Statistical downscaling using principal component regression for climate change impact assessment at the Cauvery river basin

Parthiban Loganathan and Amit Baburao Mahindrakar

ABSTRACT

Climate change impact studies are generally carried out with higher resolution general circulation model (GCM) outputs, which are usually for a global scale, and it is difficult to use the same for a regional scale. GCM simulations require downscaling to get a coarser scale output for local climate impact studies. In this study, an improvised principal component regression (PCR) downscaling technique is adapted to downscale 26 Coupled Model Intercomparison Project Phase 5 (CMIP5) GCM historical outputs. A massive river basin named Cauvery with 35 observation stations is categorized into three subbasins to study the regional climate impacts. In this case, the PCR model performed remarkably well compared to other conventional machine learning models with half the computational time than usual. The test statistics state that the validation of the proposed model illustrates a variance in calibration results of the PCR model, which ranges between 2 and 5%, and a variance in validation, which is less than 7% throughout the study area. Since it is desired to prioritize GCMs to choose the merely suitable models for a strategic climate study, the models were selected based on the PCR model performance. Furthermore, CCSM4, inmcm4, and EC-EARTH model's performance in recreating precipitation statistics over the study area are exceptional.

Key words | climate change, performance evaluation, statistical downscaling

Parthiban Loganathan
Amit Baburao Mahindrakar (corresponding author)
Department of Environmental and Water Resources Engineering, School of Civil, Engineering,
Vellore Institute of Technology (VIT),
Vellore,
Tamil Nadu 632014,
India
E-mail: *amahindrakarlab@gmail.com*

HIGHLIGHTS

- The evaluation and comparison of downscaling Coupled Model Intercomparison Project Phase 5 (CMIP5) general circulation models (GCMs) based on renowned machine learning (ML) techniques.
- The suggestion of an alternate ML (principal component regression) approach for improvised downscaling.
- Subbasin-wise climate assessment on a large-scale river basin.
- The intercomparison of ML models with respect to calibration and validation periods.
- The outcome suggests an improvised climate downscaling approach with an appropriate CMIP5 GCM.

doi: 10.2166/wcc.2021.223

INTRODUCTION

Climate change is denoted as a change in the long-term statistics of atmospheric variables over a specific region. Urbanization and increased industrial activities led to changes in the concentration of greenhouse gases in the atmosphere. The Fifth Assessment Report (AR5) of the Intergovernmental Panel on Climate Change (IPCC) (IPCC 2014) stated that the role of greenhouse-gas emissions and the severity of potential risks and impacts remain unmanageable, especially for the developing countries with their limited ability. The climate change induced by CO_2 has affected the globe in enormous ways and is likely to become more furious shortly. Therefore, it is essential to assess the impacts of climate change over the globe to illustrate the evidence to decision-makers and stakeholders.

Climate change impact assessment studies require long-term climate data at various spatio-temporal scales. Finer spatial resolution-observed data for the historical time scale are required for such studies and they are available for certain regions over the past few decades. However, future climate projections are required to predict and analyze future climate changes. General circulation models (GCMs) are mathematical models projecting future climate scenarios. But the spatial resolution of these model outputs is coarser and can not be used directly for regional climate impact studies (Oo *et al.* 2019, 2020; Burciaga 2020; Javadinejad *et al.* 2020). Additionally, the GCM model outputs contain significant errors and require correction to ensure an appropriate outcome. Consequently, the downscaling of GCM model output is necessary to make use of it for the regional climate impact assessment (Ahmed *et al.* 2013; Wilby *et al.* 2014; Yang *et al.* 2020).

There are two different methods to downscale the GCMs, which are dynamic downscaling and statistical downscaling. Dynamic downscaling requires high-end computational power and consumes a lot of time (Busuioc *et al.* 2001; Brown *et al.* 2017; Li *et al.* 2020). Thus, it is not feasible to perform dynamic downscaling in the required region, whereas statistical downscaling requires less computational power and can be performed in a short span (Aribarg *et al.* 2016; Asong *et al.* 2016; Tiwari *et al.* 2019). Numerous studies stated that the performance of dynamic and statistical downscaling was comparable in regional-scale climate studies over historical time-scales (Wilby *et al.* 1998; Feddersen & Andersen 2005). A detailed study comparing the dynamic and statistical downscaling stated that the dynamic downscaling did not perform any better than the statistical downscaling (Chávez-Arroyo *et al.* 2013, 2015).

Statistical downscaling is developed based on the assumption that the statistical relationship between the historical observed and historical GCM output will remain constant in future climate projections (Wilby & Dawson 2013). There are numerous atmospheric parameters to consider for climate change impact studies. But the primary weather variables, such as precipitation (Pr), mean temperature (Tas), maximum temperature (Tasmax), and minimum temperature (Tasmin), play a major role in representing the hydrological system in a given region. It is also evident from previous studies that the different GCMs' output performance varies significantly in representing the regional climate scenario (Hessami *et al.* 2008; Meaurio *et al.* 2017; Shamir *et al.* 2019). Various downscaling models have been developed to handle the bias and variance in projecting future climate for a regional scale (Sarr *et al.* 2015; Yue *et al.* 2016; Preethi *et al.* 2019).

In this study, an improvised principal component regression (PCR)-based downscaling approach is carried out using a dataset of daily precipitation and temperature (mean, minimum, and maximum) for the climate change impact assessment over the Cauvery river basin. Also, 26 different Coupled Model Intercomparison Project Phase 5 (CMIP5) GCM outputs were used to project the future scenarios and rank their performance for the study area. The designed model performance is evaluated using calibration and validation. Besides, the outcomes are compared with the existing methods to state the pick over other techniques in representing the regional climate scenario.

The primary objectives of the present study are:

(i) To perform statistical downscaling using improvised PCR and compare it with the existing methods for its leads.
(ii) To model climate scenarios using different CMIP5 GCMs and prioritize models based on performance evaluation.

