the future water resources budget so as to assist decision makers improve their decisions by taking into consideration the available and future water resources (8). Analyzing the rainfall process is vital for the resolving numerous regional environmental problems related with integrated water resource management at the regional level, with suggestions for agriculture, natural threats such as drought and floods, and climate change. (1). Rainfall time series are usually characterized by complicated variability, and the effects of autocorrelation and seasonality can be easily misinterpreted with changes in the mean or variance. Such changes may take the form of trends over time or be more abrupt, and a revealing technique must be accurately considered to account for these mechanisms (15). (6) found that the accuracy, reliability drought and flood modeling and water resources planning models differed due to the quality of the data used. (14) have shown that the historical rainfall records may be suffered from non-climatic factors that give rise to inhomogeneity of such records. Non-homogeneous climatological time series can lead to inconsistent conclusions (34). Factors like missing values, seasonal fluctuations and lack homogeneity are responsible for complicating studies of hydrological change. There are also additional problems like censored data and data series that are not sufficiently long (19). Data recorded must be examined and checked for homogeneity before their use in the research studies. Each subsequent analysis is influenced by the elimination of false discovered inhomogeneities and the acceptance of inhomogeneous series. As a result, it is critical that homogenization methods are carefully followed (33). Time series forecasting approaches are typically based on previous data analysis. It is assumed that historical data patterns can be used to foretell future events. The Box-Jenkins methodology uses a three-step iterative methodology of model identification, parameter estimation, and diagnostic checking to identify the best precise model from a general category of ARIMA models (9). This three-steps procedure were repeated until the suitable model is obtained. The model could be used to forecast future time series values. According to (5), the first step before applying the Box- Jenkins model is to determine even if the historical data is stationary and whether there is any signification seasonality that require to be modeled. Forecasting and time series analysis are the two-essential element of Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF). One of the most significant natural phenomena in the studied area is rainfall. (22) Over this region, agriculture production is highly affected by the rainfall variability and intensity. Consequently, an initial warning of probable rainfall can help to solve numerous problems related to agriculture, climate change and natural risks like flood and drought. Fitting Regression Models (RM) to time series with trend and seasonality components is also part of the forecasting approach (26). Autoregressive moving average ARMA (p,q) is a model which combines p autoregressive terms with q moving average terms. If the object series is differenced d times to attain stationarity, the model is categorized as ARIMA (p,d,q) where I denotes "Integrated" (23). In addition to the non-seasonal ARIMA (p,d,q), one can recognize Seasonal ARIMA model (P,D,Q) parameters for the time series data known as SARIMA model. The parameters are seasonal autoregressive (P), seasonal differencing (D) and seasonal moving average (Q). The best fit of the historical data can be established and forecasting could be used by using the Box-Jenkins methodology, which applies ARMA, ARIMA or SARIMA. The methodology includes four stages: estimation of model parameters, model identification, forecasting and diagnostic checking (31). Momani and Naill (24) concluded that ARIMA (1, 0, 0)(0, 1, 1)$_{12}$ model identified formerly is convenient to represent their data and could be used to forecast the forthcoming rainfall data for Amman airport station for the period from 1922-1999. On the other hand, Ali (5) was tested several ARIMA models and observed that the seasonal ARIMA model of the orders SARIMA (2,1,3)(0,1,1) is the best suite model for predicting monthly rainfall data of Baghdad meteorological station. Considerable fluctuation, very slight increasing trend and significant seasonality were also noticed. Additionally, (10) forecasted rainfall for 9 cities in Turkey's Marmara region using ARIMA, ARMA, and SARIMA models, and chose the ARIMA model as the best fit with the lowest prediction error. Most people believe that the amount of rainfall had a tendency to decrease over the second half of the 20th century (16). Accordingly, Investigating the influence of predicted climate change on rainfall across the research area is beneficial. As there is an abrupt change in water management strategies in most of the countries which are facing water shortage, there is a necessity to carry out studies on trend analysis and forecasting of precipitation (3). A few researches of studies on rainfall time series analysis has been performed over Iraq, particularly over Iraqi Kurdistan Region. This has been linked to the lack of historical rainfall data in addition to the unreachability and scarcity of well-distributed meteorological stations throughout the study area. This study was aimed to investigate data homogeneity characteristics of rainfall time series collected at **22** meteorological stations; analyze the rainfall trend in the existing meteorological stations using parametric and non-parametric tests; to detect abrupt changes in rainfall pattern; and to predict the future rainfall pattern using time series models in the Box-Jenkins method.

MATERIAL AND METHODS

The study area is located at Duhok governorate, Iraqi Kurdistan Region. It has a spread of 10955.92 km^2 and lies between 36^o 18$''$ 42.64$'$ and 37^o 20$''$ 33.55$'$ N latitudes and 42^o 20$''$ 25.36$'$ and 44^{o}17″40.5 ′ E longitudes (**Fig.1**) the study area has an altitude range of 244 to 2551 meters. Historical records of rainfall belong to 22 gauging and non-gauging stations distributed all over the study area. The data set includes long time series of monthly and annual rainfall totals (mm) covered the period 1998–2020. They were provided from the Directorate of Duhok meteorology, Ministry of Agriculture and Water Resources, and the General Director of Meteorology and Seismology. In addition, the climate is characterized by the occurrence of a distinct dry season of about 5 months extending from June to October and by recurrent droughts. The rainy season is mostly between November and April. Descriptive statistics of annual rainfall records of the study region is presented in Table1. Overall, the climate at this region is Mediterranean, with rainy and cold winters and dry and warm summers. The average yearly temperature is expected to be 19 degrees Celsius. January and August are the coldest and hottest months of the year. Based on the Koppen climatic classification system, the top part of the investigated site along the Iraqi-Turkish border can be classified as type DSa, suggesting a cool wet climate in the winter and a dry season in the summer with a yearly rainfall of 500 to 800 mm. On the other hand, the majority of area in the middle and lower parts of the study region is classified as a temperate, dry summer, hot summer (Csa) according to the aforementioned scheme Adamo et al. (12). To discover the variability of the data and show whether the historical rainfall records are suffered from non-climatic factors, four homogeneity tests were employed at each station. The tests encompassed standard normal homogeneity test (SNHT), Buishand range (BR) test, Pettitt test (PT), and von Neumann ratio (VNR). Simple linear regression was used as a parametric test for detecting trends when the regression assumptions were met. In addition, the Mann-Kendall (MK) test and the Cumulative Sum (CUSUM) chart were used to detect trend and sequential shifts in time series of rainfall parameters.

