

## Investigation of impact of climate change on small catchments using different climate models and statistical approaches

Aytac Guven <sup>\*</sup>, Abdulhadi Pala  and Mohamad Sheikhvaisi 

Civil Engineering Department, Gaziantep University, Gaziantep, Turkey

\*Corresponding author. E-mail: aguven@gantep.edu.tr

 AG, 0000-0002-1935-1375; AP, 0000-0003-3090-2171; MS, 0000-0002-1605-0040

### ABSTRACT

The use of a statistical downscaling technique is needed to investigate the hydrological consequences of climate change on the local hydropower capacity. Global Circulation Models (GCMs) are crucial tools used in various simulations for potential climate change effects, including precipitation and temperature. Statistical downscaling methods comprise the improvement of relations between the large-scale climatic parameters and the local variables. This study presents the trend analysis of the observed variables compared to the statistically downscaled emission scenarios that are adopted from the Canadian Second Generation Earth Systems Model (CanESM2) in the basin of Göksu River which is located in Turkey. The key purpose of the research is to evaluate both the predicted monthly precipitation and the projections of GCMs within the three simulated scenarios of RCP2.6, RCP4.5, and RCP8.5 by Gene Expression Programming (GEP). In addition, the findings of statistical downscaling of monthly mean precipitation will be compared to the Linear Regression (LR) model. The R-value is 0.827 and 0.755 for precipitation of the GEP model for the periods of calibrating and validation. In comparison with the LR model for the validation and calibration periods (1971–2005), the results of the GEP model prove its applicability in projecting the data for monthly mean rainfall. Generally, in the simulated periods of 2021–2100, the mentioned scenarios forecast a decline in the monthly mean precipitation in the basin. Moreover, the scenario of RCP8.5 projected more suitably for the case study than expected under the scenarios RCP4.5 and RCP2.6. The mean statistically downscaled CanESM2 model was compared with the trend analysis of the areal mean precipitation over the case study area, and the trend was shown as decreasing. However, the RCP 8.5 scenario was the more quasi-asymptotic for trend.

**Key words:** GCM, GEP, precipitation, statistical downscaling, trend analysis

### HIGHLIGHTS

- Climate change impact is applied in catchment hydrology.
- New methodology is proposed.
- Future projections of water resources are done.
- Artificial intelligence is used.
- Statistical downscaling is cooperated with artificial intelligence.

## 1. INTRODUCTION

Over different time scales, the climate might vary in many ways. At this time, the issue of global warming is worrying scientists. Here, it is worth mentioning that the problem of global warming is the result of some human activities that increase the natural greenhouse impacts. In the next century, the climate will witness dangerous changes as humans continue their activities that increase greenhouse gases. This will result in influencing many areas of our life, such as affecting the runoff of the river catchment along with hydroelectric power stations. All of these need efforts to measure the impacts of such a change, for example to show computer-based modeling of how the electric power and hydroelectric power sectors will be massively influenced by climate change (Harrison *et al.* 1998). Based on Dibike's and Coulibaly's study (2005), the Global Circulation Models (GCMs) access on a worldwide scale the numerical/physical features in the atmosphere, and the system of land cover. Moreover, the GCMs have been designed in such a way to reproduce the climatic variables of the present, the past, and the future. The best methods presently offered for simulating the climate change system's reactions to the growing

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (<http://creativecommons.org/licenses/by/4.0/>).

greenhouse gases are the numerical models (GCMs) that describe physical procedures in the cryosphere, atmosphere, ocean, and even on the land's surface (IPCC 2013). GCMs reflect physical mechanisms in the atmosphere, oceans, and ground surface and are presently able to predict the global climate system's response to growing concentrations of greenhouse (Hulme & Carter 1999). GCM outputs cannot generate local climate details in fine spatial resolution due to uncertainty in the spatial resolution between the GCM and hydrological models (Hadipour *et al.* 2013). Subsequently, downscaling can be defined as a method that delivers from the variables of the large-scale atmosphere to the local-scale climate at the level of the land, and was developed to transform the GCM outputs from coarse spatial resolution to a finer spatial resolution (Anandhi *et al.* 2013).

Two primary techniques, namely empirical, also called statistical, and dynamical downscaling, have been developed for downscaling. Statistical downscaling adopts statistical relationships between the regional climate and carefully selected large-scale parameters (Wilby *et al.* 2004), while dynamical downscaling makes use of regional climate models (RCMs) to simulate finer scale physical processes consistent with the large-scale weather evolution prescribed from a GCM, and is more expensive (Mearns *et al.* 2013). Zorita & von Storch (1999) stated that statistical downscaling rests on the principle that the local climate's relation with the large-scale movement continues to be effective through the varying future climatic conditions.

Gene Expression Programming (GEP), Genetic Algorithms (GAs) and Genetic Programming (GP) are algorithms that use populations of individuals, chosen according to fitness, and introduces genetic differences with the help of one or more genetic operators (Rechenberg 1973). The central difference between the three algorithms is in the nature of the individuals: in GAs the individuals are linear strings of fixed length (chromosomes); in GP, the individuals are nonlinear entities of different sizes and shapes (parse trees); and in GEP, the individuals are encoded as linear strings of fixed length (the genome or chromosomes) that are afterwards expressed as nonlinear entities of different sizes and shapes (i.e., simple diagram representations or expression trees). GEP was designed by Ferreira in 1999 (Ferreira 2001) and it includes both the simple, linear chromosomes of fixed length parallel to the ones used in genetic algorithms GAs and the ramified structures of different sizes and shapes like the parse trees of GP. This is equivalent to say that in GEP, the genotype and phenotype are lastly disjointed and the system can now benefit from all the merits this brings about. GEP is commonly applied in hydrologic engineering, particularly in the last decade. Especially, it has been applied in the prediction of hydro-meteorological variables (Güven 2009; Güven & Aytekin 2009; Güven & Talu 2010; Seckin & Güven 2012; Traore & Güven 2013; Al-Juboori & Güven 2016).

