Journal of
Water & Climate Change

IWA PUBLISHING

Journal of Water and Climate Change Vol 15 No 2, 759 doi: 10.2166/wcc.2024.562

Hydroclimatic projection: statistical learning and downscaling model for rainfall and runoff forecasting

Shweta Kodihal [a], M. P. Akhtar [a,*] and Satya Prakash Maurya [b]

[a] Department of Civil Engineering, SCCE, Manipal University Jaipur, Jaipur, India
[b] Department of Computer Science & Engineering, Graphic Era (Deemed to be) University, Dehradun, Uttarakhand, India
*Corresponding author. E-mail: parwez.wrd@gmail.com; parwez.akhtar@jaipur.manipal.edu

SK, 0000-0002-8022-261X; MPA, 0000-0002-3647-6356; SPM, 0000-0002-4320-2038

ABSTRACT

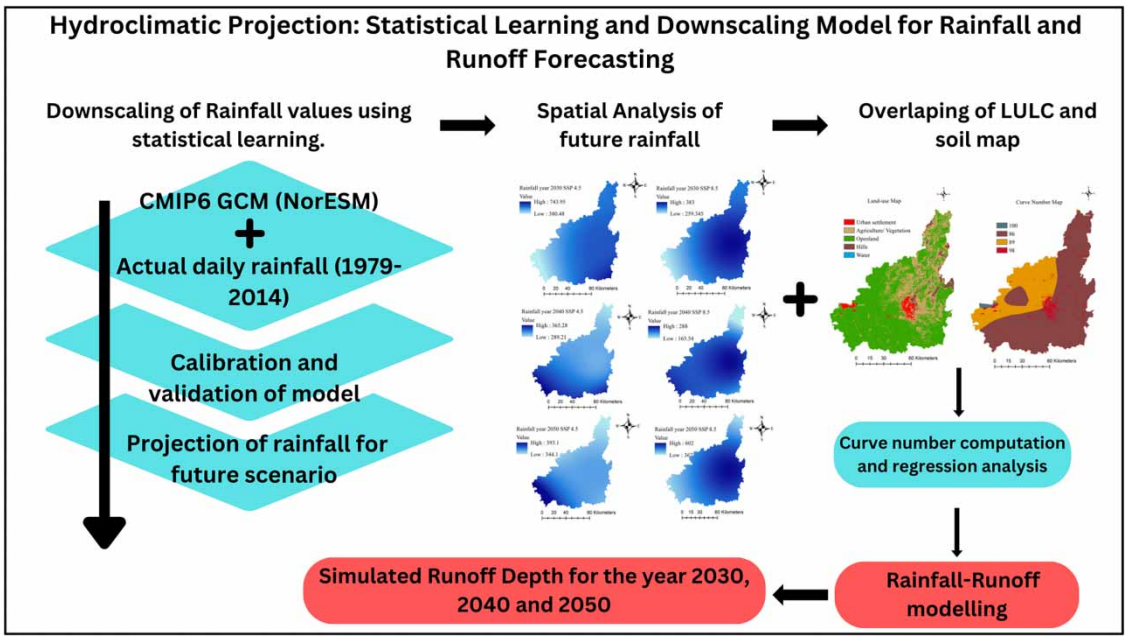
The study is carried out to investigate the surface runoff depth with changing precipitation due to climate change in the study area where sandy loam and loamy soil are dominant. In this study, future rainfall is projected by a statistical downscaling model (SDSM) using a set of predictors derived from a Coupled Model Intercomparison Project Phase 6 (CMIP6) global climate model (GCM) [the Norwegian Earth System Model (NorESM)] with updated scenarios SSP 4.5 and SSP 8.5. Daily rainfall values for the observed period (1981 to 2014) are validated using statistical learning and evaluated with matrices, namely, root mean square error (RMSE), coefficient of correlation, and Nash–Sutcliffe efficiency (NSE), which are found to be valid for further predictions. Rainfall projections show a decrease in rainfall trend of 50% from 2030 to 2040 for scenario SSP 4.5 and an increase of 7% from 2040 to 2050. Predicted rainfall for scenario SSP 8.5 shows a similar trend of decreasing rainfall of 24% for the period 2030–2040 and an increase of 19% in the period 2040–2050. Furthermore, these rainfall values are spatially modelled in a geographic information system (GIS) and rainfall maps are obtained. The obtained rainfall map, land-use map, and soil map are overlaid to compute curve numbers and runoff depths. A similar trend of decrease in runoff is observed for the period 2030–2050. The overall trend of climate change shows a water-stressed scenario.

Key words: climate change, future surface runoff, hydrological assessment, NorESM, SDSM, statistical learning

HIGHLIGHTS

- The study downscaled a CMIP6 GCM to evaluate the effect of climate change on rainfall.
- The study conducts hydrological assessments to predict surface runoff for future scenarios.
- The SDSM model, when subjected to statistical learning analysis, shows good performance in simulating rainfall values.
- The study predicts a decrease in rainfall and runoff in the years 2030, 2040, and 2050.
- Machine learning-boosted SDSM is a strong tool.

GRAPHICAL ABSTRACT



1. INTRODUCTION

Surface runoff is an important hydrological parameter for water management and storage. Surface runoff refers to the movement of water on the surface resulting from precipitation. The most studied problem in hydrology is the prediction of runoff depth, which is influenced by rainfall predictions. The statistical downscaling model (SDSM) offers significant reliability, provides efficient results, and is a widely established model for statistical downscaling of climatic data (Phuong *et al.* 2020; Ma *et al.* 2022; Mahdaoui *et al.* 2023). Statistical downscaling involves establishing a relationship between large-scale climatic models and regional/local-scale climatic models. These large-scale climatic models are derived from Global Climate Models (GCMs), also known as 'predictors'. The GCMs represent the spatial distribution of climate variables across the Earth's surface by dividing it into grid cells. Data from GCMs provide a comprehensive understanding of the past, current, and future climates. The regional or local climate variables are known as 'predictands' (Wilby & Dawson 2013). Once the GCMs perform well in establishing a relationship with the Regional Circulation Models (RCMs), they are subjected to simulation for future scenarios. The Intergovernmental Panel on Climate Change (IPCC) Sixth Assessment Report (AR6) has new versions of the Coupled Model Intercomparison Project (CMIP), with updated scenarios called SSP1-2.6, SSP2-4.5, SSP4-6.0, and SSP5-8.5. The Shared Socioeconomic Pathways (SSPs) are predicated upon five distinct narratives that elucidate overarching socioeconomic patterns capable of exerting influence over the trajectory of future societies. The present set of scenarios has been designed to encompass a comprehensive spectrum of potential future outcomes.