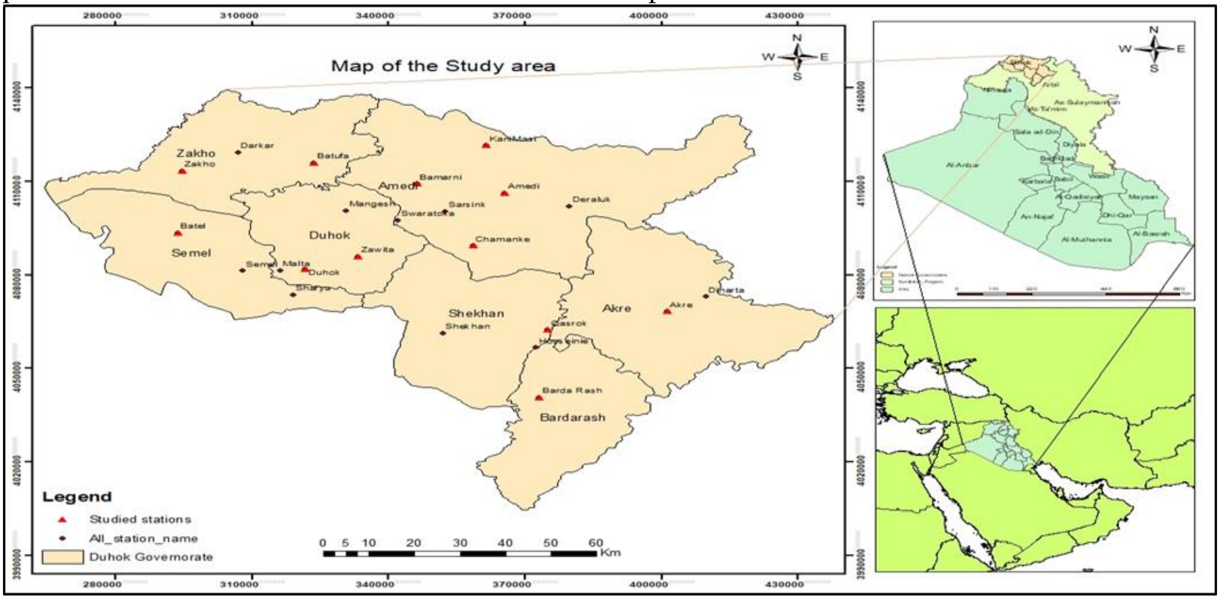

Fig.1. Location map showing the distribution of the meteorological stations map over the study area

The Mann-Kendall (MK) test, a nonparametric rank-based test, was used to analyze trend in historical rainfall data. It has an advantage over other tests as it is distribution free and robust against outliers Hess et al. (**18**). Because the presence of serial correlation can increase or decrease the likelihood of identifying significant trends, the test was done to uncorrelated data (17). When serial correlation was present, pre-whitening was an alternative test for detecting a trend in a time series (11). Sen's slope approach was also used to determine the magnitude of the trend line. The Box-Jenkins technique was utilized to create an Auto regressive Moving Average (ARMA) model for rainfall data taken from existing meteorological stations. The historical data at a given station were examined for stationarity and seasonality at the start of modeling. The time series' temporal correlation structure was detected using autocorrelation (ACF) and partial autocorrelation (PACF) functions. Augmented Dickey-Fuller Unit Root test was also conducted to confirm non-stationarity of the data using EViews software version. Differences of lag K=1 and 12 were taken to remove seasonality and nonstationarity for monthly and annual data respectively. The ACF and PACF were reexamined after differencing to check for non-stationarity and seasonality. When the initial differencing did not remove the non-stationarity, the procedure was repeated with the second differencing. Several trials were made, in which several P and Q ranging from zero to two were examined to determine the best ARIMA model from the candidate models. The model which gave the best combination of minimum RMSE, maximum R-squared, and least AIC was selected as the best fit model. To determine the model's form, autocorrelation function (ACF) and partial autocorrelation function (PACF) plots were used. The guessed model's parameters were then estimated using maximum likelihood estimation, which is a method for calculating the parameters that maximize the likelihood of observations. During the diagnostic checking, the residuals from the fitted models were checked for adequacy. It was accomplished by correlation analysis using the residual ACF/PACF functions and the goodness-of-fit test using the Ljung-Box test. When residuals were correlated, then the model was dropped and a new trial was made. Histograms, the Q-Q plot, the residuals histogram, and the residuals plot against fitted values were also employed to assess the model's adequacy. After identifying the best suitable model from the historical data, forecasting was done for the next incoming five years using IBM SPSS version 22. The suitability of the models was confirmed by investigating a set of performance indicators such as R^2, MAE, and MAPE.

RESULTS AND DISCUSSION

Table 1 displays the summary of statistics of annual rainfall recorded during the period from 1980 to 2020 at district level in Duhok governorate. The mean annual rainfall in the study area varied from as low as 415.54 at Bardarash to as high as 874.73 mm at Chamanke. The coefficient of variability the CV value varies from 26.77% to 42.42% in all stations of Duhok governorate, with a maximum at Akre and a minimum at Swaratoka, indicating high spatial variability of rainfall in the studied area. This result is concordant with the finding of (36), they observed that precipitation fluctuate greatly with wide variation in Sinjar area, which is close to the study area. Ninteen out of 22

Description of the datasets

(86.36%) stations fell in the high variable class, while 4.54% of the stations fell in the very high class and the rest fell in the moderate variability class. According to the literature, CV is used to classify the degree of variability as less ($CV < 20\%$), moderate ($20 < CV < 30\%$), high ($CV > 30\%$), very high ($CV > 40\%$) and $CV > 70\%$ indicating exceptionally high inter-annual variability of rainfall (4). A variability analysis of rainfall and other characteristics is critical for policymakers in making decisions since rainfall is a fundamental factor in determining how much water is available in a given region, (29).

Table 1. Descriptive statistics of annual precipitation recorded at the stations during 1998 – 2020.