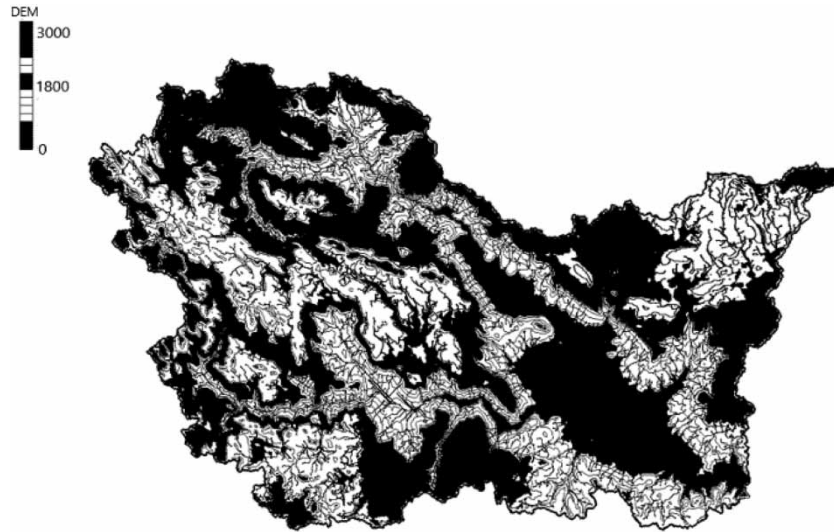
The Mann-Kendall (MK) test, which was initially advanced by Mann (1945), and the test statistic distribution was subsequently derived by Kendall (1975), is a rank-based non-parametric test for identifying a monotonic trend in a time series Yue *et al.* (2002). Hydro-meteorological data (such as rainfall, streamflow, temperature, etc.) generally have a skewed distribution; it is more appropriate for non-parametric methods than parametric methods for trend detection. Therefore, the MK test has been generally utilized to evaluate the statistical significance of trends in these data series (Önöz & Bayazit 2003; Modarres & da Silva 2007; Kumar *et al.* 2010).

The key purpose of the research was to evaluate both the predicted monthly precipitation and the projections of the GCM within the three simulated scenarios by GEP. Moreover, for the observed climatic data between 1971 and 2005, this study has also used the MK trend analysis and the Sen slope estimator to compare the different pathways RCP2.6, RCP4.5, and RCP8.5 for the period 2021–2100.

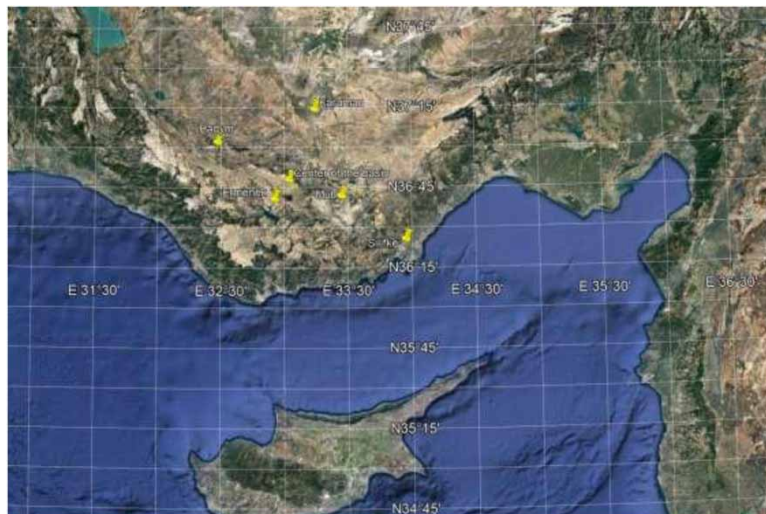
## 2. STUDY AREA AND DATA USED

The study area is the Göksu River basin, which is one of the sub-basins of the Dogu Akdeniz basin. The basin is placed in the southern part of Turkey, near the Mediterranean coast; more specifically, it is 95 km from Mersin city. Moreover, the area of this basin is approximately 17,048 km<sup>2</sup>. In addition, the site of a flow gauge station, named Karahacılı, is used for the outlet point of the basin. Furthermore, the map shows that the coordinates of the mentioned basin are 36°24'06.5" N latitude and 33°48'56.1" E longitude. Figures 1 and 2 show the location of the study area according to the Göksu River basin. There are five local gauge stations that are going to be used in the study: Ermenek, Mut, Karman, Hadim, and Silifke. In addition, Tables 1 and 2 offer essential information about the locations and properties of the above-mentioned gauge stations.

The monthly areal mean precipitation was classified into two periods: the calibration and the validation periods. The former period, which lasted from 1971 to 1995, was used for developing statistical downscaling models, while the latter period, which was from 1996 to 2005, was used for examining the model's performance and compare downscaling results.



**Figure 1** | The boundary of the Göksu River basin and Digital Elevation Model (DEM) contours.



**Figure 2** | Location of the stations and centroid of the basin.

**Table 1** | Properties of the gauge stations

	Station no.	Station name	Latitude (N)	Longitude (E)
A	18,210	ERMENEK	36°38'01.0"	32°54'27.0"
B	17,956	MUT	36°39'05.0"	33°26'02.0"
C	17,246	KARAMAN	37°11'35.5"	33°13'12.7"
D	17,928	HADIM	36°59'21.5"	32°27'20.5"
E	17,330	SILIFKE	36°22'56.6"	33°56'14.3"

The mean precipitation data of these five stations were converted to the first dataset (predictand set) and areal mean precipitation ( $P_M$ ) of the basin, which was measured by using the Thiessen polygons technique. The second dataset is the large-scale predictor variables data, taken from the Canadian Center for Climate Modeling and Analysis scenarios (CCCMA),

**Table 2** | Summary of the data used in this study for each station

Item	Silifke	Mut	Hadim	Ermenek	Karaman
Mean annual rainfall (mm)	549.83	375.16	638.10	477.53	326.12
Mean monthly rainfall (mm)	46.25	31.43	53.42	40.04	27.34
Maximum annual rainfall (mm)	1,007.70	608.70	1,074.20	778.80	478.30
Minimum annual rainfall (mm)	300.50	135.80	421.30	268.00	212.60
Maximum monthly rainfall (mm)	504.50	248.80	405.80	289.00	144.10
Duration of data (year)	34	34	34	34	34
Number of datasets (days)	12,775	12,775	12,775	12,755	12,755

where various GCMs and future scenarios have been evolved. This study has chosen only the Canadian Second Generation Earth System Model (CanESM2) with RCP 2.6, RCP 4.5, and RCP 8.5 scenarios. These have been also obtained from the Canadian climate data and scenarios official website (<http://climate-modelling.canada.ca/climatemodeldata/cgcm4/CanESM2/>). The GCM helps us to get a digital depiction of the climate system of the Earth. This model is designed to have vertical levels and a horizontal mesh of grid boxes for inland, ocean, and atmosphere. The modeling process for every GCM may encompass many points, such as the size of the field, as the large areas are recognized more than the smaller ones. Along with the size, the field's site also matters as the agreement level among GCMs outputs varies a lot from one place to another.

