DOI:10.28974/idojaras.2024.3.6

IDŐJÁRÁS

Quarterly Journal of the HungaroMet Hungarian Meteorological Service
Vol. 128, No. 3, July – September, 2024, pp. 379–398

Forecasting extreme precipitations by using polynomial regression

Fatih Dikbaş

Pamukkale University, Civil Engineering Department, Denizli, Turkey

Author E-mail: f_dikbas@pau.edu.tr

(Manuscript received in final form September 18, 2023)

Abstract— It is well known that the recent global warming intensifies the magnitude of rainfalls due to the increase in water content in the atmosphere. Therefore, the probability of exceeding the previously observed extreme precipitation values also increases with the experienced climate change, and forecasting extreme weather events is becoming more important. This paper presents a new polynomial regression approach and software (PolReg), where future extreme precipitations exceeding all previous observations are estimated for each month of year by using prediction bounds with a level of certainty at 95%. The presented method determines the degrees and coefficients of best-fitting polynomials for each precipitation station and forecasts the expected extreme value for each month of year by using the determined polynomials. The performance of the method is tested by removing and estimating a total of 792 highest observed monthly total precipitation values of 66 precipitation stations in Turkey (the highest observation for each month of year for each station). The results show that the proposed method and the provided software have a high performance and accuracy in estimating future precipitation extremes and might be applied in many disciplines dealing with forecasting probable extreme values.

Key-words: forecasting extreme precipitations, polynomial regression, data-driven modeling, hydrometeorology, Turkey

1. Introduction

Precipitation is one of the principal atmospheric forcing parameters required for hydrologic modeling (*Liu and Coulibaly* 2011). The forecasting of precipitation extremes is becoming more important with the worldwide increase in the frequency and intensity of water-related disasters like floods and droughts and dwindling water supplies (*Baxevani and Wilson* 2018; *Fowler et al.* 2010; *Leconte et al.* 2013; *McElroy* 2016). These forecasts can provide estimates of probabilities of having more (or less) precipitation than certain specified amounts (*Unkašević et al.* 2004). Understanding the variability and forecasting the probable extreme values of precipitation are also highly important for the efficient prevention of potential natural disasters (*Beguería and Vicente-Serrano* 2006; *Bhatia et al.* 2019; *Block and Rajagopalan* 2007; *Keupp et al.* 2019; *Tian et al.* 2014; *Yuan et al.* 2017). Accurate determination of design rainfall directly influences the selection of appropriate dimensions for water structures and prevents loss of lives and environmental damage (*Zhang et al.* 2021). With the impact of climate change and abnormal weather caused by the recent global warming, the magnitude of precipitation is likely to intensify in most regions of the world as indicated by both observations and climate model simulations (*Hou et al.* 2014; *Ibrahim* 2019; *Lazoglou et al.* 2019; *Li et al.* 2019; *Rai et al.* 2019; *Reager and Famiglietti* 2009; *Zhang et al.* 2013). For example, *Schönwiese et al.* (2003) have demonstrated that the increase in extreme wet months is reflected in a systematic increase in the variance and the Weibull probability density function parameters, respectively. Although there is a relative agreement in climate change scenarios about changes in extremes, significant differences in the magnitudes of extremes have been reported (*Kyselý and Beranová* 2009; *Meena et al.* 2019; *Zhao et al.* 2019). Observations indicate that there is a continuous change in the climate (*Knox* 1993) and consequently in the hydrological cycle and associated rainfall patterns (*Zhang et al.* 2009). A study on the changes in flood frequency over paleo-timescales demonstrated that the estimated flood exceedance probability can increase quite rapidly over time (*Porporato and Ridolfi* 1998). Due to these reasons, the changing character of return periods of precipitation extremes should be considered in the hydrological design and water resources management studies (*Su et al.* 2009).

Hydrologic variables depict a two-dimensional periodic behavior because of seasonality associated with the hydrologic cycle (*Dikbas* 2017a; *Dikbas* 2017b). For example, precipitation generally shows significant variations throughout a year, but the observations in the sub-periods (months, seasons, etc.) tend to be in a definite range, and alternating trends might be observed for a station in different months of the year (*Trömel and Schönwiese* 2007). In Turkey, minimum precipitations are generally observed in the July-September (summer) period varying in a low range, while the highest precipitations are mostly experienced within the November-January (winter) period. Though precipitation is normally

seasonal, its uncertain (*Kent et al.* 2015), nonstationary, and sometimes chaotic behavior (*Sivakumar* 2000; *Sivakumar et al.* 1999; *Wilks* 2012) makes the observation, quantification, estimation, and forecasting of precipitation challenging (*Gao et al.* 2017; *Schliep et al.* 2010; *Wang and Lin* 2015). The two-dimensional behavior of precipitation can be observed when the data series are placed on a matrix so that each row contains observations for each month (12 rows). This approach provides significant advantages in data-driven modeling studies over the one-dimensional approach and allows obtaining accurate modeling results and estimations as in the 3D imputation (*Dikbas* 2016b), frequency-based imputation (*Dikbas* 2016a; *Dikbas* 2017b), and two-dimensional correlation methods (*Dikbas* 2017a, 2018a; *Dikbas* 2018b) which use data located on a two-dimensional matrix.

It can be visually observed that the time series graphs of sorted monthly total precipitation series generally depict a nonlinear behavior and the slope of the curve always increases towards the higher values. This property of the precipitation series was the main reason for the selection of polynomial regression as the method for estimating future extreme values. Polynomial regression can be used to describe trend curves to model complicated patterns of sorted temporal data. Together with its applications on various hydrological data, some studies making use of polynomial regression for modeling precipitation also exist in literature (*Acock* and *Pachepsky* 2000; *Adnan et al.* 2016; *Block and Rajagopalan* 2007; *George et al.* 2016; *Goodale et al.* 1998; *Hwang et al.* 2012; *Stefanescu et al.* 2014; *Tian et al.* 2014). This study presents a methodology and software for reliably estimating expected future extreme monthly total precipitation values exceeding all previous observations of a station for each month of the year by applying univariate polynomial regression on the observed precipitation values of the investigated station itself. The prediction bounds for each monthly series are determined by using a 95% confidence level. The observations are located on two-dimensional matrices where months are in rows. The application of the developed method on 66 precipitation observation stations in Turkey has shown that the method successfully estimates the removed extreme observations. The extreme precipitation values expected with a high occurrence probability in the future were also calculated for all months for all stations. The forecasted extreme precipitation values have values exceeding all previous observations and provide crucial information especially to be considered in water resources management projects. The proposed method has a flexible structure in that new observations can be easily appended to the existing input dataset, and the best-fitting polynomials might be updated accordingly allowing consideration of the non-stationarity of the precipitation series. The results obtained for the station 18-003 are reported to illustrate the details of the developed methodology as presented in Section 3, and the summary of obtained results for all stations are presented in Section 4.

2. Study area and data

The developed method and the provided software were tested on 66 precipitation observation stations across 20 different basins in Turkey (*Fig. 1*). The figure shows the average annual precipitation values in Turkey between 1981 and 2010. The selected stations represent the majority of the climate and elevation zones and cover most of the hydrological basins in Turkey. The climate in Turkey is moderately dry with higher average precipitation in the coastal regions. Average annual total precipitation varies between 250 mm (Salt Lake region) and 2300 mm (Rize region on the coast of the Black Sea). The Black Sea coastline receives precipitation throughout the year and the Aegean and Mediterranean coasts are wet in winter but dry during the summer seasons.