Rainfall Station	Observations	Minimum (mm)	Maximum (mm)	Mean (mm)	Std. deviation	Skewness	Kurtosis
Duhok	23	251.74	1001.20	558.43	197.25	0.690	-0.058
Zakho	23	299.45	955.80	582.19	169.07	0.204	-0.303
Zawita	23	312.64	1284.70	783.91	259.25	0.184	-0.860
Akre	23	328.80	1044.90	661.23	192.18	0.216	-0.229
Mangeshk	23	228.86	1660.70	702.04	297.78	1.490	3.939
Bamarni	23	326.11	1259.00	787.87	248.61	0.092	-0.508
Amedi	23	418.90	1282.20	807.91	234.62	0.473	-0.506
Semel	23	190.09	841.10	453.81	160.87	0.616	0.048
Malta	23	245.24	1037.40	500.50	180.74	1.170	2.154
Batufa	23	303.04	1695.50	725.64	282.80	1.716	5.559
KaniMasi	23	214.88	1397.50	786.95	291.92	-0.073	0.152
Batel	22	254.46	890.40	462.02	150.29	1.313	2.245
Darkar	22	266.47	901.20	535.65	175.38	0.519	-0.593
Deraluk	22	296.58	1189.54	763.60	235.45	-0.048	-0.522
Sarsing	22	199.65	1390.90	831.49	283.46	-0.195	0.136
Bardarash	22	197.88	889.50	415.54	156.45	1.560	3.059
Qasrok	22	243.64	998.60	521.55	187.67	0.901	0.773
Swaratoka	20	328	1018	673.97	180.43	0.157	0.196
Hosseinie	20	258	985	542.13	181.39	0.964	1.038
Dinarta	20	373	1351	808.23	265.38	0.176	-0.674
Chamanke	23	266.70	1580.00	874.73	355.59	0.250	-0.463
Shekhan	16	320	1061	604.61	221.80	0.943	0.123

Table (10 shows that, with a few exceptions, the annual rainfall time series recorded at the study stations are positively skewed, i.e., skewed to the right. Similarly, it was noticed that the majority of the calculated kurtosis values are positive. Additionally, it was observed that most of the skewness values are sandwiched between -1 and +1 Based on the values of skewness, it could be concluded that the data belonging to different stations are normally distributed. The data with a range of -1 to +1 skewness were considered as normally distributed data (13); (30).

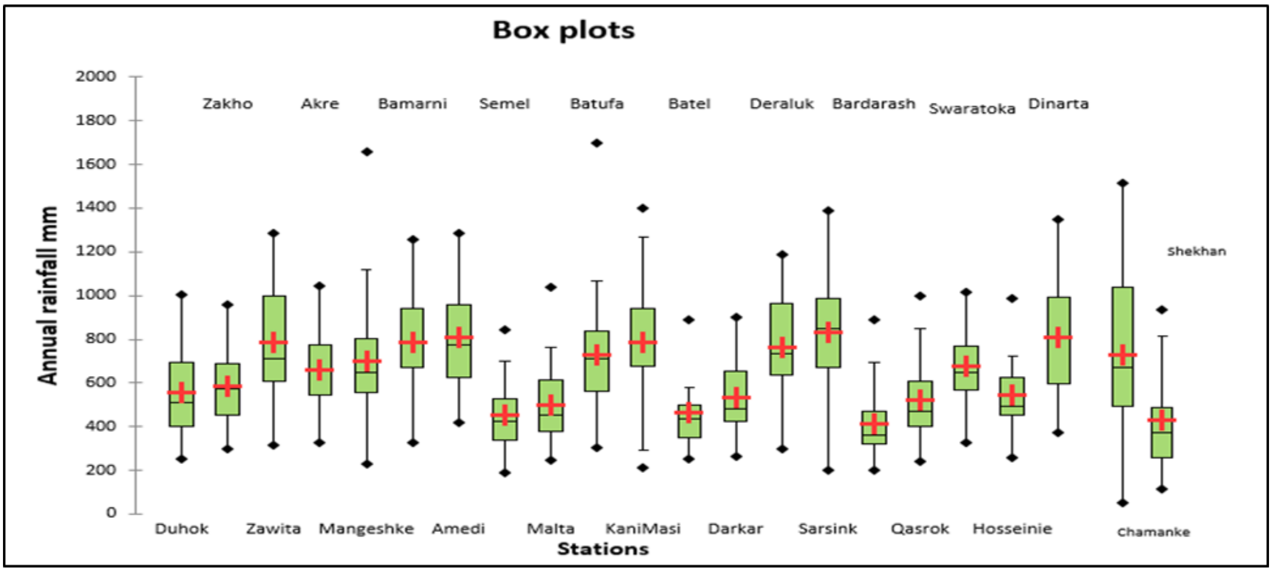

Fig.2. Box and whisker plots for 22 rainfall stations

The annual rainfall data have been revealed using box and whisker plots to further confirm the normality of the data sets, as demonstrated in **Fig.(2**) that the median line did not shift from the center of the box for the majority of the stations, showing small deviations from normality. The dataset is obviously not normally distributed, with the majority of the data falling in the upper whisker, i.e., in the fourth quartile.

Homogeneity Tests

The findings of homogeneity tests conducted on annual and monthly rainfall data records from various districts within the Duhok governorate are show in Table 2. The standard normal homogeneity test (SNHT), Pettitt's test, the Buishand range (BR) test, and the von Neumann ratio (VNR) test were all homogeneity tests. These tests were assessed at 5% significant level. The generated outputs were classified into three classes. The classification was defined as ‘useful’ when one or none of the tests reject the null hypothesis; ‘doubtful’ when the series reject two null hypotheses out of four tests; ‘suspect’ when 3 or 4 tests are rejected. The homogeneity test results of monthly and annual rainfall time series for the selected 22 stations show that 100% of stations may be classified as 'useful' for the January months, April, and October, which represent the winter, spring, and autumn seasons, respectively. Conversely, about 55% of the stations could be assigned to useful class for July as a representative of the summer season. However, the analysis of this month can be dropped from this study because, it is not a rainy month in the area under study. The overall results of yearly time series indicated that there were 17 out of 22 annual series (77.27%) were “useful”, 3 series (13.04%) were labeled as “doubtful” and the remaining (9.09%) were found to be “suspect”. Among all the testing variables, monthly rainfall data showed the highest percentage of homogeneity compared with the annual data. It is also is evident from the above results that most of the rainfall data in the existing rainfall stations of the Duhok governorate are homogeneous with only small percentage of break detected in monthly and annual time series. While the doubtful time series require careful consideration because there is an indication of inhomogeneity within the series, the useful time series can be used for more hydrological modeling. The “suspect” series should be strictly discarded from further analysis because the level of in homogeneity has exceed the normal limit and therefore is not suitable for any analysis (20). These findings are congruent with those of (2) they investigated the homogeneity of the annual and seasonal precipitation data throughout the north of Iraq and observed that stations 2, 3, and 1 out of 9 were assessed as doubtful for annual, winter, and spring precipitation data, respectively. They also highlighted that homogeneity tests appear as a useful tool to control the data reliability and quality before a study in water resources, hydrological processes, and climate change fields. It is commendable to mention that those series that contain in homogeneity should be examined properly before applying them in any hydrological application. Further work needs to be done to investigate the possible causes of the in homogeneity of rainfall data at a few stations in the region under study and subsequently, suitable adjustment techniques can be used to improve the quality of hydrological time series.