The process of presenting the access data from the website was done through the use of grid cells. Every grid cell represents a web neighboring the corresponding model grid points. This mechanism works by input of decimal latitude and longitudinal coordinates of any location. In this study, we used the centroid of the basin; when it is entered, and the data retrieved. Moreover, these variables consist of 26 variable datasets (predictors) that were inserted in the CanESM2 folders. The Canadian climate website offers the variables of the model in a daily time scale data. Applying the downscaling process to large-scale data, the variables of the daily 26 GCM had to be transformed to the corresponding maximum monthly mean data as in the predictand set. Table 3 shows the description of the 26 large-scale weather factors.

### 3. METHODOLOGY

The large-scale weather factors (26 input sets) have been used as predictors, and have been given with the assistance of the CanESM2. Also, as a predictand (output), the areal mean precipitation was used on the basin. For the groups of the data, it is important to mention that a monthly time scale is the basis of all data groups. Both data classes are split into two subsets i.e. calibration and validation. The first is the period between 1971 and 1995 for calibration, and the second is the period of validation between the years 1996 and 2005. Moreover, to forecast the downscaled monthly areal mean precipitation in the basin, the GEP techniques have been applied. Additionally, this research also compares the results of the GEP with those of Linear Regression (LR). Moreover, the MK trend analysis and Sen's slope estimator of  $P_M$  over the case study area were compared with the different pathways RCP2.6, RCP4.5, and RCP8.5 for 2021–2100.

#### 3.1. Selection of the most effective predictors

What is regarded to be a fundamental part of the statistical downscaling methods is the selection of the large-scale weather GCM factors (predictors), 26 variables. Based on certain devices, variables have been updated to maximize the association with predictor variables and predictands. Hence, the predictand and the predictors' datasets have been changed to Ln (taking the natural logarithm of each data) variables. Using Pearson's rank correlation coefficient technique, correlation analysis has been utilized to select the most powerful predictors and to analyze the linear relationship between inputs (26 different large-scale weather factors) and outputs (areal mean precipitation) separately; moreover, as predictors for downscaling models, variables with a higher correlation coefficient were chosen. This study uses predictors with a correlation of  $R > 0.5$  in order to refer to the typical situation of the variables of predictor and predictands. Tables 3 and 4 show the description of the 26 variables and variables of  $R > 0.5$  (9 in this case) which are the uppermost correlating GCM variables and areal mean precipitation separately; the correlation coefficient is bigger than 0.5. The degree of interaction between variables is identified by the size of the correlation coefficient, and the signal shows if the relation is direct or contrary.

**Table 3** | The description of the 26 CanESM2 variables (predictors)

No.	CanESM2 predictors	Description	Short name
1	ceshmslpgl	Mean sea level pressure	V1
2	ceshp1_fgl	1,000 hPa Wind speed	V2
3	ceshp1_ugl	1,000 hPa Zonal velocity	V3
4	ceshp1_vgl	1,000 hPa Meridional velocity	V4
5	ceshp1_zgl	1,000 hPa Vorticity	V5
6	ceshp1thgl	1,000 hPa Wind direction	V6
7	ceshp1zhgl	1,000 hPa Divergence	V7
8	ceshp5_fgl	500 hPa Wind speed	V8
9	ceshp5_ugl	500 hPa Zonal velocity	V9
10	ceshp5_vgl	500 hPa Meridional velocity	V10
11	ceshp5_zgl	500 hPa Vorticity	V11
12	ceshp5thgl	500 hPa Wind direction	V12
13	ceshp5zhgl	500 hPa Divergence	V13
14	ceshp8_fgl	850 hPa Wind speed	V14
15	ceshp8_ugl	800 hPa Zonal velocity	V15
16	ceshp8_vgl	800 hPa Meridional velocity	V16
17	ceshp8_zgl	800 hPa Vorticity	V17
18	ceshp8thgl	800 hPa Wind direction	V18
19	ceshp8zhgl	800 hPa Divergence	V19
20	ceshp500gl	Relative humidity at 500 hPa	V20
21	ceshp850gl	Relative humidity at 850 hPa	V21
22	ceshprcpgl	Total rainfall	V22
23	ceshs500gl	Specific humidity 500 hPa	V23
24	ceshs850gl	Specific humidity 500 hPa	V24
25	ceshshumgl	Surface-specific humidity	V25
26	ceshtempgl	Mean temperature at 2 m height	V26

**Table 4** | Summary of final most effective GCM set predictors

No.	GCM predictors	Description	R	No.	GCM predictors	Description	R
V1	ceshmslpgl	Mean sea level pressure	0.634	V20	ceshp500gl	Relative humidity at 500 hPa	-0.539
V5	ceshp1_zgl	1,000 hPa vorticity	-0.581	V24	ceshs850gl	Specific humidity 500 hPa	-0.577
V7	ceshp1zhgl	1,000 hPa divergence	0.509	V25	ceshshumgl	Surface-specific humidity	-0.566
V17	ceshp8_zgl	800 hPa vorticity	-0.571	V26	ceshtempgl	Mean temperature at 2 m height	-0.662
V19	ceshp8zhgl	800 hPa divergence	0.596				

Nine large-scale weather factors: Mean sea level pressure (V1), 1,000 hPa Vorticity (V5), 1,000 hPa Divergence (V7), 800 hPa Vorticity (V17), 800 hPa Divergence (V19), Relative humidity at 500 hPa (V20), Specific humidity 500 hPa (V24), Surface-specific humidity (V25), Mean temperature at 2 m height (V26), were used as the input data in LR model.

### 3.2. Downscaling using GEP technique

To get the answer to the problem in GEP, there are four essential stages. Defining a series of functions is the first important step in this research, and a set of functions (+, -, \*, /, sqrt,  $x^2$ ,  $x^3$ ,  $x^4$ ,  $x^5$ , 3Rt, 4Rt, 5Rt, Ln) was chosen in the



