

Development of statistical downscaling methods for the assessment of rainfall characteristics under climate change scenarios

Thanipa Onarun, Chaiyapong Theprasit and Ketvara Sittichok *

Department of Irrigation Engineering, Faculty of Engineering at Kamphaengsaen Campus, Kasetsart University, Nakhon Pathom 73140, Thailand

*Corresponding author. E-mail: fengkrs@ku.ac.th

 KS, 0000-0003-4986-2907

ABSTRACT

The objective of this research was to develop a statistical downscaling approach in the Phetchaburi River Basin, Thailand, consisting of two main processes: predictor selection and relationship construction between predictors and local rainfall. Super predictor (SP) and stepwise regression (SR) were employed with principal component analysis (PCA) and the statistical downscaling model (SDSM) was applied later. Four statistical models were finally used: M1 (SP-PCA), M2 (SP-SDSM), M3 (SR-PCA), and M4 (SR-SDSM) with 26 large circulation indices generated by CanESM2 (CMIP5). Finally, the characteristics of extreme rainfall events were observed under three climate change scenarios during three different periods (2020–2040, 2041–2070, and 2071–2100). The results revealed that rainfall and geostrophic airflow velocity were the best predictors for rainfall downscaling, followed by divergence, meridional velocity, and relative humidity. Three objective functions (R^2 , NSE, and RMSE) were applied to evaluate model performance. The M4 model presented the highest performance while M1 showed the lowest skill. The average annual rainfall specifically increased compared with the historical rainfall for RCP2.6, RCP4.5, and RCP8.5 scenarios in future periods. Very to extremely wet years determined by the standardized anomaly index (SAI) occurred more often in the far-future while severe to extremely dry years frequently occurred in the mid- and far-future.

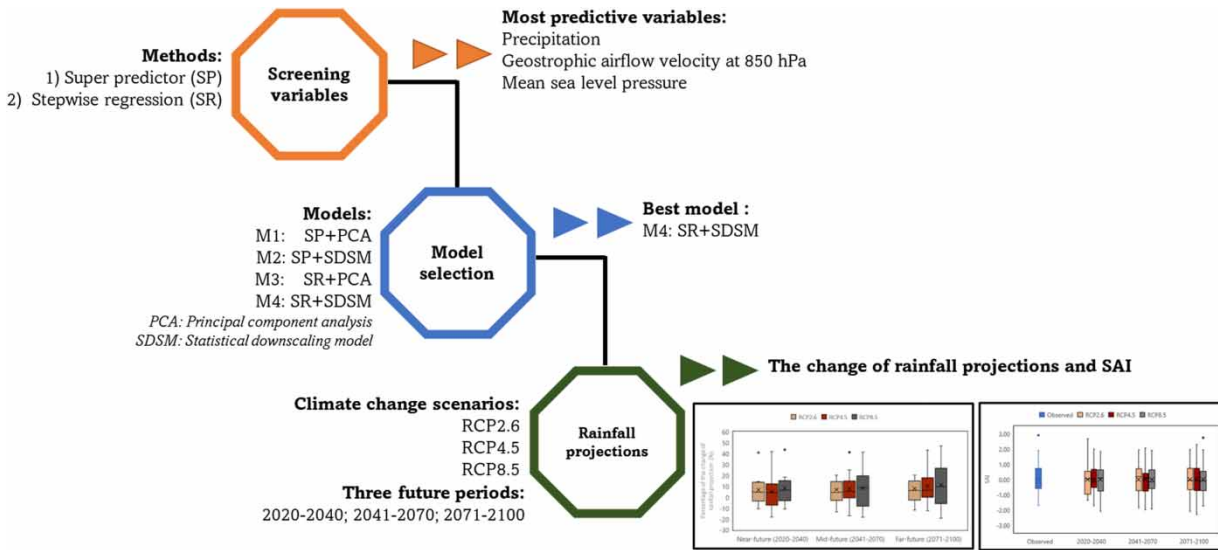
Key words: climate change, principal component analysis, SDSM, statistical downscaling

HIGHLIGHTS

- The use of stepwise regression and SDSM presented the highest performance for rainfall downscaling.
- The two most predictive variables generated by GCMs for rainfall downscaling in this area were precipitation and geostrophic airflow velocity at 850 hPa.
- Future rainfall projection trends and their characteristics using the SAI method were presented under three climate change scenarios.