A brief picture of the study area is delivered in the next section. Particulars of the datasets used in this study are presented in the 'Data description' section. The model design and performance evaluation are provided in the 'Methodology' section, and the penultimate section explains the results and discussions. The last section explains the 'Summary and conclusion' of the study.

STUDY AREA

The Cauvery river basin is a part of peninsular India which lies between 75°27′–79°54′E and 10°9′–13°30′N. It is estimated to be 85,000 km^2 with many streams, including Kabini, Bhavani, and Noyyal. Cauvery covers three states (Tamil Nadu, Karnataka, and Kerala) and a Union Territory (Puducherry). It is guarded by the Western Ghats on the west and by the Eastern Ghats on the east and south. A map demonstrating the location and the border of the study area is shown in Figure 1. The main regions of the Cauvery basin are covered with agricultural land up to 67% of the total area, and 20% of the basin is covered by the forest area (CWC & NRSC 2014). The Cauvery river basin has four seasons, namely winter (December–February), summer (March–June), South-West monsoon (July–September), and North-East monsoon (October–November). The basin remains dry in the majority of the area except for the monsoon duration. The basin has both tropical and subtropical climate zones where the temperature variation in the upper reaches of the basin is less compared to the lower reaches. April is the hottest and January turns out to be the coldest month in the basin, where the average monthly temperature ranges from 18 to 33 °C. Northern parts of the basin are comparatively colder than the southern parts of the basin.

DATA DESCRIPTION

GCM data

The CMIP5 GCM was used in this study considering the daily ensemble realization run r1i1p1. Twenty-six models from various institutions were selected, namely ACCESS1-0, ACCESS1-3, bcc-csm1-1-m, BNU-ESM, CanCM4, CanESM2, CCSM4, CMCC-CESM, CNRM-CM5, CSIRO-Mk3-6-0, EC-EARTH, FGOALS-g2, GFDL-CM3, HadGEM2-AO, HadGEM2-C, HadGEM2-ES, inmcm4, IPSL-CM5A-MR, MIROC5, MIROC-ESM, MIROC-ESM-CHEM, MPI-ESM-LR, MPI-ESM-MR, MRI-CGCM3, MRI-ESM1, and NorESM1-M. The description of the selected models is presented in Table 1. The weather parameters considered in the present

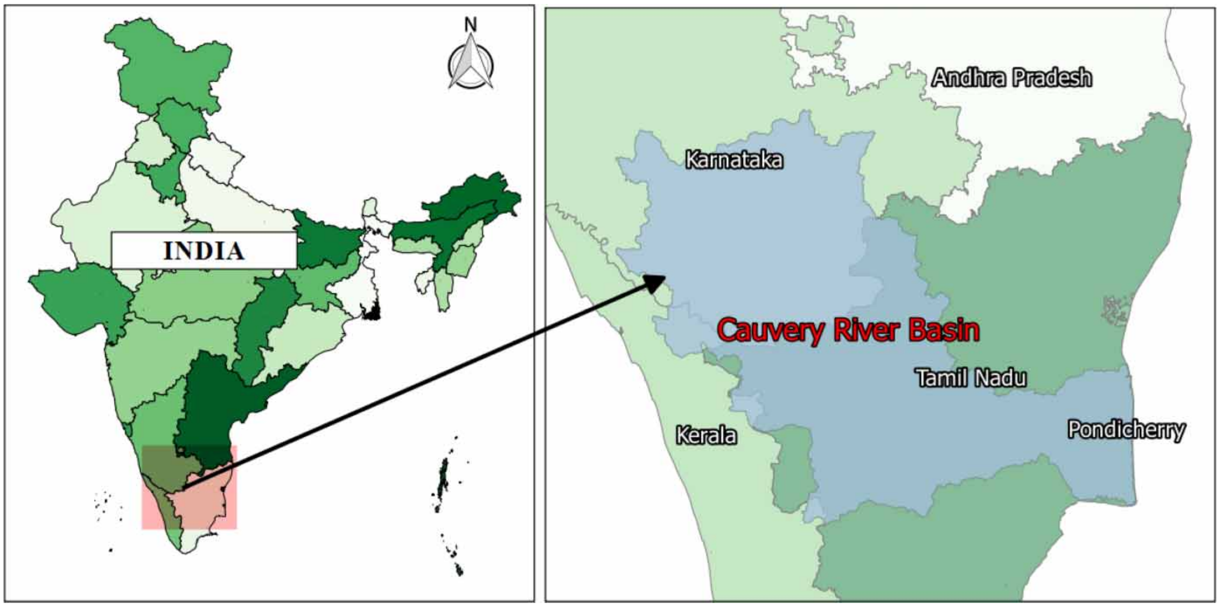

Figure 1 | Cauvery river basin extent and boundary.