Fig. 1. Map of 1981–2010 average annual precipitation in Turkey, including the locations of the 66 stations used in this study.

The descriptive statistics for all stations, including percentiles and best-fitting distributions are outlined in *Table 1*. The majority of precipitation series (46/66) fit the Wakeby distribution. The probability distributions for all stations are positively skewed and leptokurtic (except for station 21-007) as shown by the skewness and excess kurtosis measures. The maximum observations of all stations constitute the lower limit for the estimations of the expected next extreme precipitation, and the maximum observations vary between 123.0 mm (station 15-010) and 893.5 mm (station 08-006).

Table 1. Descriptive statistics, percentiles, and best-fitting distributions for the first ten stations. The complete table for all stations is provided as an online supplement because of space restrictions.

	Station	01-004	01-005	01-008	02-004	02-009	02-011	02-012	03-009	03-013	03-027
	Elevation (m)	90	395	395	35	40	10	180	20	770	240
Statistic	**Sample Size**	492	492	480	456	492	468	456	480	528	516
	Mean	42.3	54.0	46.4	69.3	77.8	47.2	82.5	46.3	55.6	52.2
	Variance	1139	1627	1331	3015	4301	1541	5073	1755	2680	1420
	Std. Error	1.52	1.83	1.68	2.60	2.99	1.82	3.39	1.93	2.25	1.69
	Skewness	1.11	1.04	1.45	1.18	1.55	1.27	1.66	1.50	1.44	0.78
	Excess Kurtosis	1.06	1.01	3.56	2.07	2.76	2.21	4.02	4.14	2.91	0.28
Percentiles	**Min**	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0
	5%	1.2	3.0	4.6	4.2	7.0	1.7	4.6	0.0	0.3	1.2
	10%	4.6	8.0	7.8	7.8	13.4	4.4	10.2	2.3	2.3	7.7
	25% (Q1)	17.0	24.2	18.5	26.8	30.1	17.1	31.8	13.2	15.2	22.6
	50% (Median)	33.8	46.1	38.7	56.5	61.8	38.0	64.5	36.2	42.9	46.0
	75% (Q3)	60.0	77.9	65.0	101.1	103.3	70.5	114.5	66.8	83.0	74.7
	90%	89.6	107.6	93.3	142.0	165.6	98.0	171.1	107.8	129.6	105.0
	95%	111.1	134.3	117.6	174.5	211.2	117.8	225.2	123.6	154.3	122.4
	Max	170.1	205.0	264.3	365.5	351.0	242.5	476.5	315.9	320.5	184.1
	Best-fit Distr.	WAK	WAK	WAK	WAK	WAK	WAK	WAK	WAK	BETA	WAK

https://www.dropbox.com/scl/fi/48xw7tbif7qoeiuocqhvh/Table.1.Descriptive.Statistics.and.Quantiles.-Full.Version.Supplement.xlsx

3. Polynomial regression

Polynomial regression is intrinsically a form of multiple linear regression. The polynomial regression model can be expressed as follows:

$$y_i = a_0 x_i^m + a_1 x_i^{m-1} + a_2 x_i^{m-2} + \cdots + a_{m-2} x_i^2 + a_{m-1} x_i + a_m + \varepsilon_i \quad (i = 1, 2, \ldots, n), \tag{1}$$

where n is the number of observations to which the polynomial function is fit, m is the degree of the polynomial to be fit to the observations, y_i is the ith element of the response vector containing the observations (the dependent variable), a_j is the coefficients (parameters) of the fitted polynomial (j = 1, 2, … , m), ε_j is the

random error (the difference between the estimation and observation), and x_i is the ith element of the vector containing the independent variable.

Eq.(1) can be expressed in matrix form as follows:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1^m & x_1^{m-1} & x_1^{m-2} & \dots & x_1^2 & x_1 & 1 \\ x_2^m & x_2^{m-1} & x_2^{m-2} & \dots & x_2^2 & x_2 & 1 \\ x_3^m & x_3^{m-1} & x_3^{m-2} & \dots & x_3^2 & x_3 & 1 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ x_n^m & x_n^{m-1} & x_n^{m-2} & \dots & x_n^2 & x_n & 1 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_n \end{bmatrix} . \quad (2)$$

As shown with the examples below, the sorted precipitation series mostly have curved shapes far from being linear and the sections with high precipitation values generally have a higher slope than the lower values. The polynomial regression method presented in this paper determines the degrees (m) and coefficients (a_0, a_1, … , a_m) of the polynomials best fitting to the precipitation series of a station for each month of the year (12 observation series and 12 polynomials for each station) with a level of certainty at 95%. The series in each month are sorted in ascending order before fitting the polynomial function. Thus, in Eq.(2), y_1 stands for the minimum and y_n stands for the maximum observation in an individual month for which the polynomial is being fit. The precipitation series are not evaluated as a whole because of the seasonal variations in the behavior of precipitation. Therefore, 12 polynomial functions are determined for each station. Then, for each month, estimation of the next probable extreme precipitation (y_{n+1}) becomes possible by using the determined best-fitting polynomial function. The details of the implemented approach are presented below for the selected station 18-003 (Uzunpınar) located in the central region of Kayseri, Sivas, and Malatya cities.

4. Results

4.1. Forecasting extreme precipitation for June for station 18-003

Estimation of the expected highest monthly precipitation exceeding all the previous observations for each month of year is the main aim of the method presented in this manuscript. The raw time series data of precipitation generally has complicated quantitative and temporal associations as shown in the heatmap of the observations of the station 18-003 for the 46 years between 1960 and 2005 (top panel in *Fig. 2*). This complicated behavior makes estimation of precipitation more challenging than other hydrologic variables like streamflow. When the return period is not the primary interest, sorting the observations is one of the mostly applied methods for assessing quantitative associations in precipitation

observations. Therefore, prior to making estimations of extreme precipitation values, all rows containing the observations for each month are sorted in ascending order. The bottom panel in *Fig. 2* shows the heatmap of the sorted precipitation observations of the station 18-003. The sorted values generate monotonically increasing series making estimation of the next highest value probable by using polynomial regression. The values to be estimated will constitute the 47th column in the bottom panel in *Fig. 2,* and they are calculated by separately determining the polynomial equations representing the relationships for all the series in the 12 rows of the horizontally sorted data matrix. Consequently, 12 polynomials fitting to the sorted precipitation series for each month of year are obtained for each station, because there are significant differences in the behavior of the precipitation series in view of both expected and observed precipitation amounts for each month of year.