In contrast to weather forecasts, which provide a comprehensive depiction of the anticipated daily progression of atmospheric conditions commencing from the current moment, climate models adopt a probabilistic approach (www.climate.gov). Many researchers have used statistical prediction approaches in their studies to better analyze different climate-related parameters such as drought, catchment modelling, land susceptibility, agricultural production, and soil erosion trends (Huntingford *et al.* 2005; Meenu *et al.* 2013; Chuenchum *et al.* 2020; Hürlimann *et al.* 2022). Ma *et al.* (2022), who compared downscaled climate projections, namely, for precipitation and temperature, state that SDSM results are more efficient compared with the Long Ashton Research Station Weather Generator (LARS-WG). Kreienkamp *et al.* (2020) analyze the difference between CMIP5 and CMIP6 GCMs. Precipitation changes are reportedly less dramatic, and they only differ significantly on a seasonal basis. In another study, Verma *et al.* (2023) compared the performance of CMIP5 and CMIP6 GCMs. It is reported that CMIP6 GCMs provide a reduced margin of error for future climate prediction and more reliability for analysis of the hydrological effects. Hussain *et al.* (2015) state the need to improve the correlation between predictor and predictand for model performance. Kumar *et al.* (2023) conduct a comparative analysis of various machine learning techniques and

suggest adopting an ensemble approach to improve the quality of prediction. Doulabian *et al.* (2021) report uncertainty in projected precipitation assessed using 25 GCMs. Nasidi *et al.* (2021) validate the modelled precipitation by the coefficient of determination, Nash–Sutcliffe efficiency (NSE), percent bias, root mean square error (RMSE), standard error, and mean absolute error.

The runoff curve number is an empirical parameter used to compute the runoff. The curve number method was initially formulated by the United States Department of Agriculture (USDA) Natural Resources Conservation Service, previously known as the Soil Conservation Service (SCS). In scholarly literature, it is commonly referred to as the 'SCS runoff curve number' (SCS-CN). The runoff curve number was derived through an empirical examination of runoff patterns observed in small catchments and hillslope plots monitored by the USDA. The method is extensively employed and demonstrates high efficacy in estimating the approximate depth of direct runoff resulting from a specific rainfall event within a given geographical region. Patil *et al.* (2023) reported the use of SCS-CN to compute the potential of rainwater harvesting sites in the extended urban area of Jaipur. The CN method incorporates several factors, including soil properties, land use, and surface condition, to account for runoff in watersheds. This comprehensive approach enhances the method's applicability and acceptance within the scientific community. Mishra & Singh (2003) present a tabular representation of the CN values for various combinations of soil, vegetation cover, and land use, which has been included in the study. Similarly, Mehta et al. (2023) implemented the Hydrologic Engineering Centre's Hydrologic Modelling System (HEC-HMS) for rainfall–runoff modelling, which shows more accurate flow values when compared with the observed flow values.

As discussed, several related studies have been conducted in the study area; however, proper scientific discussions regarding future trends of rainfall and its impact on projected surface runoff are hardly found in the literature. The information about the availability of water resources is crucial to water resources engineers in planning, management, storage as well as groundwater recharge. Notably, downscaling of climatic parameters provides reliable results for improved quality of rainfall prediction. To address the gap, the current investigation endeavours to downscale the rainfall for a forthcoming period and compute the surface runoff depth with respect to the projected rainfall for a semi-arid water-stressed region. The study incorporates a more recent CMIP6 GCM, namely, the Norwegian Earth System Model (NorESM) for a realistic assessment of the parameters for better prediction. Furthermore, this projected rainfall is employed in an analysis of the runoff resulting in future runoff depths. The present study offers valuable insights for researchers seeking to strategize their approach toward future water utilization. Thus, the objectives of the study are as follows:

1. To construct a statistical learning model using CMIP6 GCM predictors to replicate historical rainfall patterns. To validate the model against observed data and utilize a calibrated model to project future rainfall for a specified period.
2. To analyze spatially the projected daily rainfall for the future period (2030–2050) to discern distinct temporal changes.
3. To integrate a geographic information system (GIS) technique to model spatially the rainfall changes and overlay them on soil and land use/land cover (LULC) maps to compute curve numbers.
4. To investigate the temporal evaluation of runoff depth in climatically sensitive regions focusing on the impact of changing rainfall patterns due to climate change.

2. STUDY AREA

The study area covers Jaipur city and the adjoining rural areas located in the state of Rajasthan in the Northwestern region of India, as shown in Figure 1. The annual rainfall reportedly analyzed for 30 years (1990–2020) is 593 mm per the district report on Jaipur in the national aquifer mapping and management plan document. The normal annual rainfall for the period 1901 to 1971 is 527 mm. Furthermore, rainfall of a declining trend at the rate of 1.18 mm/year is reported (Jaipur district 2022). Rainfall over the timeframe has frequently been recorded as low as 200–300 mm. Notably, the land-use pattern of the study area includes both urban and rural settlements. The soil formation is alluvial sandy loam in the major part of the study area and the remaining part consists of loamy desert soil. The area is naturally drained by six rivers, namely, Sabi, Banganga, Bandi, Mendha, Mashi, and Sota rivers. Rainfall is the prime source of water for these rivers. Hence, it is important to study the runoff variation with changing rainfall patterns.

3. DATA

The predictand (actual daily rainfall) data are acquired from the India Meteorological Department (IMD) https://www.imdpune.gov.in/lrfindex.php in network Common Data Form (NetCDF), gridded ($0.25^{o} \times 0.25^{o}$) in the latitude and longitude