Table 1 | Description of CMIP5 models used in this study

				Atmosphere eq. resolution	
S. No.	CMIP5 Model ID	Institute and country of origin	Atmosphere horizontal resolution (°latitude × °longitude)	Latitude (km)	Longitude (km)
1	ACCESS-1.0	CSIRO-BOM, Australia	1.9 × 1.2	210	130
2	ACCESS-1.3	CSIRO-BOM, Australia	1.9 × 1.2	210	130
3	BCC-CSM1-1-M	BCC, CMA, China	1.1 × 1.1	120	120
4	BNU-ESM	BNU, China	2.8 × 2.8	310	310
5	CanCM4	CCCMA, Canada	2.8 × 2.8	310	310
6	CanESM2	CCCMA, Canada	2.8 × 2.8	310	310
7	CCSM4	NCAR, USA	1.2 × 0.9	130	100
8	CMCC-CESM	CMCC, Italy	3.7 × 3.7	410	410
9	CNRM-CM5	CNRM-CERFACS, France	1.4 × 1.4	155	155
10	CSIRO-Mk3-6-0	CSIRO-QCCCE, Australia	1.9 × 1.9	210	210
11	EC-EARTH	EC-EARTH, Europe	1.1 × 1.1	120	120
12	FGOALS-g2	IAP/LASG, China	2.8 × 2.8	310	310
13	GFDL-CM3	NOAA, GFDL, USA	2.5 × 2.0	275	220
14	HadGEM2-AO	NIMR-KMA, Korea	1.9 × 1.2	210	130
15	HadGEM2-CC	MOHC, UK	1.9 × 1.2	210	130
16	HadGEM2-ES	MOHC, UK	1.9 × 1.2	210	130
17	INMCM4	INM, Russia	2.0 × 1.5	220	165
18	IPSL-CM5A-MR	IPSL, France	2.5 × 1.3	275	145
19	MIROC5	JAMSTEC, Japan	1.4 × 1.4	155	155
20	MIROC-ESM	JAMSTEC, Japan	2.8 × 2.8	310	310
21	MIROC-ESM-CHEM	JAMSTEC, Japan	2.8 × 2.8	310	310
22	MPI-ESM-LR	MPI-N, Germany	1.9 × 1.9	210	210
23	MPI-ESM-MR	MPI-N, Germany	1.9 × 1.9	210	210
24	MRI-CGCM3	MRI, Japan	1.1 × 1.1	120	120
25	MRI-ESM1	MRI, Japan	1.1 × 1.1	120	120
26	NorESM1-M	NCC, Norway	2.5 × 1.9	275	210

study are precipitation (Pr – mm/day), mean temperature (Tas – °C/day), maximum temperature (Tasmax – °C/day), and minimum temperature (Tasmin – °C/day).

Observed data

The daily observed station data are obtained from the India Meteorological Department (IMD) for the 35 stations located in the Cauvery river basin from 1951 to 2005. The description of the stations is presented in Supplementary Table S1. The river basin is divided into the upper, middle, and lower Cauvery river basin based on its weather pattern and discharge statistics. Furthermore, the historical GCM datasets are re-gridded to the station scale and trimmed to the observed data specifications. The observation stations and their location are represented as a thematic map in Figure 2.

METHODOLOGY

The observed station-wise climate data (1951–2005) and historical simulation from various CMIP5 GCMs are collected and re-gridded using the Bicubic Spline Interpolation technique to ensure that all datasets are in the same dimension and co-ordinates. Furthermore, the re-gridded data are separated into calibration (1951–1990) and validation (1991–2005) for the

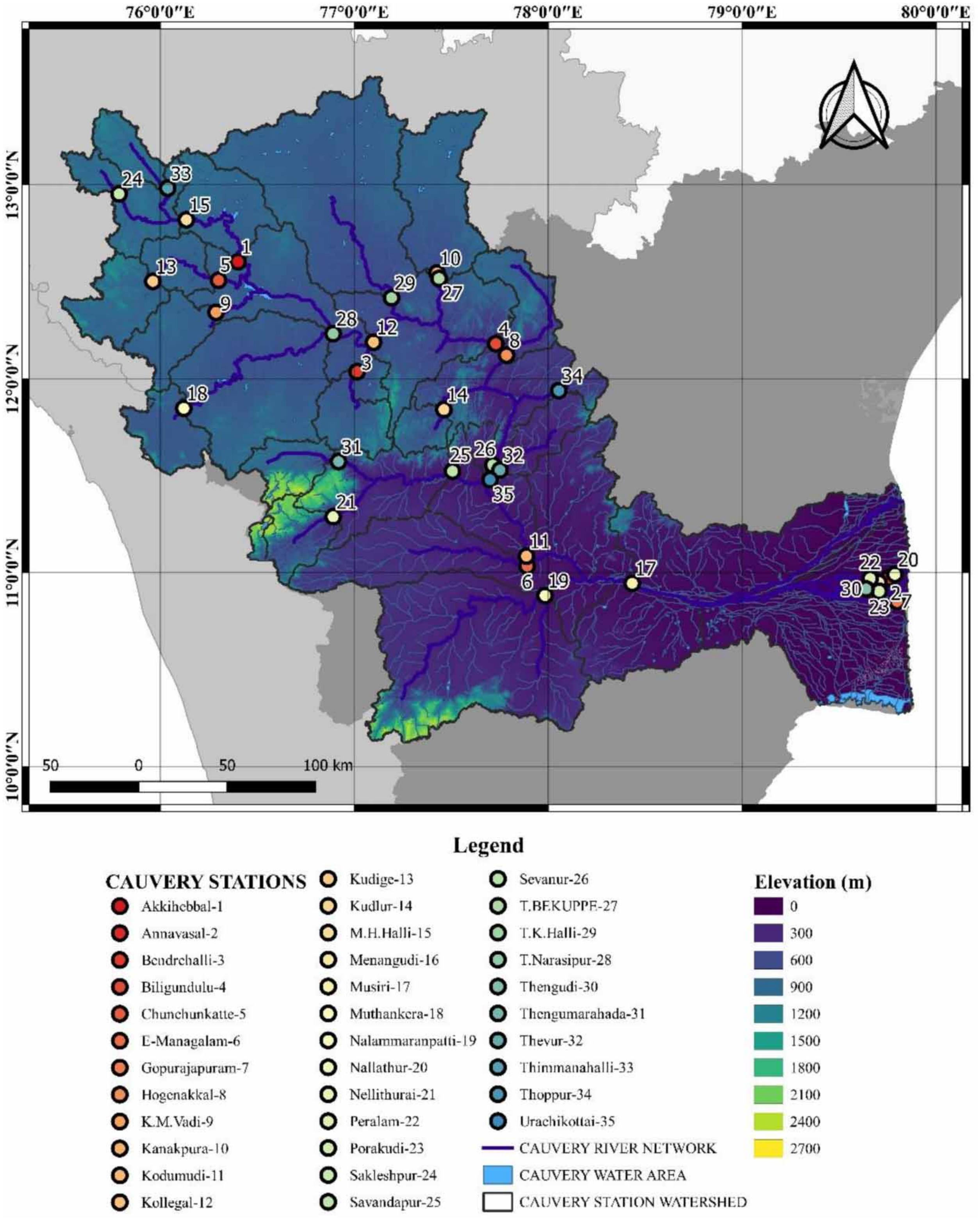

Figure 2 | Cauvery river basin with observation stations.