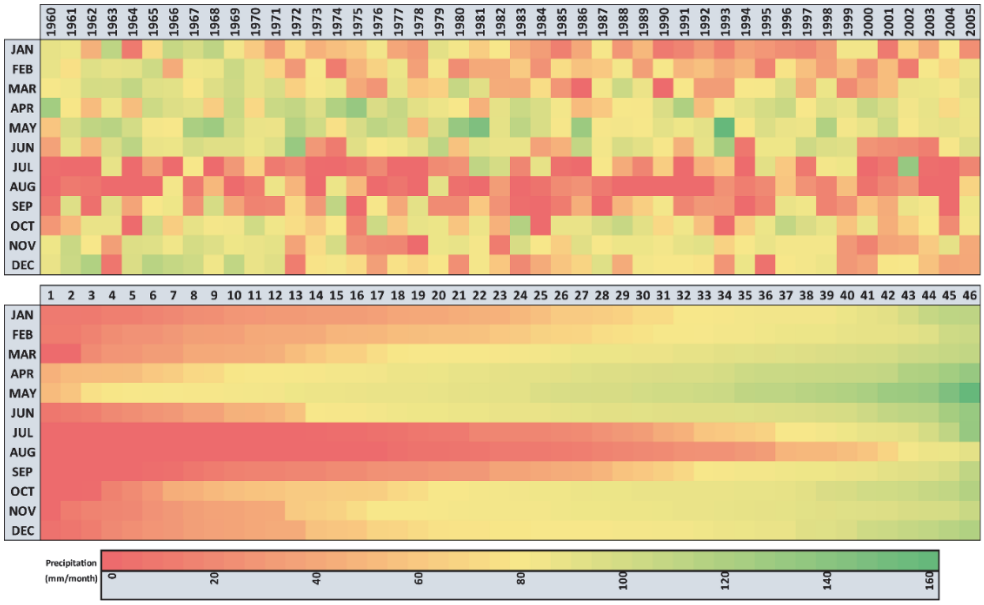

Fig. 2. Heatmaps of the monthly total precipitation observations of station 18-003 (top) and the row-wise sorted values (bottom).

For finding the best fitting polynomial for each row, the developed PolReg software (the link to the freely distributed code of the software is provided at the end of the manuscript above the References) fits polynomials from 2nd to 8th degree and calculates the performance of each fit by using 11 different statistical measures (mean squared error (MSE), normalized mean squared error (NMSE), root mean squared error (RMSE), normalized root mean squared error (NRMSE), mean absolute error (MAE), mean bias error (MBE), coefficient of correlation (r), coefficient of determination (D), coefficient of efficiency (E), maximum absolute

error (MaxAE), and mean absolute scaled error (MASE)). The software automatically generates plots of all fitted polynomials.

Fig. 3 shows the plots of the polynomials generated for the sorted observations of station 18-003 in the month of June during the 46 years. The blue points are the observed values and the grey point on the right of each panel shows the next expected value determined by simply calculating the y-value of the function of the fitted polynomial corresponding to the 47th observation on the x-axis. The plots in each panel show the fitted polynomials starting from the 2nd degree up to the 8th degree. The green dashed lines indicate the prediction bounds with a level of certainty at 95%. This interval indicates that there is a 95% chance that the new observation is contained within the lower and upper prediction bounds.

The highest observed values are out of the prediction bounds of the fitted polynomials in the plots for the second- and third-degree polynomials. Also, the expected next-extreme precipitations are lower than the highest observed values for the second- and third-degree polynomials. Consequently, the plots obtained for the polynomials with a degree of 4 and higher fit the sorted series better and provide more accurate estimates for the next highest expected precipitation as shown in the plots for June.

The plots of the polynomials do not provide sufficient clues for deciding on the best-fitting polynomial. The PolReg software calculates 11 statistical performance measures for each fitted polynomial, and the user determines the best-fitting polynomial by evaluating the plots and the performance tables together. The software generates a results table for each month and summarizes the results of the fitting procedure. *Table 2* shows the outputs obtained for the month of June. The table contains the coefficients of each fitted polynomial (a_0 to a_8); the maximum precipitation value observed in the evaluated month (Max: 107.60 mm for June which was observed in 1972); the estimated maximum value by the fitted polynomial (EstMax); the expected next highest precipitation forecasted by the fitted polynomial (EstNextMax); the upper (PredIntU(n)) and the lower (PredIntL(n)) limits of the prediction interval for the estimated maximum precipitation corresponding to the observed maximum precipitation, and the upper (PredIntU(n+1)) and lower (PredIntL(n+1)) limits of the prediction interval for the expected next maximum precipitation. Here, n is the number of observations used for fitting the polynomials and making the estimations (n = 46 for the station 18-003). The polynomials from the 5th to the 8th degree produced estimates between 106.25 mm and 108.03 mm for the observed maximum (which is 107.60 mm for June) and between 122.03 mm and 127.17 mm for the expected next maximum precipitation. The software also reports the statistical performance indicators between the observed values and the fitted polynomials. The best performances are indicated with green background and the worst performances are indicated with a red background color. For the month of June, for the station 18-003, the performance measures point out that the 8th-degree polynomial best fits the observations.

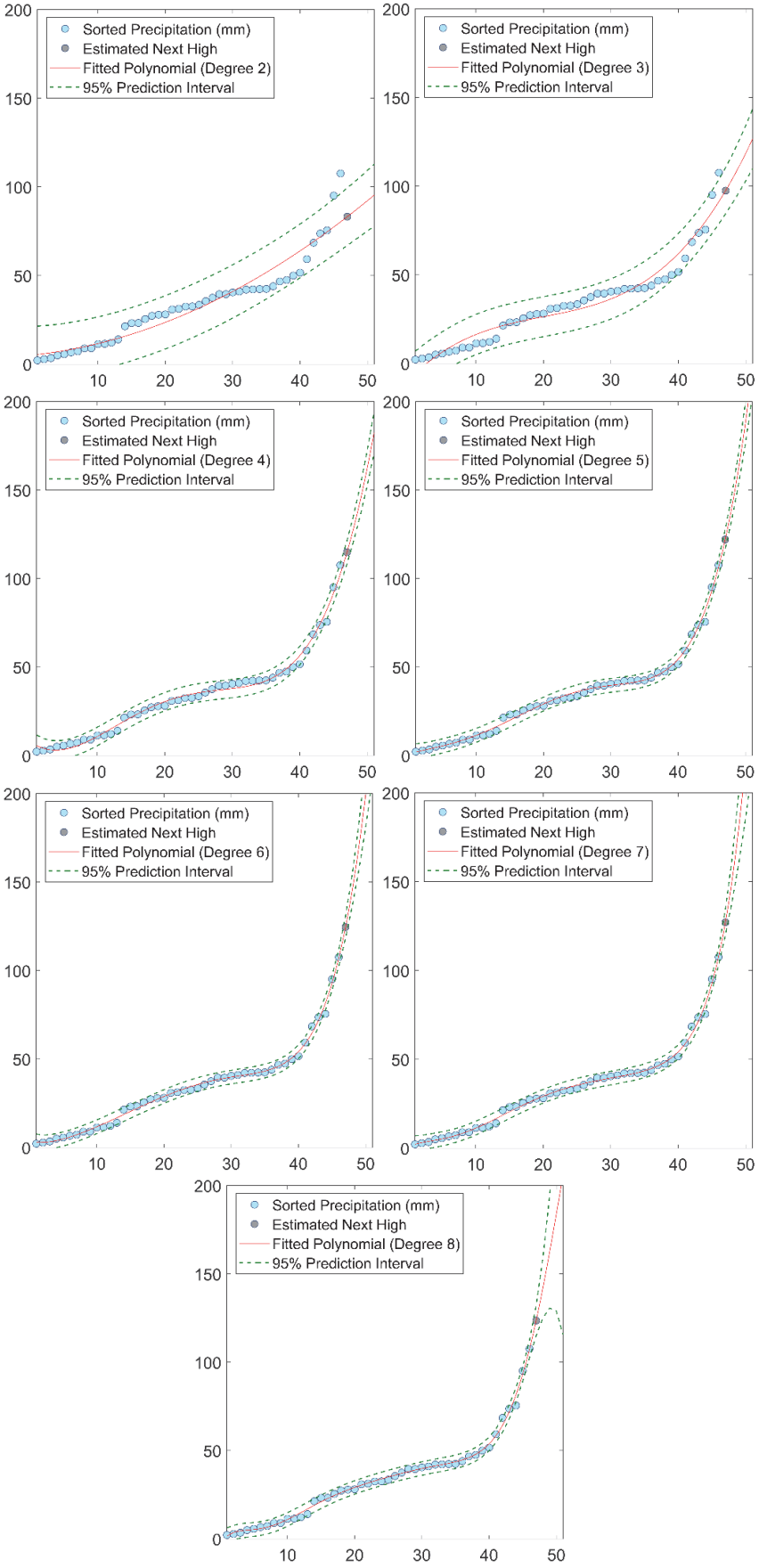

Fig. 3. Plots of the 2nd to the 8th degree polynomials fitted to the sorted 46-year-long monthly total precipitation observations of the station 18-003 in June.