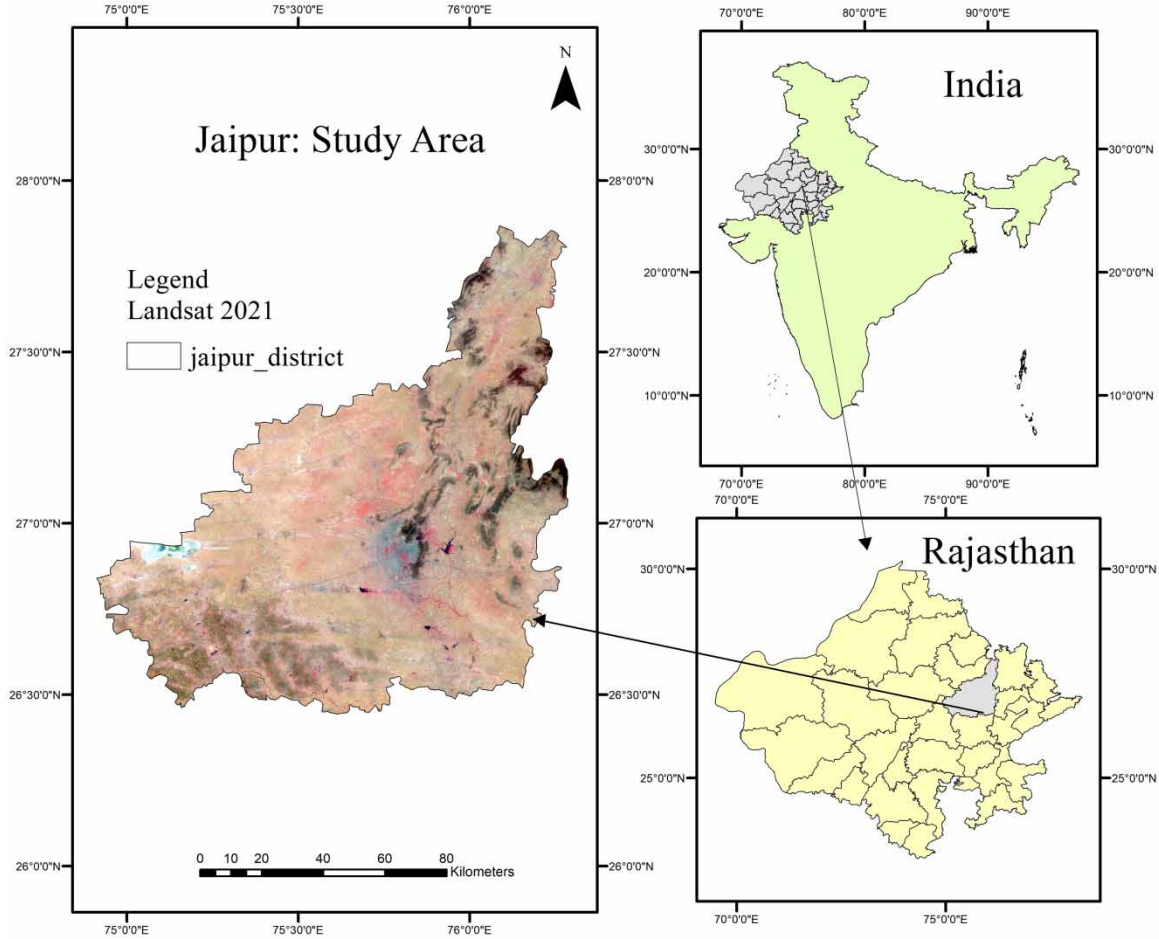


Figure 1 | Location map of the study area.

format and extracted for five stations as presented in Table 1. The actual daily rainfall is considered from 1979 to 2014, which is identical to the period of the climate model. The list of the acquired data for the study is detailed in Table 2.

The NorESM, which has been under development since 2007, has emerged as a pivotal instrument for climate researchers in their comprehensive investigation of historical, contemporary, and prospective climatic conditions (Kreienkamp *et al.* 2020). The NorESM model has made significant contributions to the field of climate simulation, particularly in its application to research evaluated in the IPCC's fifth comprehensive assessment report (Seland *et al.* 2020). The climate model NorESM2-MM used in the present study is presented in Table 3 (www.climate-scenarios.canada.ca). The predictors are derived for SSP2-4.5 and SSP5-8.5 for the period 2015–2100 as provided in CMIP6. SSP2 depicts a future where societal trends largely adhere to historical patterns. It envisages a 'middle-of-the-road world' that is characterized by a continuation of existing

Table 1 | Geographical location of stations for which daily rainfall values (1979–2014) are extracted

Station number	Latitude	Longitude	Average annual rainfall (mm)
1	26.25	74.5	460.93
2	27	76	638.14
3	28	74.5	345.86
4	28	76.5	635.42
5	26.25	76.5	680.57

Table 2 | List of datasets and their sources

S.no	Data	Source/Description
1.	Actual rainfall	IMD Pune https://www.imdpune.gov.in/lrfindex.php
2.	GCMs	Copernicus Climate Change Service https://cds.climate.copernicus.eu/cdsapp#!/dataset/projections-cmip6?tab=form
3.	Landsat for 2021	Landsat 8–9 Operational Land Imager (OLI) and Thermal Infrared (TIRS) Collection 2 Level-2. United States Geological Survey (USGS) https://earthexplorer.usgs.gov
4.	Soil map	Food and Agriculture Organization of the United Nations https://www.fao.org/soils-portal/

Table 3 | CMIP6 predictors name and their corresponding variable ID as provided by https://climate-scenarios.Canada.ca/?page=pred-cmip6-notes

No.	Variable ID	Predictor variable	No.	Variable ID	Predictor variable
1	mslp	Mean sea level pressure	14	p8_f	850 hPa Wind Speed
2	p1_f	1,000 hPa Wind speed	15	p8_u	850 hPa Zonal wind component
3	p1_u	1,000 hPa Zonal wind component	16	p8_v	850 hPa Meridional wind component
4	p1_v	1,000 hPa Meridional wind component	17	p8_z	850 hPa Relative vorticity of true wind
5	p1_z	1,000 hPa Relative vorticity of true wind	18	p8th	850 hPa Wind direction
6	p1th	1,000 hPa Wind direction	19	p8zh	850 hPa Divergence of true wind
7	p1zh	1,000 hPa Divergence of true wind	20	p500	500 hPa Geopotential
8	p5_f	500 hPa Wind speed	21	p850	850 hPa Geopotential
9	p5_u	500 hPa Zonal wind component	22	prcp	Total precipitation
10	p5_v	500 hPa Meridional wind component	23	s500	500 hPa Specific humidity
11	p5_z	500 hPa Relative vorticity of true wind	24	s850	850 hPa Specific humidity
12	p5th	500 hPa Wind direction	25	shum	1,000 hPa Specific humidity
13	p5zh	500 hPa Divergence of true wind	26	temp	Air temperature at 2 m

practices and policies, without significant deviations or transformative changes. SSP5 envisions a future characterized by rapid and unconstrained growth in economic output and energy consumption. This scenario suggests a trajectory where economic expansion takes precedence, potentially leading to increased resource exploitation and environmental degradation (www.carbonbrief.org). The shapefile of the study area is downloaded from https://www.diva-gis.org/. The land-use map is prepared using the Landsat 8–9 Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) images downloaded from the United States Geological Survey (USGS) website https://earthexplorer.usgs.gov.

4. METHODOLOGY

The rainfall prediction has been based on statistical learning that can be seen as following the four steps as shown in Figure 2. The SDSM is a stochastic and hybrid downscaling model developed by Wilby & Dawson (2007). The study employed SDSM for statistical downscaling of the predictors and predictand. The first step in downscaling is to identify the most influential predictors for each station predictand (rainfall). In this study, the potential predictors are selected based on *p*-value and partial correlation analysis (Wilby & Dawson 2007). These carefully selected predictors form the foundation of the prediction model that will provide its accuracy and reliability. In the second step, the model is calibrated for the current period through the production of ensembles. The model is subjected to a comparative analysis with the observed rainfall data, wherein various statistical metrics such as correlation coefficient (Equation (1)), RMSE (Equation (2)), and NSE (Equation (3)) are employed to assess the level of concordance of the model with the actual rainfall patterns.