betterment of the model. The designed dataset is trained and tested with existing well-established statistical machine learning (ML) models such as multiple linear regression (Joshi *et al.* 2013), partial least-squares regression (PLSR) (Matulessy *et al.* 2015), artificial neural network (ANN) (Khan & Coulibaly 2010), support vector machine (SVM) (Tabari *et al.* 2012), and K-nearest neighbor (KNN) (Caraway *et al.* 2014). Later, the ML models considered are compared with the proposed improvised PCR (Sahriman *et al.* 2014) to state the advantages over other techniques. The comparison of considered models was performed with the help of performance evaluation parameters. The ranking of CMIP5 GCM models based on regions and parameters is identified using the suggested PCR model. A brief methodology for downscaling the CMIP5 GCM is represented in Figure 3.

Principal component regression

The station-wise historical simulation from GCMs and station-observed time-series datasets are used to extract the principal components. Furthermore, future scenarios are projected by the model using future GCM scenarios. The PCR model can be denoted subsequently by the typical notation if the equation is represented in a matrix form as follows:

$$Y = XB + e \tag{1}$$

where Y is the dependent variable, X is the independent variable, B is the regression coefficient, and e represents errors/residuals. The regression coefficient in ordinary least squares is signified as the following:

$$B = (X'X)^{-1}X'Y \tag{2}$$

Here, since the variables are standardized $X'X = R$, where R is the correlation matrix of independent variables. The PCR is performed by converting independent variables to their principal components:

$$X'X = PDP' = Z'Z \tag{3}$$

where D is the diagonal matrix of eigenvalues of $X'X$, P is the eigenvector matrix of $X'X$, and Z is the principal component matrix of data matrix X. The estimation formula

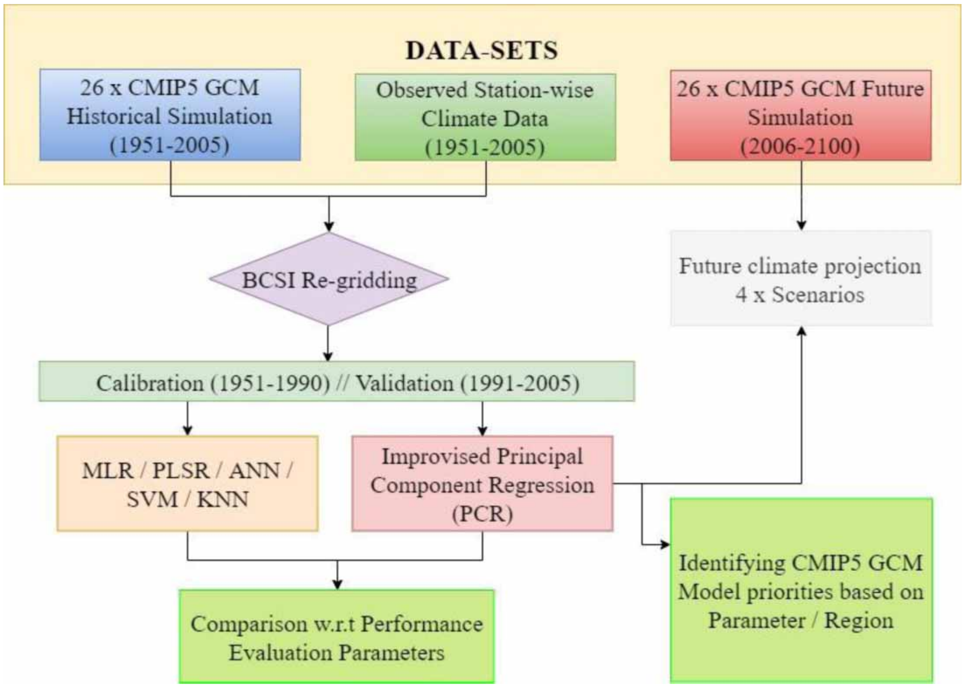
DATA-SETS
26 x CMIP5 GCM Historical Simulation (1951-2005)
Observed Station-wise Climate Data (1951-2005)
26 x CMIP5 GCM Future Simulation (2006-2100)
BCSI Re-gridding
Future climate projection 4 x Scenarios
Calibration (1951-1990) // Validation (1991-2005)
MLR / PLSR / ANN / SVM / KNN
Improvised Principal Component Regression (PCR)
Identifying CMIP5 GCM Model priorities based on Parameter / Region
Comparison w.r.t Performance Evaluation Parameters

Figure 3 | Flow chart for downscaling CMIP5 GCM.

can be mathematically represented as follows:

$$A = (Z'Z)^{-1}Z'Y = D^{-1}Z'Y \tag{4}$$

The PCR applies ordinary least-squares regression to a different set of independent variables determined by the principal components.

Performance evaluation parameters

The performance of each GCM based on the ability to capture the monotonic trend is evaluated using performance indices, such as normalized root-mean-square error (NRMSE) and the coefficient of determination (R^2) (Woldemeskel *et al.* 2012; Roozbeh 2018). The equations for computing selected performance evaluation parameters are as follows.