Table 2. Results of homogeneity tests for the annual rainfall recorded at the study stations

#	Station	N	Pettitt's test			SNHT			Buishand's test			VNR test			Classification
			K_N	P-value	K_N-critical	To	P-value	T-critical	Q	P-value	Q-critical	N	P-value	N-critical	
1	Duhok	23	70	0.203	72	9.364	0.020	7.16	5.222	0.113	1.451	1.336	0.053	1.335	Useful
2	Zakho	23	76	0.113	72	6.104	0.091	7.16	5.481	0.081	1.451	1.703	0.234	1.335	Useful
3	Zawita	23	80	0.078	72	7.208	0.042	7.16	6.167	0.038	1.451	1.277	0.034	1.335	Suspect
4	Akrah	23	56	0.559	72	7.236	0.049	7.16	4.443	0.241	1.451	1.639	0.187	1.335	Useful
5	Mangeshk	23	68	0.232	72	10.05	0.100	7.16	5.235	0.089	1.451	1.143	0.014	1.335	Useful
6	Bamarni	23	66	0.278	72	8.343	0.018	7.16	5.053	0.132	1.451	1.219	0.024	1.335	Doubtful
7	Amedi	23	76	0.122	72	6.282	0.094	7.16	6.093	0.040	1.451	1.335	0.048	1.335	Doubtful
8	Semel	23	54	0.638	72	8.643	0.017	7.16	4.855	0.164	1.451	1.532	0.127	1.335	Useful
9	Malta	23	48	0.875	72	8.484	0.108	7.16	4.025	0.346	1.451	1.464	0.088	1.335	Useful
10	Batufa	23	68	0.238	72	8.440	0.109	7.16	4.682	0.163	1.451	1.565	0.125	1.335	Useful
11	KaniMasi	23	72	0.172	72	9.479	0.005	7.16	5.671	0.060	1.451	1.350	0.050	1.335	Useful
12	Batel	22	45	0.932	67	8.511	0.049	7.09	2.990	0.660	1.444	1.586	0.152	1.324	Useful
13	Darkar	22	48	0.795	67	4.551	0.241	7.09	3.332	0.533	1.444	1.314	0.047	1.324	Useful
14	Deraluk	22	55	0.495	67	8.186	0.016	7.09	4.784	0.149	1.444	1.349	0.054	1.324	Useful
15	Sarsing	22	47	0.822	67	7.652	0.029	7.09	4.194	0.264	1.444	1.301	0.044	1.324	Doubtful
16	Bardarash	22	42	0.912	67	9.615	0.083	7.09	3.156	0.597	1.444	1.284	0.042	1.324	Useful
17	Qasrok	22	38	0.675	67	6.770	0.139	7.09	3.00	0.662	1.444	1.539	0.118	1.324	Useful
18	Swaratoka	22	34	0.777	67	4.013	0.267	7.09	2.758	0.694	1.444	1.630	0.202	1.324	Useful
19	Hosseinie	22	30	0.521	67	4.178	0.316	7.09	2.064	0.931	1.444	1.608	0.182	1.324	Useful
20	Dinarta	20	31	0.565	57	4.397	0.240	6.95	2.707	0.719	1.43	1.323	0.061	1.3	Useful
21	Chamanke	15	110	0.002	32	10.87	0.003	6.6	8.011	0.001	1.395	0.820	0.001	1.24	Suspect
22	Shekhan	23	35	0.400	72	6.361	0.067	6.6	4.067	0.138	1.451	1.729	0.284	1.335	Useful

Identification of Monotonic Trend in Annual Rainfall Time Series.

Table (3) shows the outcome of rank-based tests, namely Mann–Kendall (MK) and Spearman rank correlation (SRC) along with the results of slope-based tests, namely least squares linear regression (LR) and Sen's slope estimator. These tests were conducted at 5% significance level for detecting trends in rainfall time series data on annual basis. Annual data for (22) years spanning from 1998 to 2020 was used as input parameters for this analysis. Due to space limitation, these tests were conducted on 12 stations distributed over the whole area of Duhok governorate. Based on MK test, only four stations, namely, Duhok, Akre, Batel, Bardarash and Qasrok exhibited no trend. On the contrary, the remaining stations presented positive trends which are significant 5% level of significance. Most of these stations are situated to the north of the study area.

Table 3. Trend analysis of annual rainfall time series recorded at some selected meteorological stations over the study area

#	Station	Trend analysis indicators		Trend	Spearman's rho Test Statistic, ZSP	Regression line slope (b) mm year^{-1}
		Mann-Kendall- P value	Sen's slope			
1	Duhok	0.065	12.293	NT	1.182	12.938
2	Zakho	0.026	11.810	IT	1.302	11.505
3	Zawita	0.014	21.611	IT	1.502	21.061
4	Akre	0.092	11.382	NT	1.027	10.519
5	Bamarne	0.034	18.012	IT	1.302	17.301
6	Amedi	0.012	18.382	IT	1.451	18.269
7	Batufa	0.044	17.020	IT	1.239	20.724
8	KaniMasi	0.010	23.131	IT	1.565	25.368
9	Batel	0.956	0.545	NT	0.017	4.014
10	Bardarash	0.504	3.932	NT	0.429	5.828
11	Qasrok	0.540	5.162	NT	0.461	7.196
12	Chamanke	< 0.0001	43.027	IT	2.289	42.144

The Sen's slope ranged from as low as 0.545 mm yr^{-1} at Batel site to as high as 43.03 mm yr^{-1} at Chamanke station. It is evident from the above results there is indication of increasing the magnitude of slope with distance away from south to the north of the study region. The positive trend may due to the high annual rainfall during the last year of the study span. The Mann-Kendall test has the advantage of being applicable to data with outliers because its statistic is based on the sign of differences rather than the values of the variable (28). It is also obvious from Table (3) that both Linear LR and Sen's slope methods were comparable in estimating the magnitude of trend. The reason behind the consistency of the results of these two methods may be due the slight deviation of the rainfall data series from normality. The LR-method is a parametric test, whereas the Sen's slope method is a non-parametric test. Furthermore, the results of Table (3) highlighted that Both MK and SR methods are comparable in detecting positive trend current study. Because both the MK and SR methods failed to account for the influence of serial correlation in a time series, a discrepancy with other techniques for detecting trends that are not mentioned here is to be expected. It is worth mentioning that when the time series is serially or auto correlated, both Mann-Kendall and linear regression techniques perform poorly (32). Therefore, they should be used with cautions for trend detection.