This evaluation is conducted to ascertain the model's efficacy in accurately capturing the observed rainfall and to determine the extent to which it aligns with the empirical data. In the third phase, the model is executed to generate projections for

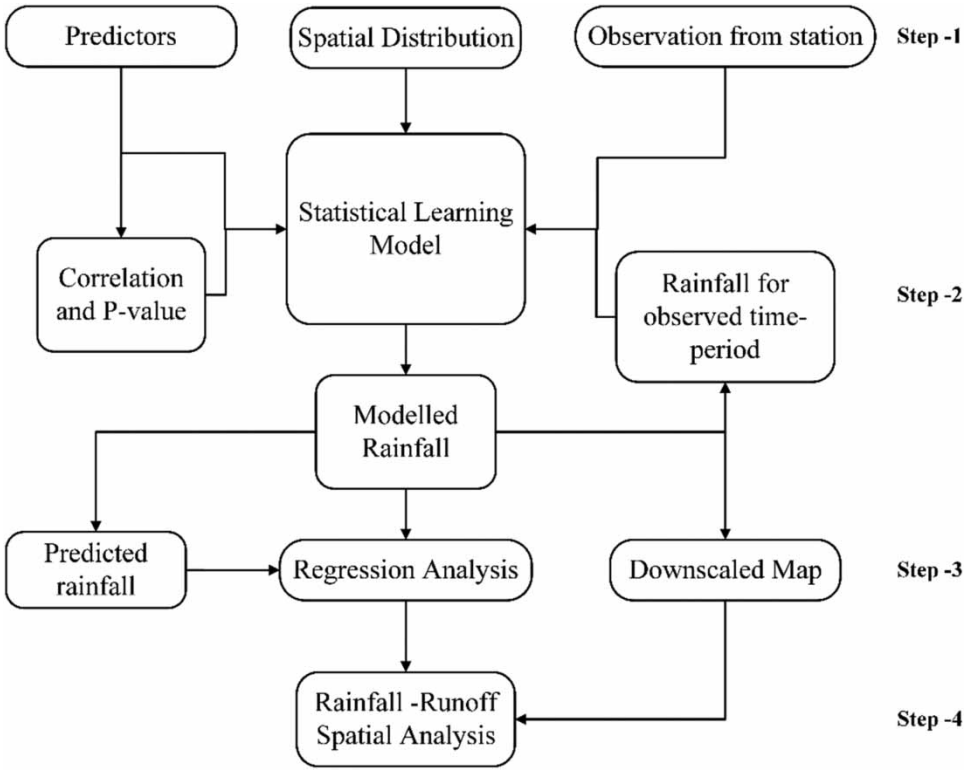


Figure 2 | Workflow of statistical learning-based rainfall prediction.

forthcoming scenarios, predicated upon the chosen predictors. In the fourth phase, spatial analysis of the projected rainfall is performed. Furthermore, the curve number map is prepared by overlaying the land-use map and the soil texture map. In the conclusion phase, surface runoff is computed using Equations (4) and (5). The approach of statistical method in conjunction with spatial analysis enhances the precision and applicability of rainfall projections. This approach advances the field of rainfall prediction, which provides further valuable insights while predicting surface runoff.

$$r = \frac{\sum (P_{\text{obs}} - \overline{P_{\text{obs}}})(P_{\text{mod}} - \overline{P_{\text{mod}}})}{\sqrt{(P_{\text{obs}} - \overline{P_{\text{obs}}})^2 (P_{\text{mod}} - \overline{P_{\text{mod}}})^2}} \tag{1}$$

where r is the correlation coefficient, P_{obs} is the observed rainfall, P_{mod} is the modelled rainfall value, $\overline{P_{\text{obs}}}$ is the mean of the observed rainfall, $\overline{P_{\text{mod}}}$ is the mean of the modelled rainfall.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (P_{\text{mod}} - P_{\text{obs}})^2}{N}} \tag{2}$$

RMSE is the root mean square error, N is the number of values, P_{obs} is the observed rainfall, P_{mod} is the modelled rainfall value.

$$NSE = 1 - \frac{\sum_{i=1}^{N} (P^i_{\text{obs}} - P^i_{\text{mod}})^2}{\sum_{i=1}^{N} (P^i_{\text{obs}} - \overline{P_{\text{obs}}})^2} \tag{3}$$

NSE is the Nash–Sutcliffe efficiency, N is the number of values, P_{obs} is the observed rainfall, P_{mod} is the modelled rainfall value, and $\overline{P_{obs}}$ is the average of the observed rainfall.

The curve number is related to soil moisture retention (S) using the following equation:

$$S = \frac{25{,}400}{\mathrm{CN}} - 254 \tag{4}$$

where S is the potential maximum storage and CN is the curve number.

Runoff quantity is estimated using the equation

$$Q = \frac{(P - 0.2S)^2}{(P + 0.8S)} \tag{5}$$

where Q is the runoff depth in millimetres, P is the precipitation in millimetres, and S is the potential maximum storage.

5. RESULTS AND DISCUSSIONS

5.1. Selection of predictors

The utilization of the correlation analysis is employed in the process of selecting appropriate predictors to construct a model. The selection of predictors is dependent upon the examination of partial correlation coefficients (partial-*r*) and their corresponding *p*-values. The selection process involves identifying the partial-*r* values that exhibit the highest degree of strength, with a particular emphasis on those that possess *p*-values in close proximity to zero (Anandhi *et al.* 2008; Ncoyini-Manciya & Savage 2022). The station-wise selection of predictors along with their partial-*r* and *p*-value are mentioned in Table 4.

5.2. Calibration and validation of the model

The model has been calibrated using the selected predictors to estimate daily rainfall values within the time frame spanning from 1979 to 2014. It is calibrated by taking one predictand and a set of predictors and estimates the parameters of the regression equation using the ordinary least square approach. The model creates various model parameters for each month of the chosen period (1979–2014). Multiple ensembles were employed to compute the mean precipitation of each month, which was subsequently cross-validated against the actual data (Table 5). The examination of the correlation coefficients from Equation (1) across all stations reveals that values ranging from 0.7 to 0.9 denote a strong association between the observed and modelled monthly rainfall, thereby signifying a high level of correlation. This high level of the correlation instils confidence in the reliability of the model predictions. RMSE from Equation (2) for all stations except stations 3 and 4 are near zero, indicating a perfect fit (Nasidi *et al.* 2021; Sharma *et al.* 2023). These values imply that the predicted model precisely matches the actual rainfall measurement, showing accuracy and precision. The acceptance criterion for the NSE from the value of Equation (3) is a threshold greater than 0.5 (Bemmoussat *et al.* 2021; Meydani *et al.* 2022). The NSE shows how well the model captures the observed variations in rainfall. A value above 0.5 indicates a satisfactory level of model efficacy, implying model capability for reproducing rainfall values with a reasonable degree of accuracy. The values of the rainfall modelled are validated against the observed rainfall data for the study period 1979–2014, as presented in Figure 3.

5.3. Prediction of rainfall

An existing trend is determined from the modelled precipitation values for the period 1979–2014, and is further used to predict the rainfall till 2050 (Shaikh *et al.* 2022). The selected predictors are subsequently employed to execute the model for the future scenarios of SSP2-4.5 and SSP5-8.5, spanning the period from 2015 to 2100. The station points are digitized, and rainfall maps are prepared by interpolating the annual rainfall value within the study area. The rainfall maps for 2030, 2040, and 2050 are shown in Figure 4. The selection of the years 2030, 2040, and 2050 is predicated upon their proximity to the present, thereby rendering them suitable hopefuls for investigation within the realm of the near future. The examination of the near future has consistently held a position of significance within the realm of scholarly inquiry, as its findings serve to prompt proactive measures and ensure a continuous state of awareness. High rainfall (743.95 mm) is predicted for the year 2030 in SSP 4.5; by contrast, low rainfall (383–259.3 mm) is predicted in SSP 8.5. In 2040, SSP 4.5 predicts more rainfall as compared with SSP 8.5. Further, in 2050, SSP 8.5 predicts high rainfall (602–367 mm) as compared with SSP 4.5, which shows less rainfall (393–344 mm). However, the rainfall pattern is apparently similar for both scenarios, for all the years. Notably,