Normalized root-mean-square error

The NRMSE is derived by normalizing the square root of the mean squared errors by its absolute range to a scale of [0,1]. The closer the value is to 0, the lesser the variance, and thus the better the performance of the model.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (\mathrm{obs_i} - \mathrm{sim}_i)^2}{n}} \tag{5}$$

$$\mathrm{NRMSE} = \frac{\mathrm{RMSE}}{\max(\mathrm{Obs}) - \min(\mathrm{Obs})} \tag{6}$$

Coefficient of determination

The R^2 of a model is a ratio of the regression sum of squares (SSR) to the total sum of squares (SSTO). Where the regression sum of squares is the arithmetic difference between the error sum of squares (SSE) and the total sum of squares.

$$R^2 = 1 - \frac{\mathrm{SSE}}{\mathrm{SSTO}} = \frac{\mathrm{SSR}}{\mathrm{SSTO}} = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2} \tag{7}$$

RESULTS AND DISCUSSION

Various research suggests several ranking techniques that were applied around the globe. However, very few studies were performed over the Indian subcontinent (Das *et al.* 2018). The present study considers the Cauvery river basin located in southern-peninsular India which is classified into the upper, middle, and lower basins for a better understanding of subbasin-level changes in climatic conditions over the past few decades. Subbasin-wise values of precipitation and temperature are analyzed to assess how various models represent the actual scenarios of the basin. A map illustrating the subbasins of the Cauvery river basin is presented in Figure 4. The subbasin-wise climate data are calculated using the station-wise-observed mean weather data for each subbasin. The reliability of each model is validated through performance evaluation parameters (NRMSE % and R^2) of model results. The following section presents the results obtained from this study and the discussion related to it.

Comparison of model performance

The performance of selected models in replicating the actual scenario based on the region and parameters is evaluated using nominated performance evaluation parameters (Ruan *et al.* 2018). The calibration and validation scores of each model are evaluated, and the performance evaluation scores of the calibration and validation dataset are represented in Table 2. The validation results show that all the CMIP5 GCMs were better at replicating temperature than precipitation in each subbasin considered (MacAdam *et al.* 2010; Das *et al.* 2018). Also, the PCR model performed remarkably well compared to other ML models (Noori *et al.* 2010; Chávez-Arroyo *et al.* 2013). The proposed technique enables the ranking of GCM concerning a local scale such as a specific river or subbasin. The variance between the observed and simulated climate data for all the regions and parameters considered in the study was below 7% in PCR, and the other models' results ranged between 12 and 30%. Moreover, the time engaged for the PCR model to train and test the data is almost half the time taken for other individual models selected in this study.

The overall length of historical subbasin-wise climate data for 55 years (1951–2005) is split into 75% calibration and 25% validation. The evaluation is carried out with a calibration period of 40 years (1951–1990) and a validation period of 15 years (1991–2005) for a better understanding

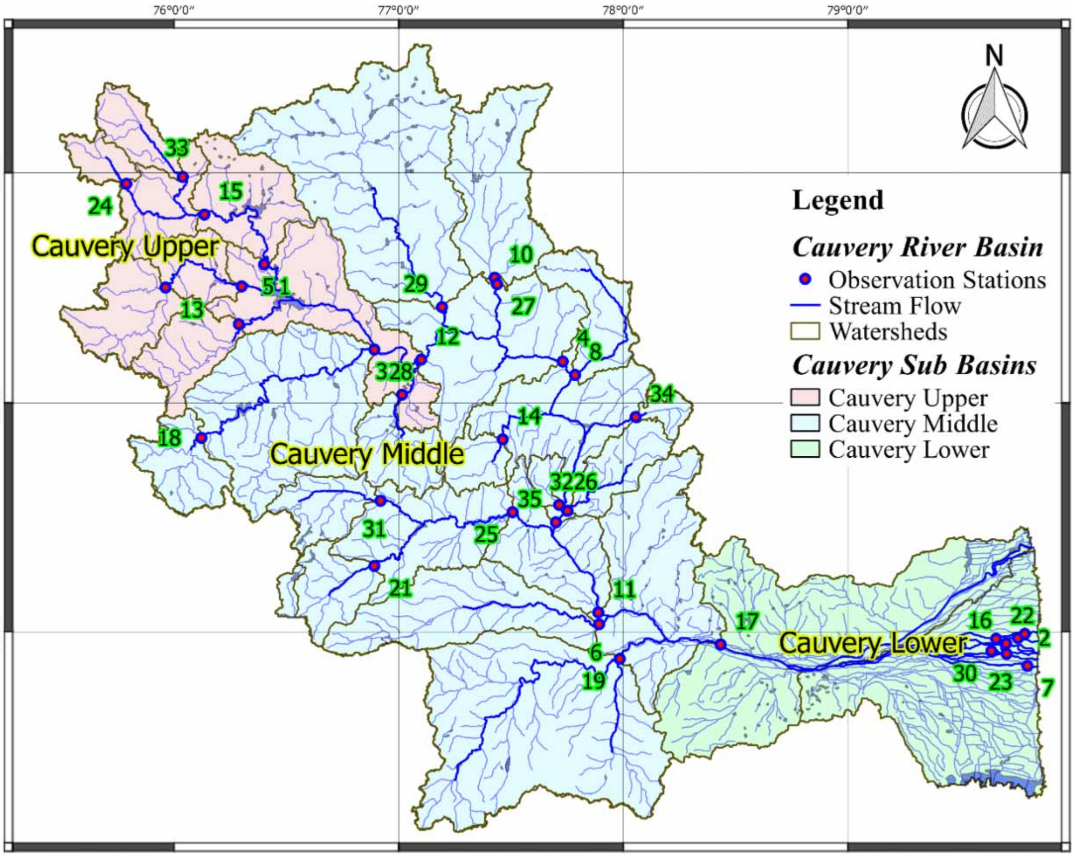

Figure 4 | Classification of the Cauvery river basin.

of the performance of each ML model. The comparative bar plots clearly explain that the performance of every model during the validation phase is slightly lower than the calibration phase. It is also evident that the RMSE % values of the PCR model are far lower than other ML models for each subbasin and parameter. Furthermore, the R^2 values of the PCR model are above 0.9 for calibration and 0.7 for validation at every combination. This illustrates that the variance of the PCR model is very less, and the model performance is higher compared to other renowned ML models.