In the case of research on extreme values, a function that generally has a good fit to most of the observed values but not to the extreme values should not be regarded as a good fit. This situation is clearly depicted by the 2nd and 3rd degree polynomials which have an r score of 0.954 and 0.975, respectively (scores which might be regarded as a high correlation), but they are far from representing the extreme values even though they generally have a good fit to the lower observations. Therefore, for providing more clues on the best fitting polynomials, the software also calculates the 11 statistical performance indicators between the highest five observations and the values of the fitted polynomials corresponding to the observed extremes and reports them at the bottom section of the output table.

The statistical measures for the extreme values in June indicate that the 7th degree polynomial fits best to the observed extremes. Consequently, by looking at the plots and the table, it might be concluded that a monthly total precipitation value up to 127.17 mm might be expected in June in the station 18-003 with a prediction bound between 119.93 mm and 134.42 mm at 95% confidence level.

Table 2. Output table summarizing the polynomial fits and statistical measures for the 46-year-long monthly total precipitation observations of the station 18-003 in June

JUNE							
Pol.Deg.:	2	3	4	5	6	7	8
a_0	0.0272	0.0026	0.0002	0.0000	0.0000	0.0000	0.0000
a_1	0.3828	-0.1558	-0.0180	-0.0005	0.0000	0.0000	0.0000
a_2	5.1129	3.8600	0.4724	0.0136	0.0007	0.0001	0.0000
a_3		-9.2332	-2.8683	-0.0955	-0.0167	-0.0035	0.0008
a_4			8.1035	1.1139	0.2708	0.0557	-0.0184
a_5				0.9243	-0.7247	-0.3679	0.2493
a_6					3.4269	1.7129	-1.6842
a_7						0.7807	5.7060
a_8							-2.8217
Max	107.60	107.60	107.60	107.60	107.60	107.60	107.60
EstMax	80.25	91.30	102.52	106.25	107.25	108.03	107.29
EstNextMax	83.16	97.51	114.85	122.03	124.53	127.17	123.57
PredIntU(n)	96.22	103.80	108.46	110.91	111.98	112.84	112.12
PredIntL(n)	64.29	78.80	96.58	101.60	102.52	103.22	102.45
PredIntU(n+1)	99.36	110.55	121.37	127.54	130.76	134.42	132.14
PredIntL(n+1)	66.97	84.47	108.33	116.51	118.30	119.93	115.00
Goodness of Fit Measures for the Whole Series:							
MSE	**49.6702**	27.0253	5.4196	2.9826	2.7899	2.6524	**2.4943**
NMSE	**0.0877**	0.0477	0.0096	0.0053	0.0049	0.0047	**0.0044**
RMSE	**7.0477**	5.1986	2.3280	1.7270	1.6703	1.6286	**1.5793**
NRMSE	**0.2961**	0.2184	0.0978	0.0726	0.0702	0.0684	**0.0664**
MAE	**4.9392**	4.2176	1.8697	1.2329	1.2182	1.2013	**1.1155**
MBE	0.0000	**0.0000**	0.0000	0.0000	0.0000	0.0000	**0.0000**
r	**0.9541**	0.9753	0.9951	0.9973	0.9975	0.9976	**0.9977**
d	**0.9104**	0.9512	0.9902	0.9946	0.9950	0.9952	**0.9955**
E	**0.9104**	0.9512	0.9902	0.9946	0.9950	0.9952	**0.9955**
MaxAE	**27.3472**	16.2979	6.8502	6.3976	5.9546	**5.3896**	6.0817
MASE	**2.1088**	1.8007	0.7983	0.5264	0.5201	0.5129	**0.4763**
Goodness of Fit Measures for the Highest 5 Values:							
MSE	**213.1992**	76.8190	17.1623	11.7163	11.4662	**11.4318**	12.1024
NMSE	**0.7762**	0.2797	0.0625	0.0427	0.0417	**0.0416**	0.0441
RMSE	**14.6013**	8.7646	4.1427	3.4229	3.3862	**3.3811**	3.4789
NRMSE	**0.8810**	0.5288	0.2500	0.2065	0.2043	**0.2040**	0.2099
MAE	**9.6929**	6.7377	3.4075	2.8099	**2.7652**	2.9008	2.8112
MBE	**9.4325**	3.6221	0.4461	**0.2509**	0.3834	0.5720	0.3786
r	**0.9532**	0.9604	0.9691	0.9732	0.9750	**0.9771**	0.9740
d	**0.9085**	0.9223	0.9391	0.9470	0.9506	**0.9547**	0.9487
E	**0.0298**	0.6504	0.9219	0.9467	0.9478	**0.9480**	0.9449
MaxAE	**27.3472**	16.2979	6.8502	6.3976	5.9546	**5.3896**	6.0817
MASE	**0.9916**	0.6893	0.3486	0.2875	**0.2829**	0.2968	0.2876

4.2. Model validation

The exceedance probabilities for the observed and estimated highest values of the station 18-003 for each month are calculated by using the best-fitting distributions (*Table 3*). The parameters of the best fitting distributions and the probabilistic estimates determined by using the inverse cumulative distribution function for $p=1-(1/46)=0.97826$ are also presented for a comparison of expected and experienced observations for the investigated period. EasyFit software was used for deciding the best fitting distributions. The used software makes goodness of fit tests according to the Kolmogorov-Smirnov, Anderson-Darling, and chi-squared tests. The software supports 55 continuous and discrete distributions and sorts the tested distributions according to their scores in the goodness of fit tests. The distribution with the highest scores is regarded as the best fitting distribution. Then, the exceedance probabilities are calculated based on the parameters of the best-fitting distribution. It must not be forgotten that the return period of any observed extreme event is not equal to the investigated period, because the observed extreme might always have a much longer (or even shorter) return period. The selected observation period and the observed extremes should never be regarded as dependent variables. This situation is also pointed out by the exceedance probabilities determined for the observations used in this study, where the exceedance probabilities are very low for some months, even though all of the highest values were experienced within the investigated 46-year period. For example, the exceedance probabilities of the observed extremes for five months (March, June, July, September, and November) vary between 1% and 1.3% based on the best-fitting distributions. The exceedance probabilities calculated for the forecasted future extreme precipitations provide clues about the probable return periods of extreme events. Even though the return periods for some forecasts seem to be long, the experienced extreme values in the evaluated period show that the forecasted future extremes are not far from that being expected.