CUSUM Analysis

Figure (3) displays a graphical representation of the CUSUM analysis. The results of the analysis presented that potential abrupt shifts appear to have taken place in 2019 and 2020 with respect to the annual rainfall time series recorded at some meteorological stations over the study area. A segment with an ascending trend in the CUSUM chart signifies a period of time where the value is above the total average and vice versa. The stations at which abrupt changes have occurred encompass Duhok, Zawita, Batufa, Amedi and Chamanke. The type of change is characterized by ascending trends in the indicated periods at the abovementioned stations. On the other hand, it was observed that the chart line gradually goes downward and intersect the lower boundary at three stations, namely, Zawita at 2001 and Bamarne at 2001 and 2002 and KaniMasi at 2000 through 2002. indicating the values tend to be beyond the limit of Lower CUSUM (-4σ). Unlike the chart lines for the aforementioned stations, the chart lines for the remaining stations do not intersect the lower and the upper CUSUM limits. This implies that all points fall within the region bounded by the lower and the upper CUSUM limits. Usually, a CUSUM chart shows an out-of-control activity when a process intersects a boundary by an ascending or descending drift of the cumulative sum. The control limits for the individual's chart are $\pm 4\sigma$.

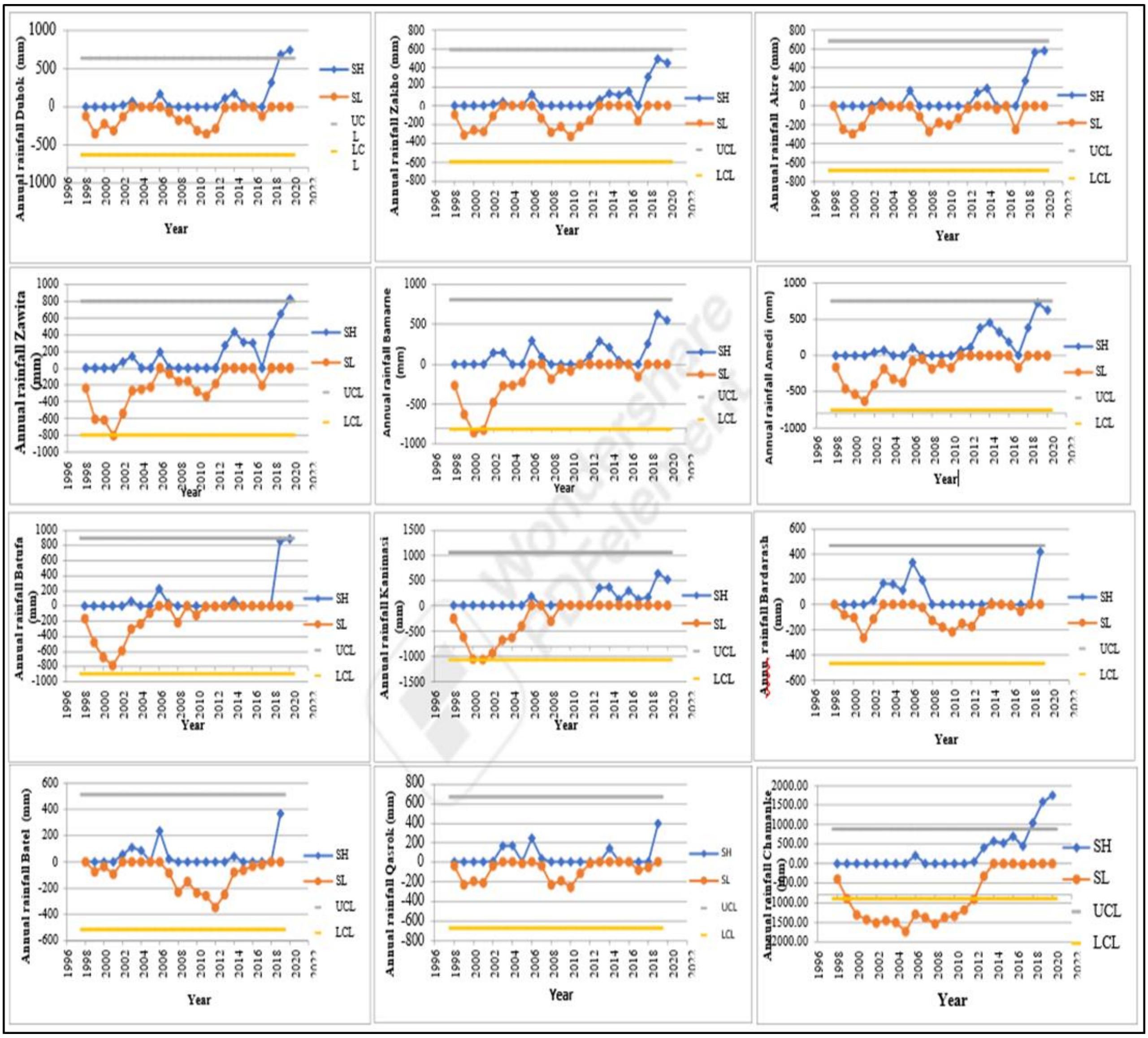

Fig.3. Charts showing the Cumulative sum of deviation of annual rainfall at 12 stations distributed over the study area

Time series modeling

Annual rainfall time series modeling: Table (4) presents the best fit ARIMA models for predicting annual rainfall at 12 stations. It is apparent from the presented results of Table 4 that the best model for predicting annual rainfall for 75% of the stations was ARIMA (0,1,1).