Table 4 | Predictors selected for modelling and prediction

Station	Predictors	Code	Partial-r	*p*-value
1	1,000 hPa Relative vorticity of true wind	p1_z	0.159	0.0000
	1,000 hPa Meridional wind component	p1_v	0.078	0.0073
	500 hPa Geopotential	p500	0.091	0.0015
	850 hPa Relative vorticity of true wind	p8_z	0.143	0.0000
	500 hPa Specific humidity	s500	0.117	0.0000
2	1,000 hPa Relative vorticity of true wind	p1_z	0.142	0.0000
	500 hPa Geopotential	p500	0.080	0.0080
	850 hPa Zonal wind component	p8_u	0.086	0.0003
	850 hPa Relative vorticity of true wind	p8_z	0.130	0.0000
	500 hPa Specific humidity	s500	0.065	0.0075
3	1,000 hPa Relative vorticity of true wind	p1_z	0.169	0.0000
	1,000 hPa Divergence of true wind	p1zh	0.121	0.0000
	500 hPa Geopotential	p500	0.080	0.0080
	850 hPa Relative vorticity of true wind	p8_z	0.101	0.0007
4	1,000 hPa Relative vorticity of true wind	p1_z	0.149	0.0000
	1,000 hPa Divergence of true wind	p1zh	0.095	0.0001
	500 hPa Wind speed	p5_f	0.084	0.0007
	500 hPa Relative vorticity of true wind	p5_z	0.092	0.0002
	500 hPa Geopotential	p500	0.113	0.0000
	500 hPa Specific humidity	s500	0.076	0.0023
5	1,000 hPa Meridional wind component	p1_v	0.072	0.0024
	1,000 hPa Relative vorticity of true wind	p1_z	0.179	0.0000
	500 hPa Relative vorticity of true wind	p5_z	0.092	0.0001
	500 hPa Geopotential	p500	0.114	0.0000
	850 hPa Zonal wind component	p8_u	0.083	0.0004
	850 hPa Relative vorticity of true wind	p8_z	0.126	0.0000
	Total precipitation	prcp	0.000	0.001
	500 hPa Specific humidity	s500	0.09	0.0001

Table 5 | Performance of model validation

Station	Correlation	RMSE	NSE
1	0.84	0.06	0.71
2	0.91	0.03	0.84
3	0.76	0.2	0.6
4	0.8	0.36	0.64
5	0.83	0.05	0.7

north-eastern areas receive more rainfall in 2030, whereas the middle and south-western areas receive more rainfall in 2040 and 2050, respectively.

5.4. Runoff computation

The computation of predicted runoff depth for the years 2030, 2040, and 2050 was conducted. The soil map has been downloaded from the Food and Agriculture Organization of the United Nations https://www.fao.org/soils-portal/ and

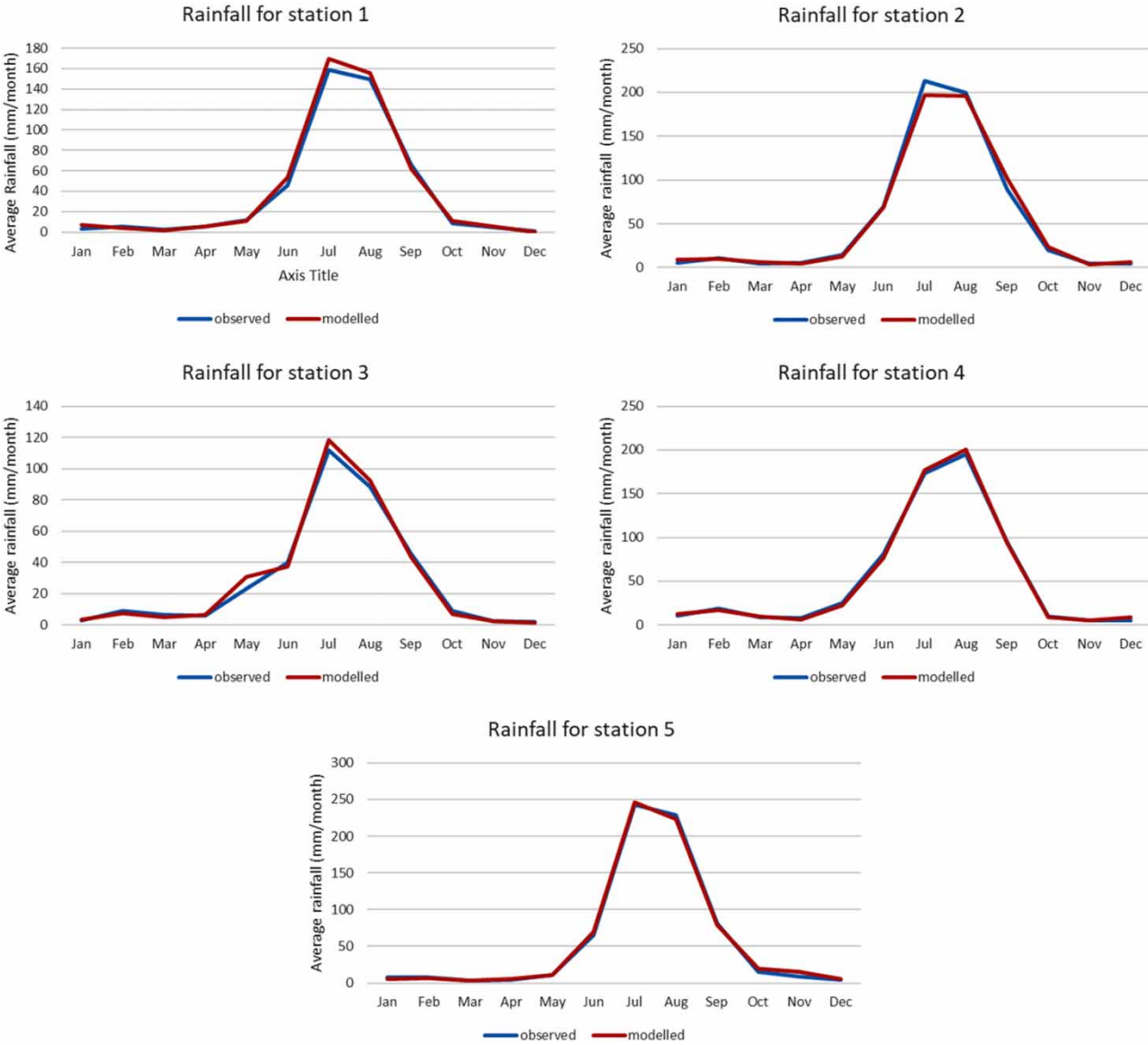


Figure 3 | Average monthly observed and modelled rainfall for the period 1979–2014.