As another method applied for testing the performance of polynomial regression in the estimation of the observed highest values, all highest observed values for each month of year in each dataset (12 values for each station and a total of 792 extremes) were removed from the datasets, and polynomial regression was used for estimating the removed extremes. An output table was also generated for the calculations (*Table 4*) presenting the results for June for the station 18-003 after the highest value observed in June 1972 (107.60 mm) was removed from the dataset.

The results obtained by the best fitting 2nd and the 3rd degree polynomials were again far from forecasting the deliberately removed extreme value, but the 5th to the 8th degree polynomials all produced forecasts close to the removed value (104.62 mm to 109.38 mm), and the removed value is within the prediction intervals of the 5th to the 8th-degree polynomials. The software does not know the value of the removed extreme during the estimation process.

Table 3. Exceedance probabilities for the observed highest precipitations and estimated future extremes of the station 18-003 for each month of year

	Best-Fitting Distribution	X1*	X(P=0.98)	P(X>X1)	X2*	P(X>X2)
January	Log-Pearson 3	73.9	77.2	0.024	81.6	0.019
February	Inv. Gaussian (3P)	60.7	56.6	0.016	77.8	0.005
March	Wakeby	69.4	69.0	0.021	74.1	0.009
April	Johnson SB	109.1	100.8	0.011	129.7	0.001
May	Fatigue Life (3P)	154.5	147.8	0.018	186.9	0.007
June	Dagum	107.6	91.4	0.010	123.6	0.006
July	Gen. Extreme Value	109.0	75.1	0.011	157.0	0.006
August	Gen. Pareto	38.6	36.1	0.019	46.0	0.013
September	Gen. Pareto	72.5	62.1	0.012	94.2	0.003
October	Wakeby	82.0	78.0	0.017	88.8	0.010
November	Wakeby	63.5	58.0	0.013	72.7	0.006
December	Gen. Extreme Value	85.1	87.1	0.023	89.3	0.020

*X1: Highest observation; X(p=0.98): The probabilistic estimate determined by using the inverse cumulative distribution function for p=0.97826; X2: Forecasted extreme precipitation

Table 4. Output table summarizing the polynomial fits and the statistical measures for the month of June for station 18-003 after 12 highest observations for each month are removed from the input data

JUNE							
Pol.Deg.:	2	3	4	5	6	7	8
a_0	0.0176	0.0020	0.0002	0.0000	0.0000	0.0000	0.0000
a_1	0.7431	-0.1175	-0.0162	-0.0005	0.0000	0.0000	0.0000
a_2	2.9390	3.2557	0.4245	0.0111	0.0005	0.0001	0.0000
a_3		-7.2179	-2.4289	-0.0553	-0.0140	-0.0042	0.0009
a_4			7.1456	0.8672	0.2423	0.0670	-0.0206
a_5				1.3130	-0.5971	-0.4583	0.2747
a_6					3.2714	2.0250	-1.8428
a_7						0.4715	6.1469
a_8							-3.1902
Max	95.10	95.10	95.10	95.10	95.10	95.10	95.10
EstMax	71.93	79.70	88.91	91.90	92.66	93.47	92.73
EstNextMax	74.27	84.42	98.79	104.62	106.58	109.38	105.72
PredIntU(n-1)	84.48	90.29	94.37	96.56	97.46	98.34	97.63
PredIntL(n-1)	59.37	69.12	83.44	87.23	87.86	88.59	87.83
PredIntU(n)	87.01	95.49	104.81	110.18	112.96	116.82	114.56
PredIntL(n)	61.53	73.36	92.77	99.06	100.20	101.94	96.87

Table 4. Continued

Pol.Deg.:	2	3	4	5	6	7	8
Goodness of Fit Measures for the Whole Series:							
MSE	**30.5169**	19.2320	4.5448	2.9596	2.8440	2.6943	**2.5366**
NMSE	**0.0673**	0.0424	0.0100	0.0065	0.0063	0.0059	**0.0056**
RMSE	**5.5242**	4.3854	2.1318	1.7204	1.6864	1.6414	**1.5927**
NRMSE	**0.2595**	0.2060	0.1001	0.0808	0.0792	0.0771	**0.0748**
MAE	**3.9968**	3.4997	1.7008	1.2483	1.2448	1.2084	**1.1288**
MBE	**0.0000**	0.0000	0.0000	0.0000	0.0000	0.0000	**0.0000**
r	**0.9649**	0.9781	0.9949	0.9967	0.9968	0.9970	**0.9971**
d	**0.9311**	0.9566	0.9897	0.9933	0.9936	0.9939	**0.9943**
E	**0.9311**	0.9566	0.9897	0.9933	0.9936	0.9939	**0.9943**
MaxAE	**23.1737**	15.3962	6.1934	5.7482	5.7831	**5.5643**	6.0303
MASE	**1.8930**	1.6575	0.8055	0.5912	0.5896	0.5723	**0.5346**
Goodness of Fit Measures for the Highest 5 Values:							
MSE	**127.2844**	52.8151	13.8332	**11.1934**	11.3853	11.7789	12.1521
NMSE	**0.7347**	0.3048	0.0798	**0.0646**	0.0657	0.0680	0.0701
RMSE	**11.2820**	7.2674	3.7193	**3.3456**	3.3742	3.4320	3.4860
NRMSE	**0.8571**	0.5521	0.2826	**0.2542**	0.2564	0.2607	0.2648
MAE	**8.4804**	4.7540	3.1171	**2.7116**	2.8329	2.9437	2.9430
MBE	**7.0343**	3.0175	0.5262	**0.4096**	0.5197	0.7180	0.5309
r	**0.9458**	0.9506	0.9569	0.9597	0.9609	**0.9632**	0.9591
d	**0.8945**	0.9037	0.9157	0.9210	0.9234	**0.9277**	0.9200
E	**0.0817**	0.6189	0.9002	**0.9192**	0.9179	0.9150	0.9123
MaxAE	**23.1737**	15.3962	6.1934	5.7482	5.7831	**5.5643**	6.0303
MASE	**0.9475**	0.5312	0.3483	**0.3030**	0.3165	0.3289	0.3288

4.3. Forecasting the extreme values for the remaining 65 precipitation stations

The above discussion was generated based on estimations and observations for a single station (18-003). A method's ability to estimate values for a single station is not sufficient to claim that it will be successful in estimating values for other data series. To test the success of the application of the proposed polynomial regression approach across multiple stations, the presented software was used to estimate extreme precipitation values for 66 stations across Turkey. For each station, expected future extreme precipitations were estimated for each month, and the performance of polynomial regression was tested by estimating the removed extremes (a total of 792 observed extremes) as explained above. The scatterplots between the forecasts of the future extremes and the observed maximums for the investigated 65 stations (01-004 to 26-005) are provided as an online supplementary material (*Fig. S1*). The link to the online supplement is at the end of the manuscript above the References.

The success of polynomial regression in forecasting extremes might only be validated when the expected precipitations occur in real life, but testing the performance by removing and estimating the observed extremes has been used in literature as a reliable practice. The scatterplots and the correlations between the removed observed highest precipitation values and the estimations of the fitted polynomial functions show that the presented method is very successful in approximating the removed precipitation values with high accuracy. The closeness of the points in the scatterplots to the straight diagonal $y = x$ line (not shown in the figures) is an indicator of the estimation performance; the closer the points to the $y = x$ line are, the higher the performance is.