Table 4. The best fit ARIMA models for forecasting the annual rainfall at 12 stations selected over the study area

#	Station	Ljung-Box Q(18)		Test of Performance						ARIMA-model
		Statistics	Sig.	RMSE	MAPE	MASE	AICc	AIC	BIC	
1	Duhok	13.23	0.721	187.90	29.573	0.902	300.658	299.324	302.597	(0,1,1)
2	Zakho	13.86	0.677	159.19	25.353	0.770	293.458	292.125	295.398	(0,1,1)
3	Batel	11.829	0.756	144.52	24.440	0.753	291.851	289.498	293.862	(1,0,1)
4	Akre	15.24	0.578	184.21	24.900	0.734	312.599	311.336	314.742	(0,0,1)
5	Bamarne	11.489	0.717	205.67	20.245	0.696	310.743	306.993	312.448	(1,1,2)
6	Chamanke	16.372	0.498	272.11	24.686	0.873	314.10	313.473	315.655	(0,1,1)
7	Amedi	20.076	0.270	210.34	23.538	0.834	305.954	304.620	307.894	(0,1,1)
8	Zawita	12.724	0.754	230.69	26.239	0.848	309.567	308.233	311.507	(0,1,1)
9	Bardarash	19.309	0.311	160.45	30.911	0.947	281.15	279.74	282.87	(0,1,1)
10	Qasrok	13.602	0.695	181.63	32.888	0.840	285.91	284.503	287.636	(0,1,1)
11	Batufa	11.926	0.805	257.39	26.869	0.759	315.125	313.792	317.065	(0,1,1)
12	KaniMasi	8.743	0.948	250.70	30.973	0.711	313.361	312.028	315.301	(0,1,1)

The models were identified using the sample autocorrelation (AC) and partial autocorrelation (PA) plot shapes. Depend on the pattern of AC and PA, there is a strong indication that the model is mixed, which is the ARMA model. The stationarity in mean was obtained by first differencing. The time series plot indicated that the time series did not require log transformation to achieve stationarity in variance. Further, both ACF and PACF were checked for stationarity after differencing. However, these plots were not shown because of limited space as mentioned earlier. These steps were taken as a guide to reach at the optimum solution in an efficient way. To identify the best fit model at each station by entertaining many tentative models. For this purpose, a grid search over selected values of p and q, varying between 0 and 3 were set up and the corresponding model orders that gave minimum values for Information values (AIC) and Bayesian information criteria were picked. The trials were made with and without differences. Table 4 shows that the Ljung-Box test achieved p values of 0.05. As a result, the presented models are white noise, indicating that the proposed ARIMA models meet the requirement and can be used for forecasting. The residual of the fitted Model was also examined for adequacy purposes. This was accomplished by comparing the ACF and PACF of these residuals at different lags. Plotting residual autocorrelation and partial autocorrelation (not displayed here) was useful in identifying misspecification (7). A good model is a model without autocorrelation and model residual from population with normal distribution Mukhaiyar et al. (25). To verify the suitability of the model, the autocorrelation values of the residual were plotted against lag (not shown here). It was noticed that, there was no spike at any lag indicating that the residual process is random. Therefore, the most appropriate models show in Table are 4 recommended for forecasting. (36) reported that a model with the less number of variables provides the best forecasting output and this goal can be achieved by depending on AIC to select the best model. After selecting the best models, their parameters were estimated. The estimated values were employed for evaluating the model performance and for future forecasting. Furthermore, the residuals of the best fit models were evaluated for normality by plotting normal Q-Q plots, histograms, and residual plots. The residual error plotted showed no pattern, indicating that the models may be used to represent the rainfall data. Additionally, it can be shows from Table 4 that the mean absolute percentage of error varied from as low as 20.25% at Bamarne station to as high as 32.89% at Qasrok station.

Monthly Rainfall Time Series Modelling

Table (**5**) portrays the best fit SARIMA models for forecasting monthly rainfall at 12 stations. The selection of the best fit model was based on the same steps taken and the same criteria considered in the previous section besides considering the seasonality of the data. It appears from the results presented in Table 5 that the model $(1,1,1)x(1, 0, 1)_{12}$ is the best suitable model for predicting and forecasting monthly rainfall for more than 83% of the selected stations (10 out of 12 stations). The selection of the best fit model was depending mainly on the minimum value Bayesian information criteria (BIC). Similar to annual rainfall modeling, the adequacy of the best fit models for predicting monthly rainfall was also based on testing the ACF and PACF residuals at deferent lags. It is also evident from table 5 that the Ljung-Box test attained p values ≥ 0.05. Accordingly, it can be decided that the proposed models are *white noise,* meaning that the proposed ARIMA models fulfill the requirement and can be used for forecasting monthly rainfall over the selected stations. Furthermore, the proposed models' adequacy was confirmed using graphical representations of the residuals. Like the residuals of the annual rainfall, the residuals of the monthly rainfall appeared without any pattern, indicating the appropriateness of the selected models. This implies the selected models were the models without autocorrelation and their residuals were normally distributed. Furthermore, it was revealed that, the mean absolute error of monthly rainfall prediction was ≤ 36 mm in most cases.

Table 5. The best fit SARIMA models for forecasting the monthly rainfall at 12 stations selected over the study area

#	Station	Ljung-Box Q (18)		Test of Performance			ARIMA-model
		Statistics	Sig.	RMSE	MAE	BIC	
1	Duhok	17.346	0.238	46.008	30.250	7.760	(1,1,1)(1,0,1)
2	Zakho	16.106	0.307	44.356	29.960	7.707	(1,1,1)(1,0,1)
3	Batel	12.560	0.549	41.123	27.973	7.560	(1,1,1)(1,0,1)
4	Zawita	17.708	0.220	65.566	42.139	8.489	(2,1,1)(0,1,1)
5	Akre	17.230	0.244	53.548	36.074	8.084	(1,1,1)(1,0,1)
6	Bamarne	17.492	0.231	60.017	40.309	8.312	(1,1,1)(1,0,1)
7	Amedi	21.047	0.100	59.879	39.565	8.287	(1,1,1)(1,0,1)
8	Batufa	19.258	0.155	52.699	35.952	8.061	(1,1,1)(1,0,1)
9	KaniMase	17.446	0.293	66.245	42.712	8.501	(1,1,1)(0,1,1)
10	Bardarash	15.889	9.321	35.645	23.802	7.285	(1,1,1)(1,0,1)
11	Qasrok	20.729	0.122	44.502	28.852	7.729	(1,1,1)(1,0,1)
12	Chamanke	16.931	0.260	83.834	53.610	9.023	(1,1,1)(1,0,1)

Rainfall Forecasting

The performance of the most relevant ARIMA models for the 12 stations was also assessed by forecasting monthly rainfall data from January 2021 to December 2025 to indicate the models' adequacy and performance (Fig.4). The forecasting values included sample period forecast and post sample period forecast. The first one was employed to develop confidence in the model, while the latter was used to generate future forecasts for use in planning and other uses such as agricultural activities. As can be noticed from Fig. the forecasted values show similar pattern of the original data series to some extent. All forecasted values lied within 95% confidence interval (not shown on the graphs). Overall, the sample period forecast for some months deviated slightly from the original data. It is also evident from Fig. that the models under forecast the monthly rainfall for Dec and January. The main reason for this inconsistency stems from the fact that during some years, for instance during 2019 the rainfall rate was quite unusual. At this stage these models can be used as the first approximation for forecasting. The accuracy of prediction can be increased in the future by expanding the database and correcting the data using the double mass curve. The autocorrelation function (ACF) and partial autocorrelation function (PACF) were also tested for residual errors resulting from the best fitting models for the monthly rainfall data to ensure that this model is representative for our data and could be used to forecast future rainfall data from 2021 to 2025. It was observed that, as previously stated, the residual errors did not serially correlated. Further, the residuals normal probability and histogram plots supported the adequacy of the models at this level of study. Additionally, it was observed that residual values have constant variance and the points come into view to be randomly scattered around zero.