delineated for the study area. The soil formation in Jaipur district consists of mainly loam and sand. Loamy soil and sandy loam were identified as hydrological soil groups D and C, respectively, as reported by Khare *et al.* (2017). To prepare the land-use map, the Landsat 8–9 image for 2021 was downloaded from the USGS website https://earthexplorer.usgs.gov. The Landsat image is mosaicked and delineated for the study area. Among several bands, land classes are identified and classified into five distinct classes, namely '*Urban settlement*', '*Agriculture/vegetation*', '*Open land*', '*Hills*', and '*Water*' (Kodihal & Akhtar 2023). The land-use map is superimposed on the HSG group soil map to identify CNs per the methodology suggested by Mishra & Singh (2003) and Patil *et al.* (2023). The CN for the LULC areas superimposing 'Agriculture' and 'Open land' on soil group HSG C is 86 and HSG D is 89. The CN for the areas superimposing 'urban settlement' on soil group HSG C and HSG D is 98, as mentioned in Table 6. The CN grid map is prepared using the raster calculator tool in the GIS platform presented in Figure 5. Further, runoff depth is computed using Equations (4) and (5) using the projected precipitation data for 2030, 2040, and 2050 (SSP 4.5 and 8.5). The maximum surface runoff depth and minimum surface runoff depth for the rainfall for the respective years are presented in Table 6. The maximum computed surface

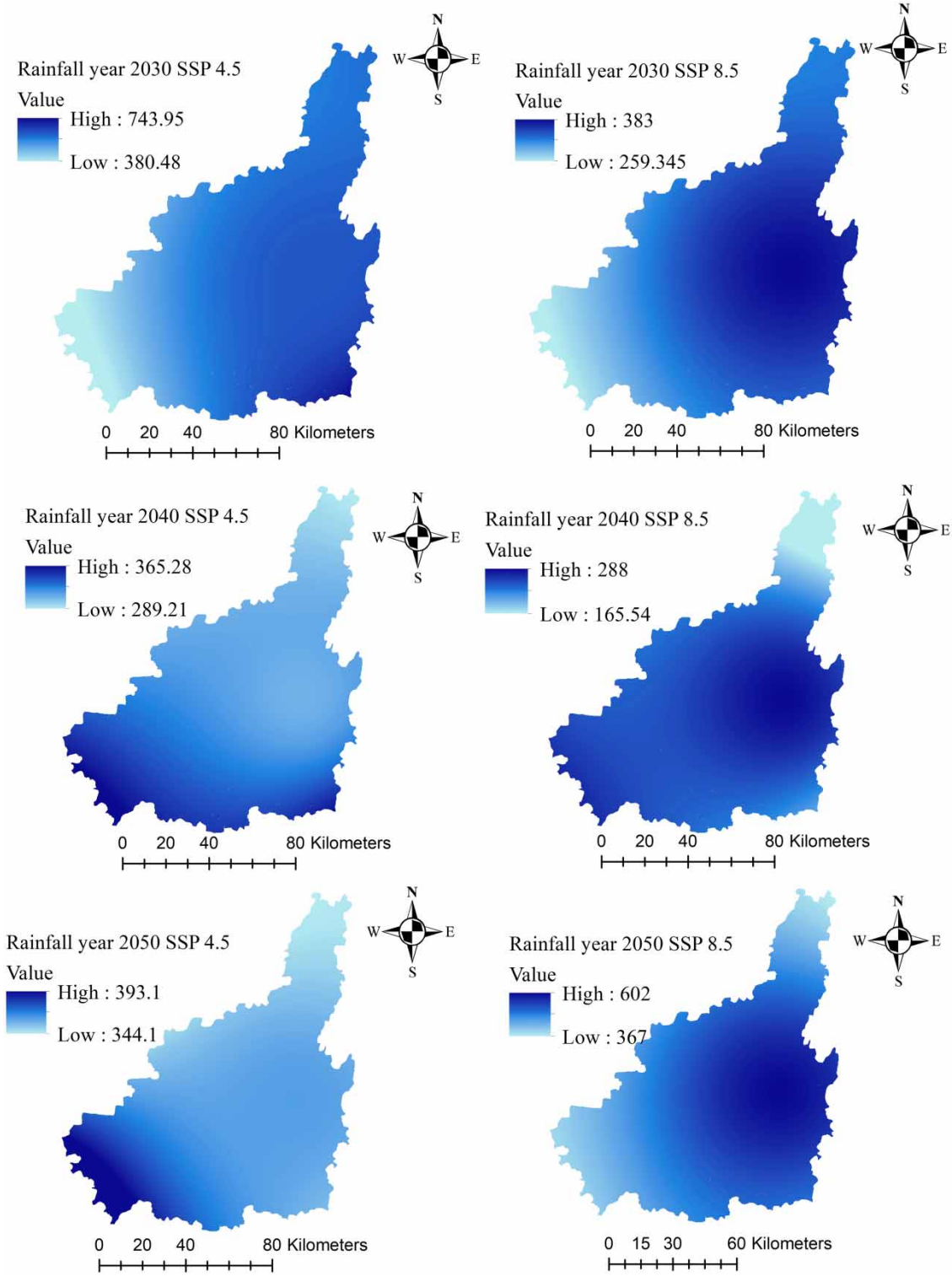


Figure 4 | Geographical representation of precipitation patterns for the scenarios SSP2-4.5 and SSP5-8.5 for 2030, 2040, and 2050.

runoff is (485.2 mm) for vegetation or open land in 2030 (SSP 4.5). The minimum computed surface runoff is at least (94.5 mm) for the vegetation or open-land class in 2040 (SSP 4.5). The 2040 SSP 8.5 shows the lowest values of computed runoff considering the worst-case scenario.

Table 6 | Rainfall values, curve numbers, and their respective surface runoff depths

Land-use class	CN		Maximum Annual Rainfall	Minimum Annual Rainfall	Maximum surface runoff	Minimum surface runoff	Maximum Annual Rainfall	Minimum Annual Rainfall	Maximum surface runoff	Minimum surface runoff
	HSG soil group C	HSG soil group D	in mm	in mm	depth in mm	depth in mm	in mm	in mm	depth in mm	depth in mm
Year 2030			SSP 4.5				SSP 8.5			
Urban settlement	98	98	658.4	618.52	431.770	395.420	383.4	340.32	190.604	156.357
Agriculture/ Vegetation/ Open land/Hills	86		686.88	431.74	485.280	259.940	383.4	292.31	173.784	106.299
Agriculture/ Vegetation/ Open land/Hills		89	603.35	215.48	414.260	94.590	383.4	292.31	178.236	109.785
Year 2040			SSP 4.5				SSP 8.5			
Urban settlement	98	98	328.3	315.76	146.958	137.283	288	260.14	116.389	96.254
Agriculture/ Vegetation/Open land/Hills	86		364.99	303.53	159.818	114.219	288	204.92	103.294	50.461
Agriculture/ Vegetation/Open land/Hills		89	364.99	315.76	164.094	126.764	288	260.14	106.730	87.522
Year 2050			SSP 4.5				SSP 8.5			
Urban settlement	98	98	358.12	351.59	170.483	165.273	288	260.14	116.389	96.254
Agriculture/ Vegetation/Open land/Hills	86		393.09	351.59	181.711	149.589	601.07	204.92	356.830	214.738
Agriculture/ Vegetation/Open land/Hills		89	393.09	351.59	186.258	153.730	601.07	434.26	362.932	219.652