The figures show that the majority of the estimations are very close to the observations as it is indicated also by the correlations. 80.3% (53/66) of the correlations are over 0.9, while 65.2% (43/66) are over 0.95, and 15.2% (10/66) are over 0.99. The highest correlation (0.998) was obtained for station 03-027 and the lowest correlation (0.554) was observed for the station 01-004. In addition to correlation, the RMSE, NRMSE, E, and MAE measures between the removed and estimated precipitations are also calculated for all stations (*Table 5*). The values are calculated for the removed and forecasted 12 values for each month of year for each station as shown in the scatterplots in the supplementary materials.

Table 5. Statistical performance measures for the forecasts of the removed highest observations for the first ten stations. The complete table for all stations is provided as an online supplement because of space restrictions.

STATION	01-004	01-005	01-008	02-004	02-009	02-011	02-012	03-009	03-013	03-027
r	0.554	0.921	0.794	0.979	0.963	0.928	0.969	0.974	0.995	0.998
RMSE	29.592	16.897	21.283	24.045	27.685	23.139	30.992	24.823	8.137	2.111
NRMSE	0.233	0.120	0.149	0.126	0.112	0.164	0.131	0.166	0.050	0.015
NSE	-0.434	0.824	0.365	0.894	0.872	0.795	0.910	0.856	0.988	0.994
MAE	13.389	6.935	13.090	16.963	17.728	11.805	19.750	14.334	5.631	1.644

https://www.dropbox.com/scl/fi/n7yd7jos4x00e0cgenh44/Table.5.Performance.Measures.-Full.Version.Supplement.xlsx

Fig. 4 shows the annual averages of the observed and estimated precipitations together with the differences between observations and estimations. In the annual scale, highest future precipitation increases are expected in the southern and northwest shoreline regions. Figures for the monthly observations, estimations, and differences are presented as an online supplement in *Figs. S2-S13*. Those figures clearly show the expected variation of precipitation throughout the seasons.

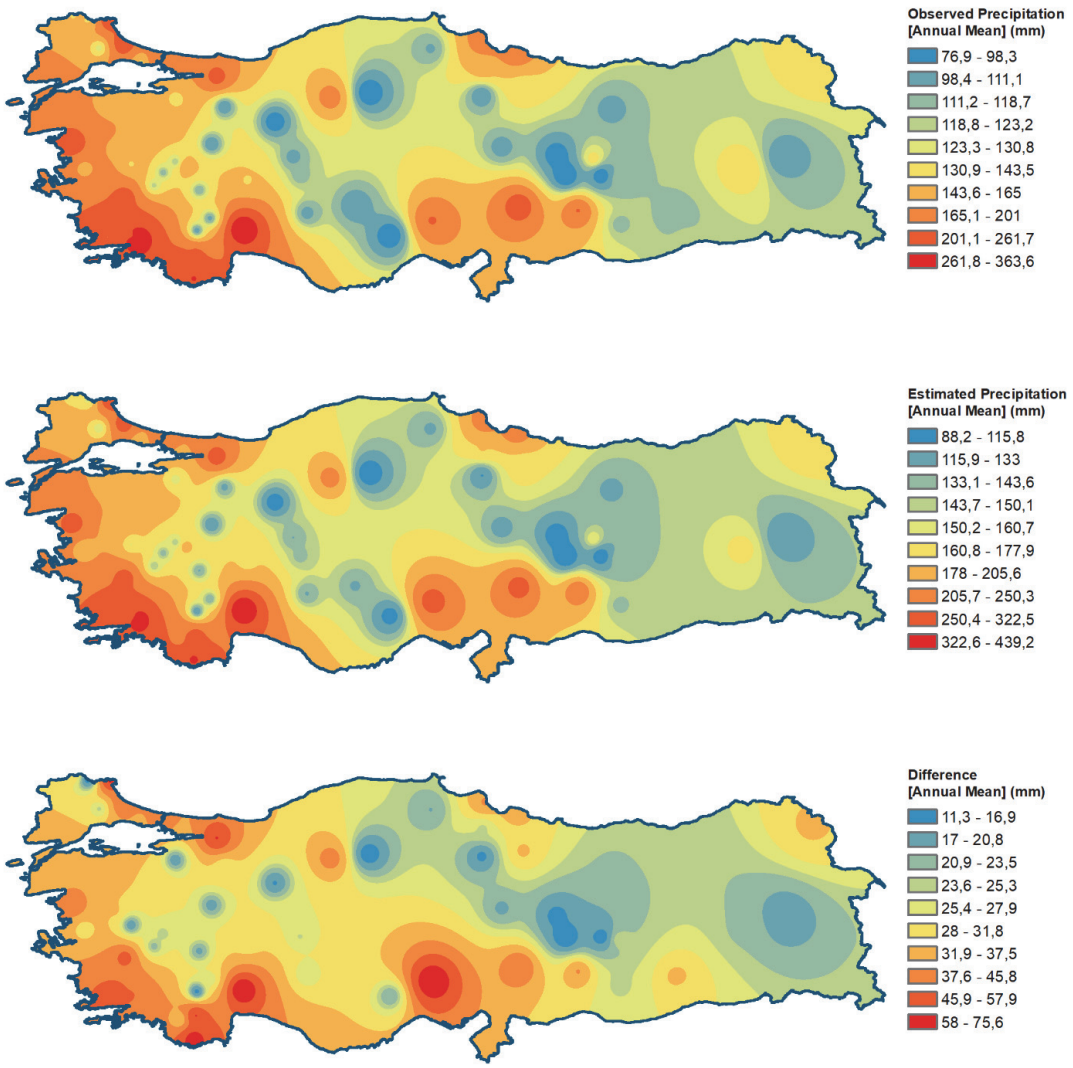

Fig. 4. Annual averages of the observed precipitation, the estimated higher precipitation, and the difference between the observed and the estimated higher precipitation.

In the winter months, extreme precipitation is mostly expected in the western half of the country, while the southern half has a risk of extreme precipitation in spring. Then, in summer, the extreme precipitation expectation moves toward the eastern half of the country, and finally, in autumn months, the northern half of the country seems to be prone to extreme precipitation. These findings fit well with the previously observed extremes showing a rotating motion over the country through the seasons.

Even though polynomials of a given degree provide an advantage in allowing the data to determine the fitted model in a somewhat more flexible way, polynomial regression sometimes suffers from various drawbacks. One disadvantage is that individual observations can exert an influence on remote parts of the curve (*Green* and *Silverman* 1993), and the polynomial regression may

suffer from severe extrapolation problems. It is also well-known that a polynomial with a higher order may fit the data better, but it may produce very weird estimators, especially for extrapolation.

In the presented case, the scatterplots show that a few highly unexpected precipitation values cause the decreases in correlations. For example, for station 01-004, the highest August precipitation was 149.3 mm (the 4th highest observation of the station among 492 observations), while all the remaining 40 August observations are under 44 mm. This very high precipitation value observed in summer was underestimated by all the fitted polynomials, and the correlation value decreased. Another difficulty in polynomial regression is that the model elaboration implicit in increasing the polynomial degree cannot be controlled continuously (*Green* and *Silverman* 1993). Therefore, the users of the presented method and software should be warned that some very rare extreme values might influence the performance of the model negatively, and both the statistical validations and the graphical outputs should be checked comparatively, especially for extremely rare cases. If these shortcomings of polynomial regression are experienced at an unacceptable rate in future uses, then the users might consider some sophisticated models, including local linear regression and penalized regression, which regulate the smoothness of the estimated mean structure but have their shortcomings like collinearity, sparsity, curse of dimensionality, and biased coefficient estimates.