GEP program. Identifying the chromosome structure, which embraces the number of genes for each chromosome and the magnitude of the gene, is the second crucial stage for the GEP program. Up to this point, it is worth stating that with the aid of four genes per chromosome for every GEP model, we achieved the best outcomes. The process of choosing the function's link is the third important stage. Concerning this study, '+' (addition) was the chosen linking function. The fitness measure is the last stage, and here it is important to note that the training set's Root-Relative Square Error (RRSE) is used as a fitness function. In order to calibrate the parameters and equations of the downscaling model by GEP, the local precipitation (monthly areal mean precipitation) and large-scale weather data of the calibration duration have been utilized. Moreover, to produce the validation forecast future predictions, the GEP model, which automatically picks the effective predictors from nine and totally eight inputs, also improved the procedure of calibration equations which is given in Equation (1). Mean sea level pressure (V1), 1,000 hPa Vorticity (V5), 1,000 hPa Divergence (V7), 800 hPa Vorticity (V17), 800 hPa Divergence (V19), Relative humidity at 500 hPa (V20), Surface-specific humidity (V25), Mean temperature at 2 m height (V26), eight large-scale weather factors were used as a input data in GEP model to produce Equation (1). The definition of each variable is given in Table 4:

$$\begin{aligned}
 Y = & \text{Sqrt} (5Rt (((3Rt (((-1.311 + [V5]) - [V24])) \times (-[V7]))^2))) \\
 & + \left( \frac{\left( (-11.320 \times [V19]) + \left( \frac{[V17]}{3.186} \right) \right) \times -10.082}{([V26]^3)} \left( (6.218[V5])^5 \right) \right) \\
 & + \left( 5Rt \left( \left( (5Rt (7.959) + [V1]) - [V1] \right) + \left( ([V25]^4)^2 \right) \right) \right)^4 \\
 & + 5Rt(\log (((5Rt ([V20] + 4.081))(3Rt ([V7]) \times ([V26] - [V25]))^2)^3)))
 \end{aligned} \tag{1}$$

As presented here, Rt is the root, × is the multiplication process between the data values, and Y represents the areal mean precipitation of basin in the equation.

### 3.3. Trend analysis

For the observed climatic data between 1971 and 2005, this study uses the MK trend analysis and the Sen's slope estimator to compare the different pathways RCP2.6, RCP4.5, and RCP8.5 for 2021–2100.

The descriptive statistics of the observed data including the number of observations and the CanESM2 model under the different scenarios of RCP2.6, RCP4.5, and RCP8.5, mean standard deviation, maximum and minimum values are listed in Table 5.

With the help of the MK test and Sen's slope, the detection of the trend was evaluated, and the MK Statistic S for the trend and the trend's magnitude were calculated using the following Equations (2)–(4):

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sgn} (X_j - X_i) \tag{2}$$

**Table 5** | The descriptive statistics of the variables

Variable	Observations (years)	Average (mm)	Minimum (mm)	Maximum (mm)	Standard deviation (mm)
Observed (1971–2005)	35	481.783	278.280	776.227	100.008
RCP2.6 (2021–2100)	80	341.312	332.098	350.526	5.420
RCP4.5 (2021–2100)	80	320.297	275.545	365.050	26.328
RCP8.5 (2021–2100)	80	281.669	203.190	360.148	46.169

The values of the sequential data are  $x_i$  and  $x_j$ , while the length of the database is presented in 'n':

$$T_i = \frac{X_j - X_k}{J - K} \text{ for } i = 1, 2, 3, \dots, N \quad (3)$$

As presented here,  $T_i$  is the size of  $i$  tie group while,  $X_j$  and  $X_k$  are the data values for  $j$  and  $k$  times for the period as  $j > k$ . For every observation, the slope is projected. The average is calculated from  $N$  observations of the slope to predict the slope for Sen estimator:

$$Q_i = T \frac{N+1}{2} \text{ for } N \text{ is odd} \quad (4)$$

$$= \frac{1}{2} \left( T \frac{N}{2} + T \frac{N+1}{2} \right) \text{ for } N \text{ is even}$$

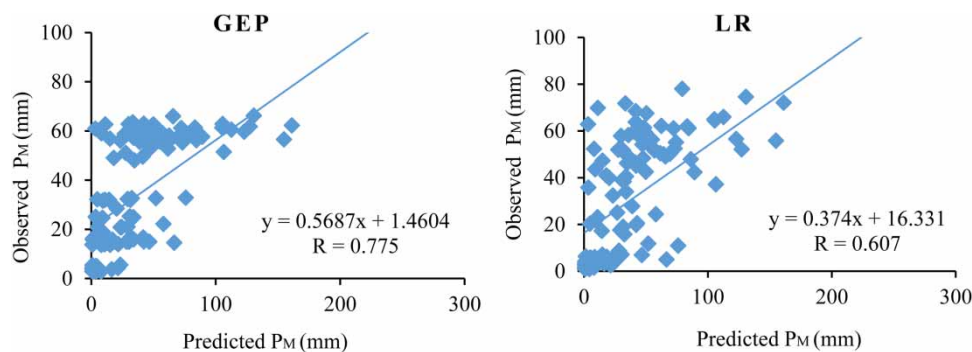
To attain the correct slope for the non-parametric test in the series, the two-sided test is accomplished at  $100(1 - \alpha)\%$  of the 0.05 confidence interval get the correct slope for a non-parametric test in the series (Mondal *et al.* 2012). Furthermore, in the form of falling and rising trends, the negative and the positive slope of  $Q_i$  is achieved.

#### 4. RESULTS AND DISCUSSION

Large-scale weather factors (26 input sets) were used as predictors which were obtained from CanESM2. The  $P_M$  was used as the predictand (output). All datasets are based on a monthly time scale. Both datasets were divided into two groups: calibration and validation. The first group is the calibration period ranging between 1971 and 1995. The second group is the validation period ranging between 1996 and 2005. GEP downscaling technique was utilized to predict the downscaled  $P_M$  of the basin. Furthermore, the prediction results of downscaling techniques were compared with each other to examine the best model performance. Also, LR was used for comparison of the proposed nonlinear models with a linear model. Moreover, the trend analysis of  $P_M$  over the case study area was compared by the mean of the statistically downscaled CanESM2.

The GEP model provided the greatest consequences depending on the performance of the offered models in the calibration period. For every model gained from the calibration times, the validation process was conducted using a simpler formulation. Equation (1), is used with the GEP model in order to generate the validation process. The findings of the models GEP and LR are presented in terms of  $R$  during the validation period. As seen in Figure 3, a scatter plot was generated between the observed (vertical axis-Y) and the predicted  $P_M$  (horizontal axis-X). When moving from left to right, as shown in Figure 3, a downhill pattern will accompany all models, and this implies, according to the observed data, that models have a great validation. At this level, it is worth saying that the LR model had the lowermost  $R$ -value (0.607), whereas the uppermost  $R$ -value (0.775) was for the GEP model.