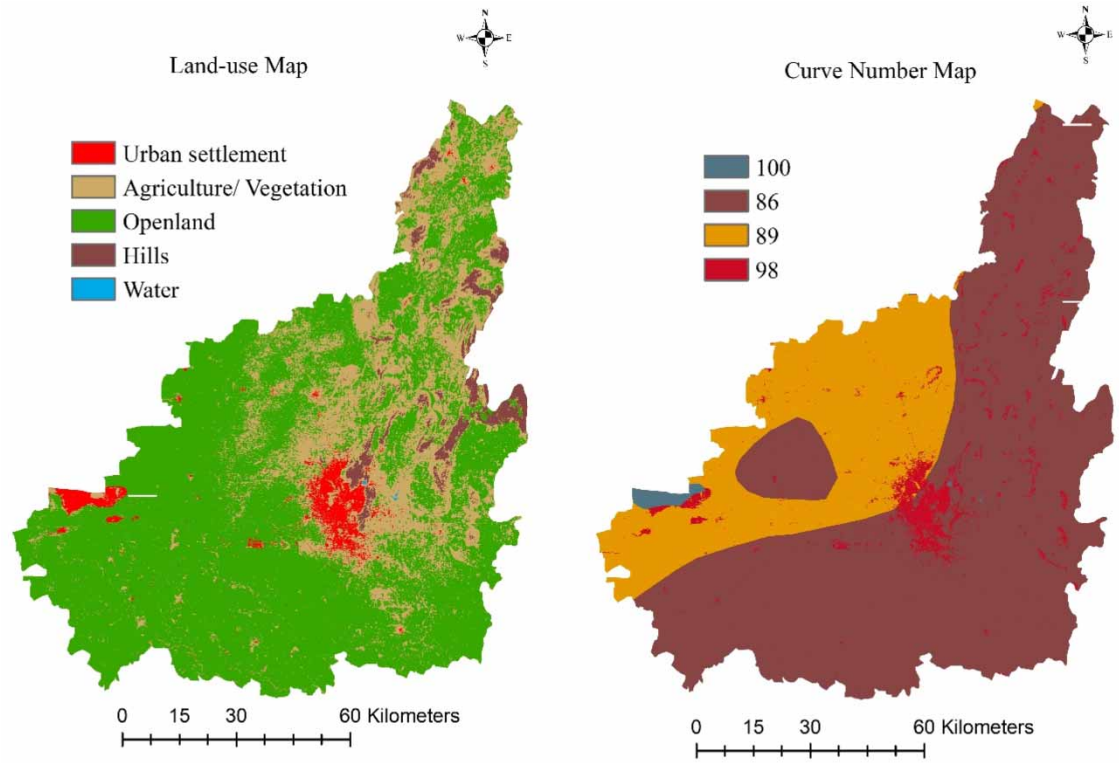


Figure 5 | Land-use map and curve number map of the study area.

6. CONCLUSION

The study is focused on assessing surface runoff depth for future scenarios. The study area has a semi-arid climate and receives comparatively low rainfall. The statistical downscaling approach is selected because it is reliable and effective in downscaling rainfall, as can be illustrated when comparing the predicted and actual rainfall using statistical matrices. The selection of predictors is crucial in downscaling as the accuracy of modelling depends on the correlation of predictors and predictand. The modelled rainfall values for the observed period (1979–2014) show less deviation from the observed rainfall values. The projected rainfall values show a decreasing rainfall trend from 2030 to 2040 and a slight increase from 2040 to 2050 for all scenarios. The statistical downscaling of future rainfall based on CMIP6 GCM (NorESM) predictors, coupled with SSP 4.5 and SSP 8.5, has proven to be a reliable approach. The meticulous validation process, employing statistical learning and assessing performance through metrics, has instilled confidence in the reliability of the generated rainfall projections. Integrating GIS has facilitated spatial modelling of the projected rainfall. Overlaying the future rainfall with land-use and soil data, a holistic understanding of the hydrological landscape is obtained. Runoff computation is challenging since it relies on land-use class, soil type, and rainfall values. Computed future runoff depth is less for sandy loam compared with loam soil. The decrease in runoff is observed from 2030 to 2040 for all scenarios. An important aspect of this study is its focus on the measurement of runoff independent of its basin area, with specific attention to runoff values presented in millimetres as depth. Overall, the scenarios show a future when the study region will be under water stress. These results will make hydrologists and water resource managers more prepared for the future.

7. LIMITATIONS AND FUTURE RECOMMENDATIONS

This model is sensitive to influential parameter selection; hence, the result may be significantly affected if predictor selection and data calibration have not been done appropriately. In this work, regression analysis has been applied as a machine learning method; however, with large data like hourly rainfall, observation may be fine-tuned with the other machine learning methods to improve the accuracy.

ACKNOWLEDGEMENT

The authors would like to acknowledge the United States Geological Survey (USGS) website, https://earthexplorer.usgs.gov, and the Food and Agriculture Organization of the United Nations https://www.fao.org/soils-portal/ for providing the required data.

DATA AVAILABILITY STATEMENT

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

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Anandhi, A., Srinivas, V. V., Nanjundiah, R. S. & Nagesh Kumar, D. 2008 Downscaling precipitation to river basin in India for IPCC SRES scenarios using support vector machine. *International Journal of Climatology* **28** (3), 401–420. https://doi.org/10.1002/JOC.1529.

Bemmoussat, A., Korichi, K., Baahmed, D., Maref, N., Djoukbala, O., Kalantari, Z. & Bateni, S. M. 2021 Contribution of satellite-based precipitation in hydrological rainfall–runoff modeling: case study of the hammam boughrara region in Algeria. *Earth Systems and Environment* **5** (4), 873–881. https://doi.org/10.1007/s41748-021-00256-z.

Chuenchum, P., Xu, M. & Tang, W. 2020 Predicted trends of soil erosion and sediment yield from future land use and climate change scenarios in the Lancang–Mekong River by using the modified RUSLE model. *International Soil and Water Conservation Research* **8** (3), 213–227. https://doi.org/10.1016/J.ISWCR.2020.06.006.

Doulabian, S., Golian, S., Toosi, A. S. & Murphy, C. 2021 Evaluating the effects of climate change on precipitation and temperature for Iran using RCP scenarios. *Journal of Water and Climate Change* **12** (1), 166–184. https://doi.org/10.2166/WCC.2020.114.

Huntingford, C., Lambert, F. H., Gash, J. H. C., Taylor, C. M. & Challinor, A. J. 2005 Aspects of climate change prediction relevant to crop productivity. *Philosophical Transactions of the Royal Society B: Biological Sciences* **360** (1463), 1999–2009. https://doi.org/10.1098/RSTB.2005.1748.

Hürlimann, M., Guo, Z., Puig-Polo, C. & Medina, V. 2022 Impacts of future climate and land cover changes on landslide susceptibility: regional scale modelling in the Val d'Aran region (Pyrenees, Spain). *Landslides* **19** (1), 99–118. https://doi.org/10.1007/S10346-021-01775-6/METRICS.