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/>).

GRAPHICAL ABSTRACT



1. INTRODUCTION

Climate change is known to create an imbalance in the global climate system. Significant efforts have been made in the study of climate change to ascertain its impact on various areas, especially hydrological and water resources management (O'Connell 2017; Padhiary *et al.* 2020; Zhang *et al.* 2021). Therefore, general circulation models or global climate models (GCMs) are widely used in the climate change community due to their satisfactory performance in simulating and projecting climate and other significant variables. GCMs are numerical models, representing the mechanisms of atmosphere, land surface, oceans, and glaciers. Various emission scenarios and their impacts corresponding to future changes in different socio-economic conditions have been thoroughly investigated by the Intergovernmental Panel on Climate Change (IPCC) using GCMs as the main tools.

To study the impact of climate change, including identifying the most effective approaches for adaptation and mitigation, it is necessary to focus on specific local areas. Unreliable results may occur if the outputs of GCMs are directly used for local study areas since they show low resolution at the spatial scale (Hernanz *et al.* 2021; Loganathan & Mahindrakar 2021). The GCM resolution is currently between 100 and 300 km and systematic biases of the outputs of GCMs regarding observations are frequently found (Tabari *et al.* 2021). Therefore, downscaling techniques are essential for the study of climate change's impact on the local or regional scale.

There are two main downscaling approaches: dynamical and statistical techniques. Dynamical downscaling relies on complex physical and mathematical calculation using a high-resolution regional climate model (RCM) within the GCM. The resolutions of RCM outputs can occur at finer scales of less than 50 km, but with the high computational cost involved in calculating the RCM (Hernanz *et al.* 2021; Tabari *et al.* 2021). Statistical downscaling methods allow the investigation of the outputs of GCMs at a single-point scale with lower cost of computation compared with RCMs. However, the limitation of the statistical downscaling methods is the assumption of the relationships between climate indices and observations remaining unchanged in a future climate (Hernanz *et al.* 2021). As predictors and predictands, respectively, large-scale climate indices and local variables (e.g., temperature and precipitation) are forced into statistical models for historical periods. This initial process is performed to identify suitable climate indices for downscaling. The predictors generated from GCMs are then applied to produce local variable projections.

Various statistical downscaling approaches have been developed and investigated for estimating the impact of climate change. Three classifications of statistical modeling are stochastic weather generation, weather typing, and transfer function (Gebrechorkos *et al.* 2019), and the two main parts of statistical downscaling are predictor selection and correlation generation. Two statistical methods for correlation generation (statistical downscaling model: SDSM/principal component analysis: PCA) and two screening processes for predictor selection (super predictor: SP and stepwise regression: SR) were

investigated in this study. Multiple linear regression together with stochastic weather classification are used in the SDSM, whereas PCA has the advantage of reducing the number of predictors and their multicollinearity. The SDSM was applied in the work of [Suo *et al.* \(2019\)](#) to downscale precipitation and temperature in China and it presented satisfactory performance in simulating these two variables. A correlation matrix and *p*-value were considered and applied in the screening process followed by SDSM by [Tahir *et al.* \(2018\)](#) for rainfall downscaling. Rainfall projections were then produced under climate change CMIP5 scenarios showing rainfall increment in the Limbang River Basin of Sarawak. Other research works using SDSM can be found in [Shahriar *et al.* \(2021\)](#) and [Barokar *et al.* \(2019\)](#). For PCA application, [Loganathan & Mahindrakar \(2021\)](#) employed PCA for precipitation downscaling compared with the conventional method using a multiple linear regression model. Results revealed that PCA performance showed better skill with less computation time. The use of SR for screening process could be seen in the research work of [Osman & Abdellatif \(2016\)](#). Eight out of 26 predictors were selected to be used in different statistical downscaling models for rainfall downscaling in the northwest of England. They remarkably recommended combining various statistical techniques for downscaling. Other research works have separately applied these methods for downscaling local variable data, as can be seen in [Benestad *et al.* \(2015\)](#); [Shukla *et al.* \(2015\)](#); [Teegavarapu & Goly \(2018\)](#) and [Rahman *et al.* \(2022\)](#).