Moreover, Table 6 compares the simulated  $P_M$  for the GEP and LR models and the observed  $P_M$  within the same validation period 1996–2005. As shown in Table 6, the GEP model's simulated mean monthly areal precipitation is 34.322 mm, whereas it is 30.763 mm for the LR model; this indicates that the means of the above models are less than the observed mean. We can say that the models underestimate the means. Nevertheless, the GEP model produced better results to simulate the mean than the results of the LR model. This leads to a state so that on the one hand, the model of GEP proves to operate well for the



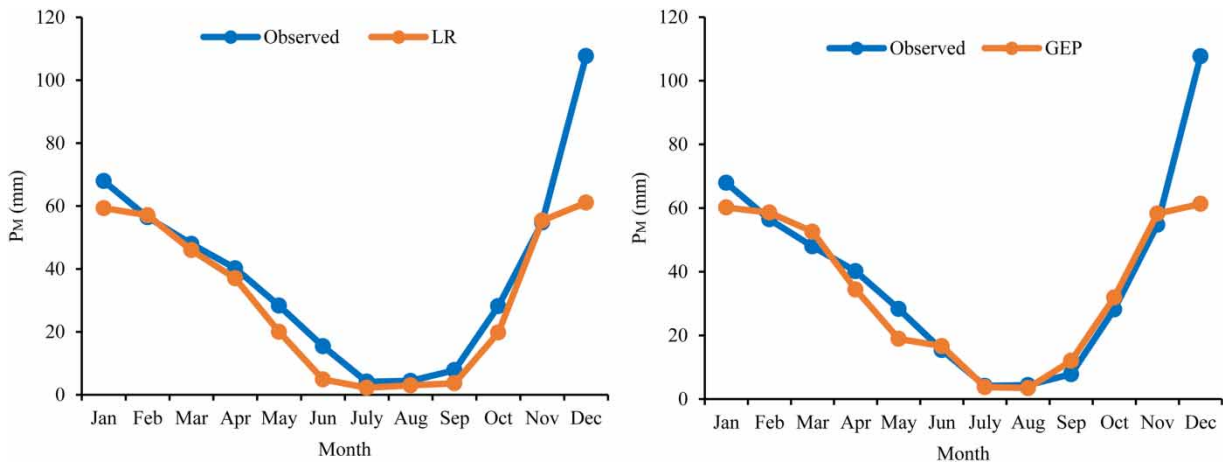
**Figure 3** | Scatter plot of the observed and predicted  $\ln P_M$  values for the validation period 1996–2005 by GEP and LR models.

**Table 6** | Comparison of statistical results of observed and simulated  $P_M$  during the validation (1996–2005)

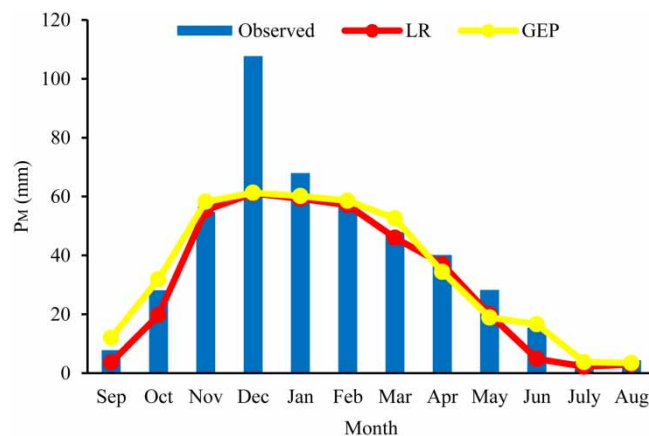
Data	Mean (mm)	Min (mm)	Max (mm)	Std. deviation (mm)	MAE (mm)	RMSE (mm)	R	R <sup>2</sup>
Observed	38.589	1.000	265.254	40.666	–	–	–	–
GEP model	34.322	2.865	66.184	23.172	19.231	31.896	0.775	0.600
Linear regression	30.763	0.934	78.036	25.014	19.871	33.091	0.607	0.369

simulated minimum precipitation; on the other hand, the inability to perform in a good way has been clearly shown in LR's simulated minimum precipitation. Here, it is important to mention that there is a major difference in the maximum value of simulated  $P_M$  when it is compared with the observed data. As for the maximum precipitation, a minimum difference of 18% has been recorded by the model of LR. This gives the LR model the feature of having the minimum underestimating difference of 8% when comparing it to the GEP model. Moreover, the GEP model has a lower value of mean absolute error (MAE) (19.231 mm) and root mean square error (RMSE) (31.896 mm) than the LR model, as shown by the numerical values given in Table 6. Additionally, the LR's MAE is nearly 3.3% and RMSE is about 3.7%, and in both cases, these are greater than those of the GEP model.

The GEP's, LR's, and observed data simulated  $P_M$  during the period of validation is presented in Figure 4. The simulated  $P_M$  of all models along with the observed data is clarified in Figure 5. Examining that figure, it is noticeable that, in August and July, the  $P_M$  has been underestimated by the GEP and LR models. However, the LR and GEP models have overestimated the



**Figure 4** | Observed and simulated mean monthly  $P_M$  for the validation period 1996–2005 by GEP and LR models.



**Figure 5** | Observed and simulated mean monthly  $P_M$  for the validation period 1996–2005 by all models.



$P_M$  in February and November as it is displayed in Figure 8. In contrast, the LR and the GEP models have misjudged the  $P_M$  in January, March, April, May, and December. The percentage of the LR's underestimated mean is 5.7% in May, which is smaller than that of the GEP model, as shown in Figure 8. Nevertheless, the GEP and LR models recorded the same precipitation and the best expectation in January, August, and December. However, the LR's prediction of mean precipitation in February and November proved to be the best in comparison with the observed data. In comparison with the LR model for the periods of validation and calibration, the GEP model has a good performance in anticipating the distinctions in monthly areal mean precipitation, as is clear in Figure 4, Figure 5 and Table 5.

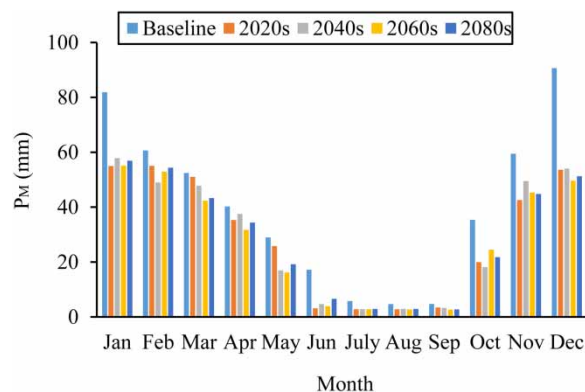
Concerning the periods of validation offered in Table 6, the GEP model demonstrated its excellence in these periods with the uppermost correlation coefficient and slightly lowermost in the values for RMSE and MAE. All of this leads us to assume that the GEP model will be suitable to use under future emission scenarios RCP 2.6, RCP 4.5, and RCP 8.5, the  $P_M$  between the years of 2021 and 2100.