Hussain, M., Yusof, K. W., Mustafa, M. R. & Afshar, N. R. 2015 Application of statistical downscaling model (SDSM) for long term prediction of rainfall in Sarawak, Malaysia. *Water Resources Management VIII* **1**, 269–278. https://doi.org/10.2495/WRM150231.

Jaipur district. 2022 *Government of India Ministry of Jal Shakti Department of Water Resources, River Development & Ganga Rejuvenation, Central Ground Water Board Western Region Jaipur Report on National Aquifer Mapping & Management Plan*.

Khare, D., Singh, L. & Jat, M. K. (2017). Impact of Landuse Change on Urban Runoff: A Case Study of Jaipur City, India. *Proceedings of the 37th IAHR World Congress*. Available from: https://www.researchgate.net/publication/319178751_Impact_of_Landuse_Change_on_Urban_Runoff_A_Case_Study_of_Jaipur_City_India.

Kodihal, S. & Akhtar, M. P. 2023 A sustainable approach to study and prediction of urban footprints of jaipur and its growth direction using satellite based-data analysis. *Materials Today: Proceedings*. https://doi.org/10.1016/j.matpr.2023.05.693.

Kreienkamp, F., Lorenz, P. & Geiger, T. 2020 Statistically downscaled CMIP6 projections show stronger warming for Germany. *Atmosphere* **11**, 1245. https://doi.org/10.3390/ATMOS11111245.

Kumar, V., Kedam, N., Sharma, K. V., Mehta, D. J. & Caloiero, T. 2023 Advanced machine learning techniques to improve hydrological prediction: a comparative analysis of streamflow prediction models. *Water* **15** (14), 2572. https://doi.org/10.3390/W15142572.

Ma, X., Zhi, X., Azam, K., Ullah, I., Scafetta, N., Munawar, S., Rahman, G., Farhan, M., Moazzam, U., Miandad, M., Ullah, K., Al-Ansari, N., Thi, N. & Linh, T. 2022 Future climate projections using SDSM and LARS-WG downscaling methods for CMIP5 GCMs over the transboundary Jhelum River Basin of the Himalayas Region. *Atmosphere* **13** (6), 898. https://doi.org/10.3390/ATMOS13060898.

Mahdaoui, K., Tahiri, M. & Asmlal, L. 2023 Downscaling future climate changes under RCP emission scenarios using CanESM2 climate model over the Bouregreg catchment, Morocco. *Modeling Earth Systems and Environment* 1–11. https://doi.org/10.1007/S40808-022-01683-1/METRICS.

Meenu, R., Rehana, S. & Mujumdar, P. P. 2013 Assessment of hydrologic impacts of climate change in Tunga–Bhadra river basin, India with HEC-HMS and SDSM. *Hydrological Processes* **27** (11), 1572–1589. https://doi.org/10.1002/HYP.9220.

Mehta, D., Dhabuwala, J., Yadav, S. M., Kumar, V. & Azamathulla, H. M. 2023 Improving flood forecasting in Narmada river basin using hierarchical clustering and hydrological modelling. *Results in Engineering* **20**, 101571. https://doi.org/10.1016/J.RINENG.2023.101571.

Meydani, A., Dehghanipour, A., Schoups, G. & Tajrishy, M. 2022 Daily reservoir inflow forecasting using weather forecast downscaling and rainfall-runoff modeling: Application to Urmia Lake basin, Iran. *Journal of Hydrology: Regional Studies* **44**, 101228. https://doi.org/10.1016/J.EJRH.2022.101228.

Mishra, S. K. & Singh, V. P. 2003 *SCS-CN Method*. 84–146. https://doi.org/10.1007/978-94-017-0147-1_2.

Nasidi, N. M., Wayayok, A., Abdullah, A. F. & Mohd Kassim, M. S. 2021 Dynamics of potential precipitation under climate change scenarios at Cameron highlands, Malaysia. *SN Applied Sciences* **3** (3), 1–17. https://doi.org/10.1007/S42452-021-04332-X/FIGURES/8.

Ncoyini-Manciya, Z. & Savage, M. J. 2022 The assessment of future Air temperature and rainfall changes based on the statistical downscaling model (SDSM): the case of the wartburg community in KZN Midlands, South Africa. *Sustainability* **14** (17), 10682. https://doi.org/10.3390/SU141710682.

Patil, D., Kumar, G., Kumar, A. & Gupta, R. 2023 A systematic basin-wide approach for locating and assessing volumetric potential of rainwater harvesting sites in the urban area. *Environmental Science and Pollution Research* **30** (6), 14707–14721. https://doi.org/10.1007/S11356-022-23039-Z/TABLES/2.

Phuong, D. N. D., Duong, T. Q., Liem, N. D., Tram, V. N. Q., Cuong, D. K. & Loi, N. K. 2020 Projections of future climate change in the Vu Gia Thu bon river basin, Vietnam by using statistical downScaling model (SDSM). *Water* **12** (3), 755. https://doi.org/10.3390/W12030755.

Seland, Ø., Bentsen, M., Olivié, D., Toniazzo, T., Gjermundsen, A., Graff, L. S., Debernard, J. B., Gupta, A. K., He, Y. C., Kirkeväg, A., Schwinger, J., Tjiputra, J., Schanke Aas, K., Bethke, I., Fan, Y., Griesfeller, J., Grini, A., Guo, C., Ilicak, M. & Schulz, M. 2020 Overview of the Norwegian earth system model (NorESM2) and key climate response of CMIP6 DECK, historical, and scenario simulations. *Geoscientific Model Development* **13** (12), 6165–6200. https://doi.org/10.5194/GMD-13-6165-2020.

Shaikh, M. M., Lodha, P., Lalwani, P. & Mehta, D. 2022 Climatic projections of Western India using global and regional climate models. *Water Practice and Technology* **17** (9), 1818–1825. https://doi.org/10.2166/WPT.2022.090.

Sharma, K. V., Kumar, V., Singh, K., Mehta, D. J., Sharma, K. V., Kumar, V., Singh, K. & Mehta, D. J. 2023 LANDSAT 8 LST pan sharpening using novel principal component based downscaling model. *RSASE* **30**, 100963. https://doi.org/10.1016/J.RSASE.2023.100963.

Verma, S., Kumar, K., Verma, M. K., Prasad, A. D., Mehta, D. & Rathnayake, U. 2023 Comparative analysis of CMIP5 and CMIP6 in conjunction with the hydrological processes of reservoir catchment, Chhattisgarh, India. *Journal of Hydrology: Regional Studies* **50**, 101533. https://doi.org/10.1016/J.EJRH.2023.101533.

Wilby, R. L. & Dawson, C. W. 2007 *SDSM 4.2-A Decision Support Tool for the Assessment of Regional Climate Change Impacts User Manual*. Available from: http://www.cics.uvic.ca/scenarios/index.cgi?Scenarios.

Wilby, R. L. & Dawson, C. W. 2013 The statistical downscaling model: insights from one decade of application. *International Journal of Climatology* **33** (7), 1707–1719. https://doi.org/10.1002/JOC.3544.

First received 30 September 2023; accepted in revised form 22 December 2023. Available online 19 January 2024