Prioritizing of CMIP5 GCMs based on PCR with respect to regions/parameters

The priority of CMIP5 models was assessed by evaluating the performance of the PCR model using individual GCM. The priorities of these models were evaluated by providing equal scores to the performance evaluation parameters. The test statistics obtained from the analysis are presented in Table 3. The results state that CCSM4, inmcm4, EC-Earth, MIROC-ESM-CHEM, and BNU-ESM are good at projecting the historical precipitation events in all three subbasins. Whereas GFDL-CM3, CNRM-CM5, IPSL-CM5A-MR, and inmcm4 are performing exceptionally well in projecting the temperature changes (average, maximum, and minimum) over the Cauvery river basin.

SUMMARY AND CONCLUSION

Downscaling of climate data over a certain region requires the selection of suitable GCM and appropriate models to replicate the regional scenario. The model adopted for downscaling is expected to be effective and computationally economical. It is necessary to identify the suitable downscaling model for the selected region and the applicable GCMs for the selected parameters to represent the actual scenario.

Table 2 | Performance evaluation scores of calibration and validation datasets

Model		**R^2**						**NRMSE (%)**					
Parameter	**Basin**	**GLM**	**PLSR**	**ANN**	**SVM**	**KNN**	**PCR**	**GLM**	**PLSR**	**ANN**	**SVM**	**KNN**	**PCR**
Calibration													
PR	Upper	0.54	0.45	0.51	0.50	0.56	**0.94**	20.40	31.68	28.88	22.55	20.14	**4.39**
	Middle	0.60	0.53	0.57	0.58	0.62	**0.94**	17.85	23.32	22.61	19.06	17.65	**6.77**
	Lower	0.52	0.45	0.49	0.49	0.54	**0.93**	18.01	24.07	22.56	24.38	18.83	**6.37**
TAS	Upper	0.76	0.58	0.76	0.75	0.87	**0.97**	11.69	17.19	11.69	11.64	10.51	**4.12**
	Middle	0.76	0.59	0.76	0.75	0.87	**0.98**	11.67	16.56	11.67	11.97	10.36	**4.13**
	Lower	0.63	0.47	0.63	0.61	0.75	**0.96**	17.07	23.06	17.07	16.36	17.06	**6.75**
TASMAX	Upper	0.76	0.66	0.76	0.75	0.78	**0.97**	12.29	17.50	12.29	12.24	16.29	**5.71**
	Middle	0.48	0.36	0.44	0.45	0.54	**0.94**	22.35	32.31	30.77	19.50	18.95	**6.25**
	Lower	0.72	0.66	0.68	0.70	0.74	**0.96**	13.85	18.57	17.27	13.70	13.60	**5.17**
TASMIN	Upper	0.77	0.63	0.77	0.76	0.88	**0.98**	12.65	15.49	12.65	12.75	10.50	**4.14**
	Middle	0.79	0.64	0.79	0.79	0.89	**0.98**	12.14	14.93	12.14	12.61	9.84	**4.08**
	Lower	0.81	0.67	0.81	0.80	0.89	**0.98**	11.88	15.00	11.88	12.42	9.82	**3.88**
Validation													
PR	Upper	0.46	0.38	0.43	0.43	0.48	**0.80**	17.34	26.93	24.55	19.17	17.12	**3.73**
	Middle	0.48	0.42	0.46	0.46	0.50	**0.75**	14.28	18.66	18.09	15.25	14.12	**5.42**
	Lower	0.39	0.34	0.37	0.37	0.41	**0.70**	13.51	18.05	16.92	18.29	14.12	**4.78**
TAS	Upper	0.65	0.49	0.65	0.64	0.74	**0.82**	9.94	14.61	9.94	9.89	8.93	**3.50**
	Middle	0.61	0.47	0.61	0.60	0.70	**0.78**	9.34	13.25	9.34	9.58	8.29	**3.30**
	Lower	0.47	0.35	0.47	0.46	0.56	**0.72**	12.80	17.30	12.80	12.27	12.80	**5.06**
TASMAX	Upper	0.65	0.56	0.65	0.64	0.66	**0.82**	10.45	14.88	10.45	10.40	13.85	**4.85**
	Middle	0.38	0.29	0.35	0.36	0.43	**0.75**	17.88	25.85	24.62	15.60	15.16	**5.00**
	Lower	0.54	0.50	0.51	0.53	0.56	**0.72**	10.39	13.93	12.95	10.28	10.20	**3.88**
TASMIN	Upper	0.65	0.54	0.65	0.65	0.75	**0.83**	10.75	13.17	10.75	10.84	8.93	**3.52**
	Middle	0.63	0.51	0.63	0.63	0.71	**0.78**	9.71	11.94	9.71	10.09	7.87	**3.26**
	Lower	0.61	0.50	0.61	0.60	0.67	**0.74**	8.91	11.25	8.91	9.32	7.37	**2.91**

The bold values signify better performance of the PCR model compared to other ML models.

Table 3 | CMIP5 GCM priority based on the PCR model performance for selected weather parameters

Parameter: precipitation/average, minimum, and maximum temperature

	Upper			**Middle**			**Lower**		
Rank	**Model**	**NRMSE %**	**R^2**	**Model**	**NRMSE %**	**R^2**	**Model**	**NRMSE %**	**R^2**
1	MRI-ESM1	4.85	0.83	CCSM4	5.42	0.78	CCSM4	5.06	0.74
2	CCSM4	3.9	0.82	inmcm4	5.00	0.78	CNRM-CM5	4.78	0.72
3	ACCESS1-3	3.73	0.82	MRI-ESM1	4.25	0.76	inmcm4	4.20	0.72
4	inmcm4	3.52	0.81	EC-EARTH	3.30	0.75	MIROC-ESM	3.88	0.72
5	EC-EARTH	3.50	0.80	BNU-ESM	3.26	0.75	EC-EARTH	2.91	0.70

The performed downscaling technique is well suitable for the Cauvery river basin, and the applicability of the same at other basins depends on the performance metrics since each river basin has unique properties. The following conclusions have been drawn from this study: (i) the PCR downscaling model performs exceptionally well compared to other conventionally used ML models. (ii) The time taken to develop a PCR downscaling model for a given sub-basin is reduced by almost half the time taken to build a conventional downscaling model. (iii) The variance in calibration results of the PCR model is ranged between 2 and 5%, whereas the validation results were <7% throughout all subbasins and for each parameter under consideration. (iv) The prioritizing of GCMs based on PCR proposed that for precipitation, CCSM4, inmcm4, and EC-Earth are best in representing the historical scenario. (v) The temperature statistics are captured well in GFDL-CM3, CNRM-CM5, and inmcm4. Also, inmcm4 stays among the top rank for precipitation and temperature for all subbasins.