The estimation performance of the presented method is directly associated with the amount and quality of the available data as it has a data-driven methodology. Therefore, it is expected that an even better performance should be expected when the data covers a longer range of observations and contains more information about the behavior of precipitation. An advantage of the method is that the best-fitting polynomials might be updated always as new observations are made. This approach allows consideration of the non-stationary structure of precipitation time series and provides a better opportunity for forecasting record values. A worldwide increase is reported in extreme precipitation values with the influence of global warming causing increased temperatures and evapotranspiration even in locations with decreasing precipitation trends. Consequently, there is an increasing requirement for alternative methods of estimating precipitation extremes showing extensive increases in frequency.

5. Conclusions

The method presented in this paper uses polynomial regression to forecast the most probable future monthly total precipitation exceeding all the previously observed precipitations for each month of the year. The method is applied to observations of 66 precipitation stations in Turkey. The results show that polynomial regression applied for the first time in literature with an approach as

presented in this paper can estimate expected precipitation extremes with high accuracy at a 95% confidence level. This result was obtained by removing and estimating the highest observed precipitations for each month of year for all stations. It is anticipated that the presented methodology and software might contribute to the overall improvement in the skill of forecasting extreme precipitation and other variables with similar behavior. Knowing probable future extreme precipitation values and their locations will allow us to take precautions against hazards in areas likely to experience these extremes, and loss of lives and property will hopefully be prevented.

The polynomial regression software developed for implementing the method is provided freely together with this manuscript. The software is distributed under the terms of the GNU General Public License version 3, and a copyright notice is provided at the beginning of the code.

Data Availability Statement:
Due to confidentiality agreements, supporting data can only be made available to bona fide researchers subject to a non-disclosure agreement. Details of the data and how to request access are available at https://www.turkiye.gov.tr/devlet-su-isleri-hidrometrik-veri-talebi.

Link to the online supplement containing Figures S1-S12:
https://www.dropbox.com/scl/fi/9ka40g5fsfiliakszifdt/Forecasting-Extreme-Precipitations.-Supplement.docx

Link to the code of PolReg software:
https://www.dropbox.com/scl/fi/iq3t8y3we3tikq8nmps79/PolReg.m

References

Acock, M.C. and *Y.A. Pachepsky*, 2000: Estimating missing weather data for agricultural simulations using group method of data handling. *J. Appl. Meteorol. 39*, 1176–1184. https://doi.org/10.1175/1520-0450(2000)039<1176:EMWDFA>2.0.CO;2

Adnan, S., K. Ullah, and *G. Shouting*, 2016: Investigations into precipitation and drought climatologies in south central Asia with special focus on Pakistan over the period 1951-2010. *J. Climate 29*, 6019–6035. https://doi.org/10.1175/JCLI-D-15-0735.1

Baxevani, A. and *R. Wilson*, 2018: Prediction of catastrophes in space over time. *Extremes 21*, 601–628. https://doi.org/10.1007/s10687-018-0314-z

Beguería, S. and *S.M. Vicente-Serrano*, 2006: Mapping the hazard of extreme rainfall by peaks over threshold extreme value analysis and spatial regression techniques. *J. Appl. Meteorol. Climatol 45*, 108–124. https://doi.org/10.1175/JAM2324.1

Bhatia, N., V.P. Singh, and *K. Lee*, 2019: Variability of extreme precipitation over Texas and its relation with climatic cycles. *Theor. Appl. Climatol. 138,* 449–467. https://doi.org/10.1007/s00704-019-02840-w

Block, P. and *B. Rajagopalan*, 2007: Interannual variability and ensemble forecast of upper Blue Nile basin Kiremt season precipitation. *J. Hydrometeorol.8*, 327–343. https://doi.org/10.1175/JHM580.1

Dikbas, F., 2016a: Frequency Based Prediction of Buyuk Menderes Flows. *Teknik Dergi 27*, *1*, 7325–7343.

Dikbas, F., 2016b: Three-dimensional imputation of missing monthly river flow data. *Scientia Iranica 23*, *1*, 45–53. https://doi.org/10.24200/sci.2016.2096

Dikbas, F., 2017a: A novel two-dimensional correlation coefficient for assessing associations in time series data. *Int. J. Climatol. 37*, 4065–4076. https://doi.org/10.1002/joc.4998

Dikbas, F., 2017b: Frequency-based imputation of precipitation. *Stoch. Environ. Res. Risk Asses. 31*, 2415–2434. https://doi.org/10.1007/s00477-016-1356-x

Dikbas, F., 2018a: A New Two-Dimensional Rank Correlation Coefficient. *Water Resour. Manage. 32*, 1539–1553. https://doi.org/10.1007/s11269-017-1886-0

Dikbas, F., 2018b: Compositional Correlation for Detecting Real Associations Among Time Series. In (eds. *Z. Prof. Yildirim, PhD, Gece Kitaplığı*) Academic Researches in Mathematic and Sciences, 27–46.

Fowler, H.J., D. Cooley, S.R. Sain, and *M. Thurston*, 2010: Detecting change in UK extreme precipitation using results from the climateprediction.net BBC climate change experiment. *Extremes 13*, 241–267. https://doi.org/10.1007/s10687-010-0101-y

Gao, L., J. Huang, X. Chen, Y. Chen, and *M. Liu*, 2017: Risk of extreme precipitation under nonstationarity conditions during the second flood season in the Southeastern Coastal Region of China. *J. Hydrometeorol.18*, 669–681. https://doi.org/10.1175/JHM-D-16-0119.1

George, J., L. Janaki, and *J. Parameswaran Gomathy*, 2016: Statistical Downscaling Using Local Polynomial Regression for Rainfall Predictions – A Case Study. *Water Resour. Manage. 30*, 183–193. https://doi.org/10.1007/s11269-015-1154-0

Goodale, C.L., J.D. Aber, and *S. V. Ollinger*, 1998: Mapping monthly precipitation, temperature, and solar radiation for Ireland with polynomial regression and a digital elevation model. *Climate Res. 10*, 35–49. https://doi.org/10.3354/cr010035

Green, P.J., and *B.W. Silverman*, 1993: Nonparametric Regression and Generalized Linear Models: A roughness penalty approach. Taylor & Francis. https://doi.org/10.1201/b15710

Hou, A.Y. and *Coauthors*, 2014: The global precipitation measurement mission. *Bull. Amer. Meteorol. Soc. 95*, 701–722. https://doi.org/10.1175/BAMS-D-13-00164.1

Hwang, Y., M. Clark, B. Rajagopalan, and *G. Leavesley*, 2012: Spatial interpolation schemes of daily precipitation for hydrologic modeling. *Stoch. Environ. Res. Risk Asses. 26*, 295–320. https://doi.org/10.1007/s00477-011-0509-1

Ibrahim, M.N., 2019: Generalized distributions for modeling precipitation extremes based on the L moment approach for the Amman Zara Basin, Jordan. *Theor. Appl. Climatol*. 138, 1075–1093 https://doi.org/10.1007/s00704-019-02863-3

Kent, C., R. Chadwick, and *D.P. Rowell*, 2015: Understanding Uncertainties in Future Projections of Seasonal Tropical Precipitation. *J. Climate 28*, 4390–4413. https://doi.org/10.1175/JCLI-D-14-00613.1