Even though various GCMs exist, the scope of this study refers to the outputs of CanESM2 as predictors to investigate all four statistical models for downscaling rainfall in the study area. This model has the advantage of being compatible and user-friendly with SDSM. Outputs of CanESM2 are promptly input to this application without requirement of any initial steps for preparing the data. Furthermore, various works on climate change agree with employing the results of CanESM2 to study the impact of climate change on different hydrological concerns. For example, [Armain *et al.* \(2021\)](#) employed the output of CanESM2 for precipitation projection while [Hussain *et al.* \(2017\)](#) also used it for both precipitation and temperature projections in Malaysia. [Javaherian *et al.* \(2021\)](#) investigated temperature and precipitation under different climate change scenarios with CanESM2 in Iran. Works using the outputs of CanESM2 for the study of climate change at the regional and local scale include [Yang & Saenko \(2012\)](#), [Hua *et al.* \(2014\)](#), and [Hassan & Hashim \(2020\)](#).

2. STUDY AREA

The Phetchaburi River Basin covers three provinces in Thailand, namely Phetchaburi, Samut Songkhram, and Ratchaburi, measuring approximately 6,255 km². The south side is adjacent to the Prachuap Khiri Khan Coastal Basin; the west border is with Myanmar; and the east side is located next to the Gulf of Thailand, as shown in [Figure 1](#). Average annual precipitation of the basin is around 1,110 mm in the rainy season from May to October. Annual minimum/maximum temperature and average relative humidity are 27.9/33.5 °C and 74.9%, respectively. During dry spells, this basin is frequently affected by drought, especially in regions isolated from water sources. Every year from January to May, the temperature is hot, causing significant issues with some regions receiving only a small amount of precipitation. In addition, the amount of water in the reservoir is dwindling, rendering it incapable of satisfying the water-needs of various places. There is also the problem of flash floods due to the short period of heavy rainfall. According to the flood statistics, Phetchaburi Province, which covers a substantial portion of the river basin, experienced severe flooding conditions in 2003, 2016, 2017, and 2018 causing damage to more than 300 km².

3. RESEARCH METHODOLOGY

3.1. Data

3.1.1. Observed rainfall

The rain gauge stations ([Table 1](#)) of the Thai Meteorological Department (TMD) collected daily precipitation statistics from 1985 to 2020. Due to the mountainous topography on the west side of the Phetchaburi River Basin, the rain gauge stations were unable to cover the distribution of the entire river basin. Of the ten representative rain gauge stations in this study, nine were located in the Phetchaburi River Basin and one in the surrounding area, with annual precipitation averaging between 778 and 1,018 mm. Missing values were filled using the inverse distance weighting (IDW) method. The rainfall data for each station were then divided into calibration and validation periods for model verification.

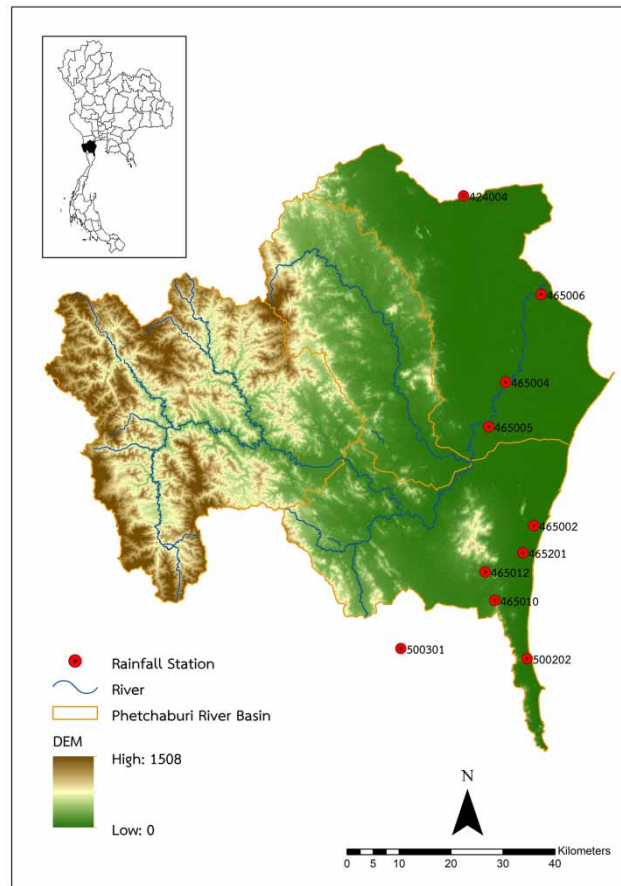


Figure 1 | Study area and measurement stations of the Phetchaburi River Basin, Thailand.