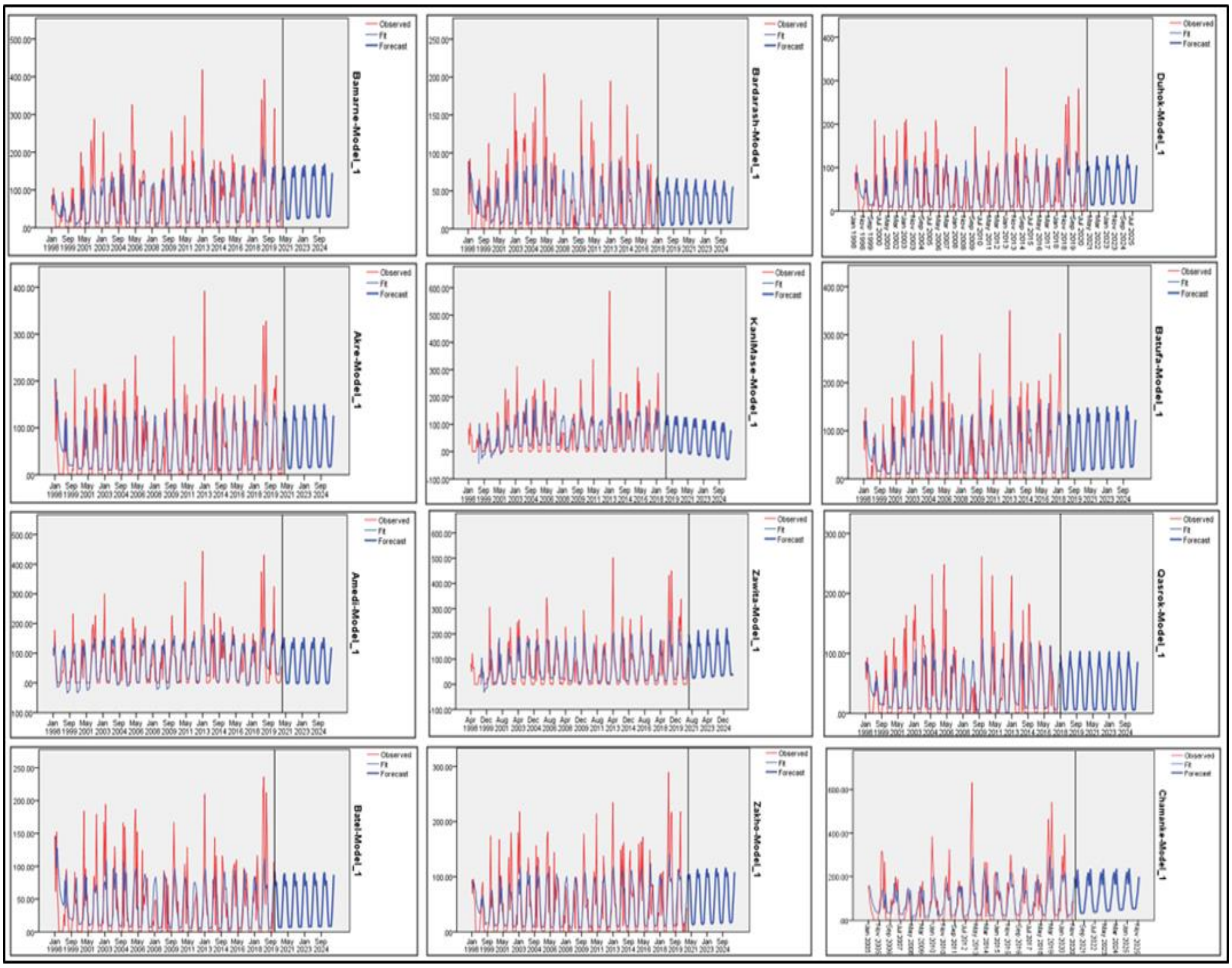

Fig.4. Comparing the observed, fitted and forecasting monthly rainfall at 12 station selected over the study area

It should be emphasized that for successful models, a model with the less variables number gives the best forecasting results. Therefore, the model selection was depend on the minimum value for Akaike's Information Criterion. It is commendable to mention that there will be improvement in accuracy of prediction with an increase in number autoregressive and moving average parameters. This trial did not highlight due to improvement was in favor of the training data (fitted values), but did not in favor of the forecasting values.

CONCLUSIONS

It can be decided from the results of this study that time series study is homogeneous and can be considered for further analysis. Unexpectedly, the monthly and annual rainfall time series recorded at most of stations exhibited positive trend during the period from 1998 to 2020 with abrupt changes during 2019 and 2020 at a few stations. Additionally, the results unveiled that the study time series can be forecast with reasonable accuracy using ARIMA and SARIMA models, but still in need of further improvement after database expansion in the future.

REFERENCES

1. Abdul-Aziz, A.R., M. Anokye, A. Kwame, L. Munyakazi and N.N.N Nsowah-Nuamah,. 2013. Modeling and forecasting rainfall pattern in Ghana as a seasonal ARIMA process: The case of Ashanti region. Int. J. Humanities Social Sci., 3(3), 224-233.

2. Agha, O.M.A.M., S.Ç. Bağçacı and N.Şarlak. 2017. Homogeneity analysis of precipitation series in North Iraq. IOSR Journal of Applied Geology and Geophysics, 5(03):57-63.

3. Al Balasmeh, O., R. Babbar and T. Karmaker. 2019. Trend analysis and ARIMA modeling for forecasting precipitation pattern in Wadi Shueib catchment area in Jordan. Arabian Journal of Geosciences, 12(2), :1-19.