## 5. FUTURE PROJECTION OF $P_M$ UNDER DIFFERENT EMISSION SCENARIOS

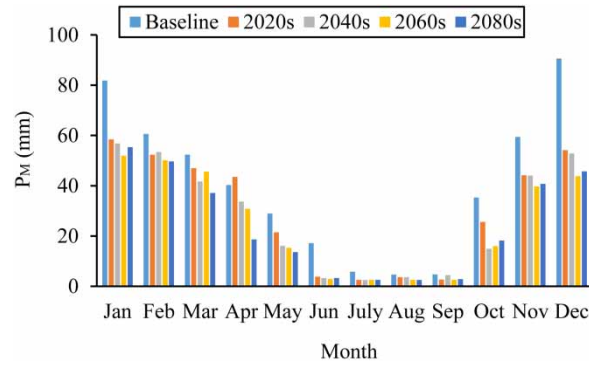
The classifications of the downscaled outcomes from scenarios RCP 2.6, RCP 4.5, and RCP 8.5 had four periods with a range of 20 years. The four durations are 2020s (2021–2040), 2040s (2041–2060), 2060s (2061–2080), and 2080s (2081–2100). A comparison was made between them and the period for the baseline (1971–2005) in order to study the upcoming change in  $P_M$  of the case study area. The GEP's different predictions of  $P_M$  under the CanESM2 scenario for various periods are provided in Figure 6; thus, it is noticeable that, when compared with the baseline mean precipitation, we find that the prediction of GEP's amount of precipitation for each different period will be reduced. Again, in comparison with the baseline mean precipitation, the predictions of precipitation for each different period provided via the GEP model will be lessened in the 2020s, 2040s, 2060s, and 2080s, and increased in April 2020s as displayed in Figure 7. Shifting to Figure 8, there will be a decrease in GEP's predicted precipitation when comparing it with the precipitation of the baseline.

Under the scenarios RCP 2.6, RCP 4.5, and RCP 8.5, there will be a decline in the monthly mean precipitation  $P_M$  of each month for the years of the 2020s, 2040s, 2060s, and 2080s.

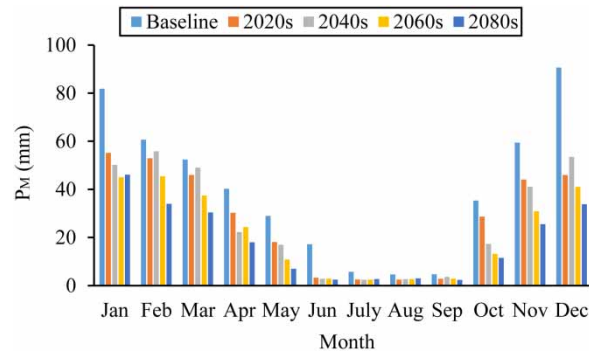
Moving to the scenario RCP2.6, the  $P_M$  will be reduced in July. In addition, according to that scenario, the decrease in the  $P_M$  will be 51.6% for the 2020s, 48.5% for the 2040s, 51.3% for the 2060s, and 49.705% for the 2080s. According to the RCP4.5 scenario, during the month of June, there will be an obvious decrease; whereas the  $P_M$  of the 2020s, 2040s, 2060s, and 2080s will be 77.7%, 81.2%, 82.9%, and 80.7%, respectively. Also, in June, the monthly mean precipitation of the 2020s, 2040s, 2060s, and 2080s will show a decline by 80.6%, 83.4%, 83%, and 85.3%, respectively, under the RCP8.5 scenario. As also offered in Figures 6–8, the  $P_M$  will decrease through the months of June, July, August, and September throughout time spans of the 2020s, 2040s, 2060s, and 2080s; this is according to the scenarios RCP2.6, RCP4.5, and RCP8.5. Moving to the scenario RCP2.6, projected mean precipitation is very near to the observed baseline mean precipitation for the duration of March and September. Similarly, under RCP4.5, the projected mean precipitation in August is also neighboring the observed baseline mean precipitation. As shown in Figures 6 and 7, all the precipitation predictions via the mentioned scenarios differ in size; nevertheless, their patterns are similar. Moreover, in the same figures, an annual decrement of the baseline period (1971–2005) is presented under all scenarios.



**Figure 6** | The projection of mean monthly  $P_M$  under CanESM2-RCP 2.6 by the GEP model.



**Figure 7** | The projection of mean monthly  $P_M$  under CanESM2-RCP 4.5 by GEP model.

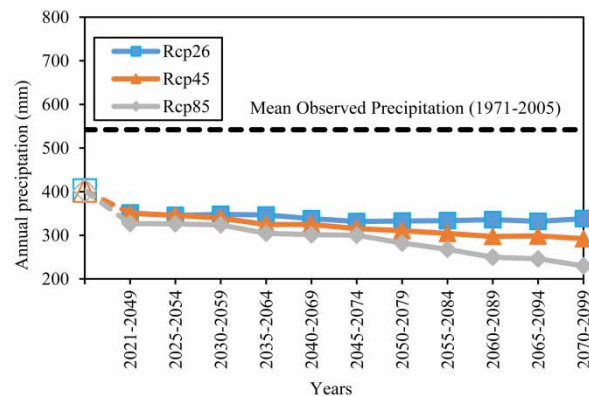


**Figure 8** | The projection of mean monthly  $P_M$  under CanESM2-RCP 8.5 by GEP model.

Concerning the scenarios RCP2.6, RCP4.5, and RCP8.5 presented in Figure 9, they show the expected  $P_M$  for various periods by suggested downscaled models. Thus, we can notice from Figure 9 that, under all scenarios from the period of 2021–2049 to the period of 2070–2099, the annual precipitation of the GEP model is the lowest. In Figure 9, the mean annual precipitation, which is downscaled by the models in the baseline period 1971–2005, is represented by small points, while the mean observed precipitation is represented by the dotted straight line.