DATA AVAILABILITY STATEMENT

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

REFERENCES

Ahmed, K. F., Wang, G., Silander, J., Wilson, A. M., Allen, J. M., Horton, R. & Anyah, R. 2013 Statistical downscaling and bias correction of climate model outputs for climate change impact assessment in the U.S. northeast. *Global and Planetary Change* **100**, 320–332. https://doi.org/10.1016/j.gloplacha.2012.11.003.

Aribarg, T., Raghavan, S. V., Vu, M. T., Liong, S.-Y. Y., Supratid, S., Aribarg, T., Supratid, S., Raghavan, S. V. & Liong, S.-Y. Y. 2016 Statistical downscaling rainfall using artificial neural network: significantly wetter Bangkok? *Theoretical and Applied Climatology* **126**, 453–467. https://doi.org/10.1007/s00704-015-1580-1.

Asong, Z. E., Khaliq, M. N. & Wheater, H. S. 2016 Projected changes in precipitation and temperature over the Canadian Prairie Provinces using the Generalized Linear Model statistical downscaling approach. *Journal of Hydrology* **539**, 429–446. https://doi.org/10.1016/j.jhydrol.2016.05.044.

Brown, K., Kamruzzaman, M. & Beecham, S. 2017 Trends in sub-daily precipitation in Tasmania using regional dynamically downscaled climate projections. *Journal of Hydrology: Regional Studies* **10**, 18–34. https://doi.org/10.1016/j.ejrh.2016.12.086.

Burciaga, U. M. 2020 Sustainability assessment in housing building organizations for the design of strategies against climate change. *HighTech and Innovation Journal* **1**, 136–147. https://doi.org/10.28991/hij-2020-01-04-01.

Busuioc, A., Chen, D. & Hellström, C. 2001 Performance of statistical downscaling models in GCM validation and regional climate change estimates: application for Swedish precipitation. *International Journal of Climatology* **21**, 557–578. https://doi.org/10.1002/joc.624.

Caraway, N. M., McCreight, J. L. & Rajagopalan, B. 2014 Multisite stochastic weather generation using cluster analysis and k-nearest neighbor time series resampling. *Journal of Hydrology* **508**, 197–213. https://doi.org/10.1016/j.jhydrol.2013.10.054.

Chávez-Arroyo, R., Lozano-Galiana, S., Sanz-Rodrigo, J. & Probst, O. 2013 On the application of principal component analysis for accurate statistical-dynamical downscaling of wind fields. *Energy Procedia* **40**, 67–76. https://doi.org/10.1016/j.egypro.2013.08.009.

Chávez-Arroyo, R., Lozano-Galiana, S., Sanz-Rodrigo, J. & Probst, O. 2015 Statistical-dynamical downscaling of wind fields using self-organizing maps. *Applied Thermal Engineering* **75**, 1201–1209. https://doi.org/10.1016/j.applthermaleng.2014.03.002.

CWC and NRSC. 2014 *Cauvery Basin Report*. Central Water Commission and National Remote Sensing Centre.

Das, L., Dutta, M., Mezghani, A. & Benestad, R. E. 2018 Use of observed temperature statistics in ranking CMIP5 model performance over the Western Himalayan Region of India. *International Journal of Climatology* **38**, 554–570. https://doi.org/10.1002/joc.5193.

Feddersen, H. & Andersen, U. 2005 A method for statistical downscaling of seasonal ensemble predictions. *Tellus A: Dynamic Meteorology and Oceanography* **57**, 398–408. https://doi.org/10.1111/j.1600-0870.2005.00102.x.

Hessami, M., Gachon, P., Ouarda, T. B. M. J. & St-Hilaire, A. 2008 Automated regression-based statistical downscaling tool. *Environmental Modelling and Software* **23**, 813–834. https://doi.org/10.1016/j.envsoft.2007.10.004.

IPCC. 2014 *Climate Change 2014*. Synthesis Report. Versión inglés.

Javadinejad, S., Dara, R. & Jafary, F. 2020 Climate change scenarios and effects on snow-melt runoff. *Civil Engineering Journal* **6**, 1715–1725. https://doi.org/10.28991/cej-2020-03091577.

Joshi, D., St-Hilaire, A., Daigle, A. & Ouarda, T. B. M. J. 2013 Data based comparison of Sparse Bayesian Learning and Multiple Linear Regression for statistical downscaling of low flow indices. *Journal of Hydrology* **488**, 136–149. https://doi.org/10.1016/j.jhydrol.2013.02.040.

Khan, M. S. & Coulibaly, P. 2010 Assessing hydrologic impact of climate change with uncertainty estimates: Bayesian neural network approach. *Journal of Hydrometeorology* **11**, 482–495. https://doi.org/10.1175/2009JHM1160.1.

Li, Y., Liang, Z., Hu, Y., Li, B., Xu, B. & Wang, D. 2020 A multi-model integration method for monthly streamflow prediction: modified stacking ensemble strategy. *Journal of Hydroinformatics* **22**, 310–326. https://doi.org/10.2166/hydro.2019.066.