4. Alahacoon, N. and M. Edirisinghe. 2021. Spatial variability of rainfall trends in Sri Lanka from 1989 to 2019 as an indication of

climate change. ISPRS International Journal of Geo-Information, 10(2), :84.
5. Ali, S.M. 2013. Time series analysis of Baghdad rainfall using ARIMA method. Iraqi Journal of Science, 54(4), :1136-1142.
6. AL-Lami, A.M., Y.K. AL-Timimi and A.M. AL-Salihi. 2014. The homogeneity analysis of rainfall time series for selected meteorological stations in Iraq. Diyala Journal for Pure Science, 10(2), :60-77.
7. Anderson, O.D. 1977. time series analysis and forecasting: Another Look at the Box-Jenkins Approach. Journal of the Royal Statistical Society: Series D (The Statistician), 26(4), :285-303.
8. Attah, D.A. and G.M. Bankole. 2012. Time series analysis model for annual rainfall data in lower Kaduna catchment Kaduna, Nigeria. International Journal of Research in Chemistry and Environment (IJRCE), 2(1), :82-87.
9. Box G.E.P. and G. Jenkins. 1970. Time Series Analysis: forecasting and control. San Francisco, Holden-Day.
10. Coban, V., Guler, E., Kilic, T. and S.Y. Kandemir. 2021. Precipitation forecasting in Marmara region of Turkey. Arabian Journal of Geosciences, 14(2), :1-10.
11. Cunderlik, J.M. and D.H. Burn. 2002. Local and regional trends in monthly maximum flows in southern British Columbia. Canadian Water Resources Journal, 27(2), :191-212.
12. D'Adamo, I., M. Gastaldi, G. Ioppolo and P. Morone. 2022. An analysis of Sustainable Development Goals in Italian Cities: Performance Measurements and Policy Implications. Land Use Policy, 120, :106278.
13. Di Virgilio, N., A. Monti and G. Venturi. 2007. Spatial variability of switchgrass (*Panicum virgatum L.*) yield as related to soil parameters in a small field. Field crops research, 101(2), :232-239.
14. Elzeiny, R., M. Khadr, S. Zahran and E. Rashwan. 2019. Homogeneity analysis of rainfall series in the upper blue nile river basin, Ethiopia. Journal of Engineering Research, 3(September), :46-53.
15. El-Jiboori, M. H., M. S, Wid, and Y. K. Al-Timimi .2024. Correlation coefficient analysis between pm2.5 concentrations and some meteorological parameters in Iraq. Iraqi Journal of Agricultural Sciences, 55(4), 1292-1302. https://doi.org/10.36103/2twexb46
16. Hafi, Z.B. 2016. Trend analysis of annual rainfall time series in Northwest Libya. The Libyan Journal of Science A, 19, :39-54.
17. Helsel, D.R. and R.M Hirsch. 1992. Statistical methods in water resources Elsevier. (Vol. 49).
18. Hess, A., H. Iyer and W. Malm. 2001. Linear trend analysis: a comparison of methods. Atmospheric environment, 35(30), :5211-5222.
19. Kundzewicz, Z.W. and A.J. Robson. 2000. Detecting Trend and Other Changes in Hydrological Data. Word Climate Program-Data and Monitoring. World Meteorological Organization, Geneva (WMO/TD-No. 1013).
20. Lin, Z., S. Stamnes, Jin, Z., I. Laszlo, S.C.Tsay, W.J. Wiscombe and K. Stamnes. 2015. Improved discrete ordinate solutions in the presence of an anisotropically reflecting lower boundary: Upgrades of the DISORT computational tool. Journal of Quantitative Spectroscopy and Radiative Transfer, 157, :119-134.
21. Lobell, D.B., G.L. Hammer, G. McLean, C. Messina, M.J. Roberts and W. Schlenker. 2013. The critical role of extreme heat for maize production in the United States. Nature climate change, 3(5), :497-501.
22. Mahmud, I., Bari, S.H. and M. Rahman. 2017. Monthly rainfall forecast of Bangladesh using autoregressive integrated moving average method. Environmental Engineering Research, 22(2), :162-168.
23. Mahsin, M., Y. Akhter and M. Begum. 2012. Modelling rainfall in Dhaka Division of Bangadesh using time series analysis, Journal of mathematical modelling and application, 1 (5) 67-73.
24. Momani, P.E.N.M. and P.E. Naill. 2009. Time series analysis model for rainfall data in Jordan: Case study for using time series analysis. American Journal of Environmental Sciences, 5(5), p.599.
25. Mukhaiyar, U., N.M. Huda, K.N. Sari and U.S. Pasaribu. 2020, March. Analysis of generalized space time autoregressive with exogenous variable (GSTARX) model with outlier factor. In Journal of Physics: Conference Series (Vol. 1496, No. 1, p. 012004). IOP Publishing.

26. Murat, M., I. Malinowska, M. Gos and J. Krzyszczak. 2018. Forecasting daily meteorological time series using ARIMA and regression models. International agrophysics, 32(2).
27. Oyamakin, S.O., F.J. Ayoola and O.T. Dare. 2010. Time series analysis of rainfall and temperature in south west Nigeria. Pac J Sc Tech, 1(2), :552-564.
28. Pal, I. and A. Al-Tabbaa. 2009. Trends in seasonal precipitation extremes–An indicator of ‘climate change’in Kerala, India. Journal of Hydrology, 367(1-2), .62-69.
29. Panda, A. and N. Sahu. 2019. Trend analysis of seasonal rainfall and temperature pattern in Kalahandi, Bolangir and Koraput districts of Odisha, India. Atmospheric Science Letters, 20(10), p.e932.
30. Paz-Gonzalez, A., S.R. Vieira and M.T.T. Castro. 2000. The effect of cultivation on the spatial variability of selected properties of an umbric horizon. Geoderma, 97(3-4), :273-292.
31. Pazvakawambwa and A. A. Ogunmokun. 2016. A time-series forecasting model for Windhoek Rainfall, Namibia. Environmental Science.
32. Sonali, P. and D.N. Kumar. 2013. Review of trend detection methods and their application to detect temperature changes in India. Journal of Hydrology, 476, :212-227.
33. Toreti, A., G. Fioravanti, W. Perconti and F. Desiato. 2009. Annual and seasonal precipitation over Italy from 1961 to 2006. International Journal of Climatology: A Journal of the Royal Meteorological Society, 29(13), :1976-1987.
34. Tuomenvirta, H. 2002. Homogeneity testing and adjustment of climatic time series in Finland. Geophysica, 38(1-2), :15-41.
35. Al-Timimi, Y., K., and F. Y. Baktash. 2024. Monitoring the shift of rainfed line of 250 mm over Iraq. Iraqi Journal of Agricultural Sciences, 55(3):931-940. https://doi.org/10.36103/h10cqh53
36. Zakaria, S., N. Al-Ansari, S. Knutsson and T. Al-Badrany. 2012. ARIMA Models for weekly rainfall in the semi-arid Sinjar District at Iraq. Journal of Earth Sciences and Geotechnical Engineering, 2(3).