### 5.1. Trend analysis of $P_M$ over the case study area by mean of statistically downscaled CanESM2 future climate projection

The trend analysis of  $P_M$  over the case study area during the periods of 1971–2005 was calculated by the MK test and Sen’s slope estimator using XLSTAT software, as well as the same for the CanESM2 model under the different scenarios RCP2.6, RCP4.5, and RCP8.5 during the periods 2021–2100. Outcomes of the MK trend test pointed out that there is a statistically



**Figure 9** | The projection of annual  $P_M$  change under CanESM2 RCP 2.6, RCP 4.5, RCP 8.5 scenarios for different periods by the GEP.

substantial decreasing trend in the series of mean yearly  $P_M$ . As the computed  $p$ -value is greater than the significance level  $\alpha = 0.05$ , one cannot reject the null hypothesis  $H_0$ . Moreover, there were statistically significant falling trends for the mean  $P_M$  for all years. Table 7 shows the findings of non-parametric analyses (Kendall's tau, Var (S),  $p$ -value, alpha, and Sen's estimator).

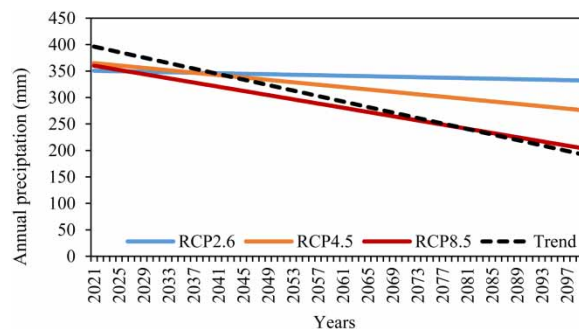
Out of these checks offered in Table 7, it was concluded that they display a downward trend in the total, mean, maximum and minimum precipitation in the study area between 1971 and 2005, but these trends were not statistically significant at 95% confidence level. In addition, the amount of precipitation was less in the period between 1996 and 2005 than in the period between 1971 and 1995. Furthermore, Sen's estimator revealed that annual total precipitations had trends of  $-2.906$  mm/year for a period of 1971–2005. In this section of the study, the trend analysis was used to compare with the series RCP2.6, RCP4.5, and RCP8.5 during the periods 2021–2100, as illustrated in Figure 10.

The annual MK test and Sen's estimator were accomplished at each scenario that was used to compute the observed values over the entire basin. As a result of the trend analysis of  $P_M$  for the periods of 2021–2100, a multidirectional dropping trend was found for each scenario during these periods. The findings of the MK trend test for RCP 2.6 pointed out that there is a statistically significant decreasing trend in the series of mean annual  $P_M$  (as the computed  $p$ -value is higher than the significance level  $\alpha = 0.05$ , one cannot abandon the null hypothesis  $H_0$ ). Moreover, the RCP 4.5 scenario produced a decreasing curve and was slightly closer to the trend for the mean  $P_M$ . However, the RCP 8.5 scenario produced a downward sloping curve and was quasi-asymptotic for trend analysis (as the calculated  $p$ -value is lower than the significance level  $\alpha = 0.05$ , one should discard the null hypothesis  $H_0$ , and agree to the alternative hypothesis  $H_a$ ). Moreover, the scenario RCP8.5 projects more suitable for the case study than expected under the scenarios RCP4.5 and RCP2.6. The trend analysis of  $P_M$  over the case study area was compared using the mean statistically downscaled CanESM2 model, and the trend was shown as decreasing.

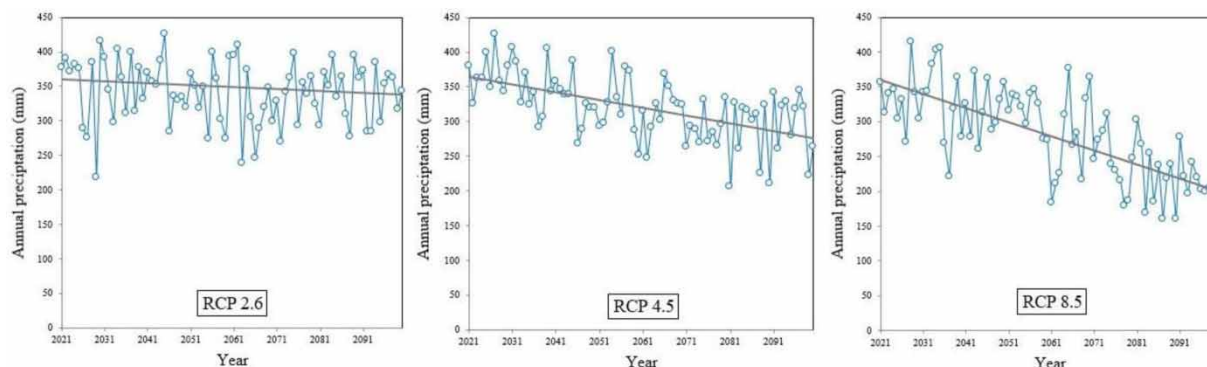
From these analyses shown in Table 7, it was identified that the values have a downward trend in the total of scenarios in the study area between 2021 and 2100, but these trends were not statistically significant at the 95% confidence level. In addition, the values were more than the expected overall mean rainfall and it was drier in the duration between 2070 and 2100 than in the period between 2021 and 2069. Moreover, Sen's estimator revealed that annual total rainfalls for RCP2.6, RCP4.5, and RCP8.5 scenarios have trends of  $-0.282$  mm/year,  $-1.110$  mm/year, and  $-2.007$  mm/year correspondingly, as shown in Figure 11.

**Table 7** | The descriptive statistics of the observed and RCPs scenarios (2021–2100)

Series\Test	Kendall's tau	Var (S)	$p$ -value	Sen's slope
Observed (1971–2005)	-0.230	4,958.333	0.053	-2.906
RCP 2.6 (2021–2100)	-0.101	57,933.33	0.188	-2.906
RCP 4.5 (2021–2100)	-0.395	57,933.33	<0.0001	-1.110
RCP 8.5 (2021–2100)	-0.504	57,933.33	<0.0001	-2.007



**Figure 10** | The trend analysis for RCP2.6, RCP4.5, and RCP8.5 scenarios during the period 2020–2100.



**Figure 11** | Sen's estimator trend for RCP 2.6, RCP4.5, and RCP8.5 for the period 2021–2100.

## 6. CONCLUSIONS

The use of the downscaling method for GEP has been used to forecast the basin's downscaled monthly areal mean precipitation. Furthermore, to analyze the outputs of the model GEP, the outcomes of prediction processes of downscaling methods have been identified and contrasted with LR. Also, the exploration of the study extends further to the downscaling outcomes of the GCM scale predicted with CanESM2 model and RCP2.6, RCP4.5, and RCP8.5 future emission scenarios.

The good application of the model of GEP in the process of evaluating  $P_M$  in the period of calibration (1971–1995) and the period of validation (1996–2005) has been based on the numerical signs between observed and downscaled data. For the calibration period, the GEP's correlation coefficient ( $R$ ) is 0.827 and for the validation period is 0.755.