MacAdam, I., Pitman, A. J., Whetton, P. H. & Abramowitz, G. 2010 Ranking climate models by performance using actual values and anomalies: implications for climate change impact assessments. *Geophysical Research Letters* **37**, 1–6. https://doi.org/10.1029/2010GL043877.

Matulessy, E. R., Wigena, A. H. & Djuraidah, A. 2015 Quantile regression with partial least squares in statistical downscaling for estimation of extreme rainfall. *Applied Mathematical Sciences* **9**, 4489–4498. https://doi.org/10.12988/ams.2015.53254.

Meaurio, M., Zabaleta, A., Boithias, L., Epelde, A. M., Sauvage, S., Sánchez-Pérez, J. M., Srinivasan, R. & Antiguedad, I. 2017 Assessing the hydrological response from an ensemble of CMIP5 climate projections in the transition zone of the Atlantic region (Bay of Biscay). *Journal of Hydrology* **548**, 46–62. https://doi.org/10.1016/j.jhydrol.2017.02.029.

Noori, R., Khakpour, A., Omidvar, B. & Farokhnia, A. 2010 Comparison of ANN and principal component analysis-multivariate linear regression models for predicting the river flow based on developed discrepancy ratio statistic. *Expert Systems with Applications* **37**, 5856–5862. https://doi.org/10.1016/j.eswa.2010.02.020.

Oo, H. T., Zin, W. W. & Thin Kyi, C. C. 2019 Assessment of future climate change projections using multiple global climate models. *Civil Engineering Journal* **5**, 2152–2166. https://doi.org/10.28991/cej-2019-03091401.

Oo, H. T., Zin, W. W. & Thin Kyi, C. C. 2020 Analysis of streamflow response to changing climate conditions using SWAT model. *Civil Engineering Journal* **6**, 194–209. https://doi.org/10.28991/cej-2020-03091464.

Preethi, B., Ramya, R., Patwardhan, S. K., Mujumdar, M. & Kripalani, R. H. 2019 Variability of Indian summer monsoon droughts in CMIP5 climate models. *Climate Dynamics* **53**, 1937–1962. https://doi.org/10.1007/s00382-019-04752-x.

Roozbeh, M. 2018 Optimal QR-based estimation in partially linear regression models with correlated errors using GCV criterion. *Computational Statistics and Data Analysis* **117**, 45–61. https://doi.org/10.1016/j.csda.2017.08.002.

Ruan, Y., Yao, Z., Wang, R. & Liu, Z. 2018 Ranking of CMIP5 GCM skills in simulating observed precipitation over the Lower Mekong Basin, using an improved score-based method. *Water* **10**. https://doi.org/10.3390/w10121868.

Sahriman, S., Djuraidah, A. & Wigena, A. H. 2014 Application of principal component regression with dummy variable in statistical downscaling to forecast rainfall. *Journal of Statistics* **2014** (4), 678–686.

Sarr, M. A., Seidou, O., Tramblay, Y. & El Adlouni, S. 2015 Comparison of downscaling methods for mean and extreme precipitation in Senegal. *Journal of Hydrology: Regional Studies* 4 (B), 369–385. https://doi.org/10.1016/j.ejrh.2015.06.005.

Shamir, E., Halper, E., Modrick, T., Georgakakos, K. P., Chang, H. I., Lahmers, T. M. & Castro, C. 2019 Statistical and dynamical downscaling impact on projected hydrologic assessment in arid environment: a case study from Bill Williams River basin and Alamo Lake, Arizona. *Journal of Hydrology X* **2**, 100019. https://doi.org/10.1016/j.hydroa.2019.100019.

Tabari, H., Kisi, O., Ezani, A. & Hosseinzadeh Talaee, P. 2012 SVM, ANFIS, regression and climate based models for reference evapotranspiration modeling using limited climatic data in a semi-arid highland environment. *Journal of Hydrology* **444–445**, 78–89. https://doi.org/10.1016/j.jhydrol.2012.04.007.

Tiwari, P. R., Kar, S. C., Mohanty, U. C., Dey, S., Sinha, P., Shekhar, M. S. & Sokhi, R. S. 2019 Comparison of statistical and dynamical downscaling methods for seasonal-scale winter precipitation predictions over north India. *International Journal of Climatology* **39**, 1504–1516. https://doi.org/10.1002/joc.5897.

Wilby, R. L., Wigley, T. M. L., Conway, D., Jones, P. D., Hewitson, B. C., Main, J. & Wilks, D. S. 1998 Statistical downscaling of general circulation model output: A comparison of methods. *Water Resources Research* 34 (11), 2995–3008. https://doi.org/10.1029/98WR02577.

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

Wilby, R. L., Dawson, C. W., Murphy, C., O'Connor, P. & Hawkins, E. 2014 The statistical downScaling model-decision centric (SDSM-DC): conceptual basis and applications. *Climate Research* **61**, 251–268. https://doi.org/10.3354/cr01254.

Woldemeskel, F. M., Sharma, A., Sivakumar, B. & Mehrotra, R. 2012 An error estimation method for precipitation and temperature projections for future climates. *Journal of Geophysical Research Atmospheres* **117**, 1–13. https://doi.org/10.1029/2012JD018062.

Yang, M., Li, X., Shi, W., Zhang, C. & Zhang, J. 2020 The Pacific-Indian Ocean associated mode in CMIP5 models. *Ocean Science* **16**, 469–482. https://doi.org/10.5194/os-16-469-2020.

Yue, T. X., Zhao, N., Fan, Z. M., Li, J., Chen, C. F., Lu, Y. M., Wang, C. L., Xu, B. & Wilson, J. 2016 CMIP5 downscaling and its uncertainty in China. *Global and Planetary Change* **146**, 30–37. https://doi.org/10.1016/j.gloplacha.2016.09.003.

First received 6 August 2020; accepted in revised form 23 February 2021. Available online 12 March 2021