Keupp, L., E. Hertig, I. Kaspar-Ott, F. Pollinger, C. Ring, H. Paeth, and *J. Jacobeit*, 2019: Weighted multi-model ensemble projection of extreme precipitation in the Mediterranean region using statistical downscaling. *Theor. Appl. Climatol*. *138*, 1269–1295. https://doi.org/10.1007/s00704-019-02851-7

Knox, J. C., 1993: Large increases in flood magnitude in response to modest changes in climate. *Nature 361*, *6411*, 430–432. https://doi.org/10.1038/361430a0

Kyselý, J., and *R. Beranová*, 2009: Climate-change effects on extreme precipitation in central Europe: uncertainties of scenarios based on regional climate models. *Theor. Appl. Climatol*. *95*, 361–374. https://doi.org/10.1007/s00704-008-0014-8

Lazoglou, G., C. Anagnostopoulou, K. Tolika, and *F. Kolyva-Machera*, 2019: A review of statistical methods to analyze extreme precipitation and temperature events in the Mediterranean region. *Theor. Appl. Climatol*. *136*, 99–117. https://doi.org/10.1007/s00704-018-2467-8

Leconte, J., F. Forget, B. Charnay, R. Wordsworth, and *A. Pottier*, 2013: Increased insolation threshold for runaway greenhouse processes on Earth-like planets. *Nature 504*, *7479*, 268–271. https://doi.org/10.1038/nature12827

Li, F., X. Ju, W. Lu, and *H. Li*, 2019: A comprehensive analysis of spatial and temporal variability of extreme precipitation in the Nenjiang River Basin, Northeast China. *Theor. Appl. Climatol*. *138*, 605-616. https://doi.org/10.1007/s00704-019-02846-4

Liu, X., and *P. Coulibaly*, 2011: Downscaling ensemble weather predictions for improved week-2 hydrologic forecasting. *J. Hydrometeorol. 12*, 1564–1580. https://doi.org/10.1175/2011JHM1366.1

McElroy, T., 2016: On the measurement and treatment of extremes in time series. *Extremes 19*, 467–490. https://doi.org/10.1007/s10687-016-0254-4

Meena, H.M., D. Machiwal, P. Santra, P.C. Moharana, and *D.V. Singh*, 2019: Trends and homogeneity of monthly, seasonal, and annual rainfall over arid region of Rajasthan, India. *Theor. Appl. Climatol*. *136*, 795–811. https://doi.org/10.1007/s00704-018-2510-9

Porporato, A. and *L. Ridolfi*, 1998: Influence of weak trends on exceedance probability. *Stoch. Hydrol. Hydraul. 12*, 1–14. https://doi.org/10.1007/s004770050006

Rai, P., A. Choudhary, and *A.P. Dimri*, 2019: Future precipitation extremes over India from the CORDEX-South Asia experiments. *Theor. Appl. Climatol*. 137, 2961–2975. https://doi.org/10.1007/s00704-019-02784-1

Reager, J.T. and *J.S. Famiglietti*, 2009: Global terrestrial water storage capacity and flood potential using GRACE. *Geophys. Res. Lett. 36*, *23*. https://doi.org/10.1029/2009GL040826

Schliep, E.M., D. Cooley, S.R. Sain, and *J.A. Hoeting*, 2010: A comparison study of extreme precipitation from six different regional climate models via spatial hierarchical modeling. *Extremes 13*, 219–239. https://doi.org/10.1007/s10687-009-0098-2

Schönwiese, C.-D., J. Grieser, and *S. Trömel*, 2003: Secular change of extreme monthly precipitation in Europe. *Theor. Appl. Climatol*. *75*, 245–250. https://doi.org/10.1007/s00704-003-0728-6

Sivakumar, B., 2000: Chaos theory in hydrology: Important issues and interpretations. *J. Hydrol. 227*, 1–20. https://doi.org/10.1016/S0022-1694(99)00186-9

Sivakumar, B., S.Y. Liong, C.Y. Liaw, and *K.K. Phoon*, 1999: Singapore rainfall behavior: Chaotic? *J. Hydrol. Engin.4*, 38–48. https://doi.org/10.1061/(ASCE)1084-0699(1999)4:1(38)

Stefanescu, V., S. Stefan, and *F. Georgescu*, 2014: Spatial distribution of heavy precipitation events in Romania between 1980 and 2009. *Meteorol. Appl. 21*, 684–694. https://doi.org/10.1002/met.1391

Su, B., Z.W. Kundzewicz, and *T. Jiang*, 2009: Simulation of extreme precipitation over the Yangtze River Basin using Wakeby distribution. *Theor. Appl. Climatol*. *96*, 209–219. https://doi.org/10.1007/s00704-008-0025-5

Tian, D., C.J. Martinez, W.D. Graham, and *S. Hwang*, 2014: Statistical Downscaling Multimodel Forecasts for Seasonal Precipitation and Surface Temperature over the Southeastern United States. *J. Climate 27*, 8384–8411. https://doi.org/10.1175/JCLI-D-13-00481.1

Trömel, S. and *C.-D. Schönwiese*, 2007: Probability change of extreme precipitation observed from 1901 to 2000 in Germany. *Theor. Appl. Climatol*. *87*, 29–39. https://doi.org/10.1007/s00704-005-0230-4

Unkašević, M., I. Tošić, and D. Vujović, 2004: Variability and probability of annual and extreme precipitation over Serbia and Montenegro. *Theor. Appl. Climatol*. *79*, 103–109. https://doi.org/10.1007/s00704-004-0060-9

Wang, X. L., and A. Lin, 2015: An algorithm for integrating satellite precipitation estimates with in situ precipitation data on a pentad time scale. *J. Geophys. Res. Atmospheres 120*, 3728–3744. https://doi.org/10.1002/2014JD022788

Wilks, D. S., 2012: Projecting "Normals" in a Nonstationary Climate. *J. Appl Meteorol Climatol.52*, 289–302. https://doi.org/10.1175/JAMC-D-11-0267.1

Yuan, Z., Z. Yang, D. Yan, and J. Yin, 2017: Historical changes and future projection of extreme precipitation in China. *Theor. Appl. Climatol*. *127*, *1*, 393-407. https://doi.org/10.1007/s00704-015-1643-3

Zhang, D., T. Wang, Y. Liu, S. Zhang, and X. Meng, 2021: Spatial and temporal characteristics of annual and seasonal precipitation variation in Shijiazhuang region, north China. *Environ. Earth Sci. 80*, *18*, 656. https://doi.org/10.1007/s12665-021-09949-0

Zhang, Q., J. Zhang, D. Yan, and Y. Wang, 2013: Extreme precipitation events identified using detrended fluctuation analysis (DFA) in Anhui, China. *Theor. Appl. Climatol*. *117*, 169–174. https://doi.org/10.1007/s00704-013-0986-x

Zhang, Q., C.-Y. Xu, H. Tao, T. Jiang, and *Y.D. Chen*, 2009: Climate changes and their impacts on water resources in the arid regions: a case study of the Tarim River basin, China. *Stoch. Environ. Res. Risk Asses. 24*, 349–358. https://doi.org/10.1007/s00477-009-0324-0

Zhao, Y., X. Xu, W. Huang, Y. Wang, Y. Xu, H. Chen, and Z. Kang, 2019: Trends in observed mean and extreme precipitation within the Yellow River Basin, China. *Theor. Appl. Climatol*. *136*, 1387–1396. https://doi.org/10.1007/s00704-018-2568-4