For the three scenarios, the predicted findings suggest a significant decrease in their patterns; however, they are distinct in the quantity of  $P_M$ . Generally speaking, during the coming century, the three scenarios expect a decline in the average yearly  $P_M$ . For predicting the potential  $P_M$  shift in the region of the basin, the downscaled outcomes of the scenarios RCP2.6, RCP4.5, and RCP8.5 have been split into four 20-year time ranges, these being: 2020s (2021–2040), 2040s (2041–2060), 2060s (2061–2080), and 2080s (2081–2100).

With the help of trend analysis, the modeled scenarios RCP2.6, RCP4.5, and RCP8.5 have been compared with the baseline period (1971–2005). Moreover, to display the trend of the potential variations in  $P_M$  in the region of the basin, the MK test and Sen's slope estimator have been developed. Nevertheless, over the period of 2021–2100, the scenario RCP8.5 indicates a sharper decline in precipitation and projects more suitably for the case study than the scenarios RCP4.5 and RCP2.6.

## DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

## REFERENCES

- Al-Juboori, A. M. & Guven, A. 2016 [Hydropower plant site assessment by integrated hydrological modeling, gene expression programming and visual basic programming](#). *Water Resources Management* **30**, 2517–2530.
- Anandhi, A., Srinivas, V. V. & Kumar, D. N. 2013 Impact of climate change on hydrometeorological variables in a river basin in India for IPCC SRES scenarios. In: *Climate Change Modeling, Mitigation, and Adaptation* (Surampalli, R. Y., Zhang, T. C., Ojha, C. S. P., Gurjar, B., Tyagi, R. D. & Kao, C. M., eds). American Society of Civil Engineers, pp. 327–356.
- Dibike, Y. B. & Coulibaly, P. 2005 [Hydrologic impact of climate change in the Saguenay watershed: comparison of downscaling methods and hydrologic models](#). *Journal of Hydrology* **307** (1–4), 145–163.
- Ferreira, C. 2001 Gene expression programming: a new adaptive algorithm for solving problems. *Complex Systems* **13** (2), 87–129.
- Guven, A. 2009 [Linear genetic programming for time-series modelling of daily flow rate](#). *Journal of Earth System Science* **118** (2), 137–146.
- Guven, A. & Aytek, A. 2009 [New approach for stage discharge relationship: gene-expression programming](#). *Journal of Hydrologic Engineering* **14** (8), 812–820.
- Guven, A. & Talu, N. E. 2010 [Gene expression programming for estimating suspended sediment yield in middle Euphrates Basin, Turkey](#). *Clean-Soil Air Water* **38** (12), 1159–1168.
- Hadipour, S., Shahid, S., Harun, S. B. & Wang, X. J. 2013 Genetic programming for downscaling extreme rainfall events. *1st International Conference on Artificial Intelligence, Modelling and Simulation (AIMS) 2013*. pp. 331–334, IEEE.

- Harrison, G. P., Whittington, H. W. & Gundry, S. W. 1998 Climate change impacts on hydroelectric power. *Proceedings of the 33rd Universities Power Engineering Conference* **1**, 391–394.
- Hulme, M. & Carter, T. R. 1999 Representing uncertainty in climate change scenarios and impact studies. In: *Proceedings of the ECLAT-2 Helsinki Workshop*, 14–16 April. Climatic Research Unit, University of East Anglia, Norwich, United Kingdom, pp. 11–37.
- IPCC 2013 Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., Qin, D. Plattner, G.K. Tignor, M. Kiehl, J.T., Trenberth, K.E. (1997)]. Earth's Annual Global Mean Energy Budget, *Bulletin of the American Meteorological Society*, **78**, pp. 197–208.
- Kendall, M. G. 1975 Rank Correlation Methods. Charles Griffin, London.
- Kumar, V., Jain, S. K. & Singh, Y. 2010 Analysis of long-term rainfall trends in India. *Hydrological Sciences Journal [Journal des Sciences Hydrologiques]* **55** (4), 484–496.
- Mann, H. B. 1945 Non-parametric tests against trend. *Econometrica* **13**, 245–259. NJ Mantua, SR Hare, Y. Zhang, JM Wallace, and RC Francis (1997), A Pacific Decadal.
- Mearns, L. O., Sain, S., Leung, L. R., Bukovsky, M. S., McGinnis, S., Biner, S. & Snyder, M. 2013 Climate change projections of the North American Regional Climate Change Assessment Program (NARCCAP). *Climatic Change* **120** (4), 965–975.
- Modarres, R. & Silva, V. D. P. R. 2007 Rainfall trends in arid and semi-arid regions of Iran. *Journal of Arid Environments* **70** (2), 344–355.
- Mondal, A., Kundu, S. & Mukhopadhyay, A. 2012 Rainfall trend analysis by Mann-Kendall test: a case study of north-eastern part of Cuttack District, Orissa. *International Journal of Geology, Earth and Environmental Sciences* **2** (1), 70–78. ISSN:227-2081.
- Önöz, B. & Bayazit, M. 2003 The power of statistical tests for trend detection. *Turkish Journal of Engineering and Environmental Sciences* **27** (4), 247–251.
- Rechenberg, I. 1973 *Evolutionsstrategie*. Holzmann-Froboog, Stuttgart.
- Seckin, N. & Guven, A. 2012 Estimation of peak flood discharges at ungauged sites across Turkey. *Water Resources Management* **26** (9), 2569–2581.
- Traore, S. & Guven, A. 2013 New algebraic formulations of evapotranspiration extracted from gene-expression programming in the tropical seasonally dry regions of West Africa. *Irrigation Science* **31**, 1–10.
- Wilby, R. L., Charles, S. P., Zorita, E., Timbal, B., Whetton, P. & Mearns, L. O. 2004 Guidelines for use of climate scenarios developed from statistical downscaling methods. In: *Supporting Material of the Intergovernmental Panel on Climate Change*, Available From the DDC of IPCC TGCIA.
- Yue, S., Pilon, P. & Cavadias, G. 2002 Power of the Mann-Kendall and Spearman's rho tests for detecting monotonic trends in hydrological series. *Journal of Hydrology* **259** (1–4), 254–271.
- Zorita, E. & von Storch, H. 1999 The analog method as a simple statistical downscaling technique: comparison with more complicated methods. *Journal of Climate* **12** (8), 2474–2489.

First received 21 January 2021; accepted in revised form 29 October 2021. Available online 5 November 2021