Table 1 | Rain gauge station data used in this study during the calibration periods and model verification

Station code	Name of station	Average rainfall (mm/year)	Geo-reference of location	
			Latitude	Longitude
424004	Pak Tho District Office	1,079	13° 22' 30"	99° 50' 38"
465002	Cha-am District Agriculture Office	899	12° 47' 56"	99° 57' 58"
465004	Ban Lat District Agriculture Office	904	13° 2' 56"	99° 55' 1"
465005	Thayang District Agriculture Office	788	12° 58' 19"	99° 53' 13"
465006	Ban Laem District Office	1,022	13° 12' 11"	99° 58' 52"
465010	Royal Forest Development Project	858	12° 40' 5"	99° 53' 49"
465012	Sri Nagarindra Park	917	12° 43' 5"	99° 52' 48"
465201	Phetchaburi Weather Station	1,020	12° 45' 7"	99° 56' 49"
500202	Hua Hin Weather Station	974	12° 34' 1"	99° 57' 11"
500301	Nong Phlub Agromet Meteorological Station	1,156	12° 35' 6"	99° 43' 48"

3.1.2. CanESM2 outputs

The CanESM2 General Circulation Model (GCM) was employed in this study due to its extensive use in various areas and beyond and its compatibility with the SDSM. The grid cell model indicated dimensions of $2.8125^\circ \times 2.8125^\circ$. The following data were used in this study: (1) NCEP global climate datasets for 1961–2005 recorded by the National Centers for

Environmental Prediction of the United States and (2) CanESM2 historical datasets from 1961 to 2005 and future climate-change data under three scenarios, namely RCP2.6, RCP4.5, and RCP8.5 from 2006 to 2100. The RCP metrics represent radiative forcing values (W/m^2). RCP2.6, RCP4.5, and RCP8.5 refer to a stringent mitigation scenario, an intermediate scenario, and a very high greenhouse gas scenario, respectively. Thus, the CanESM2 output consisted of 26 variables, as shown in Table 2.

3.2. Methods

3.2.1. Selection of significant GCM variables

Three main approaches were used in this study: selection of significant GCM variables, constructing relationships between selected variables and local daily rainfall, and projection of future rainfall under climate change scenarios. Variable selection is a vital process in statistical downscaling. It involves identifying relationships between climate variables and daily rainfall from each rain gauge station to gather an appropriate set of variables (Huang *et al.* 2011). In this study, two screening methods were employed: SP and SR, as explained in the following.

- SP method

The SP method was established by Mahmood & Babel (2012) and uses the results from the SDSM function to consider the correlation coefficient, partial correlation coefficient, and p -value. Several studies have used this method for variable selection (Singh *et al.* 2015; Pathan & Waikar 2020; Ahsan *et al.* 2021). By selecting variables using this method, multicollinearity between more than two independent variables can be reduced (Khadka & Pathak 2016). Correlation coefficients (R) were calculated between the 26 NCEP variables and local rainfall. From the 26 variables selected, the first 13 highest R values were initially selected and the highest of these was used as a SP. The selected predictors with p -value greater than 0.05 were then removed to obtain predictors presenting statistical significance. In the next step, the SP was investigated to ascertain its correlation with the other variables, and those presenting R higher than 0.7 were removed from the list to remove any multicollinearity between variables. Mahmood & Babel (2012) mentioned that an R value that is not higher than 0.7 between two predictors indicates that they can be used as independent variables in regression analysis. Finally, the remaining variables were then calculated for percentage reduction in absolute partial correlation (PRP). The lower PRP values showed a higher number of significant variables for rainfall downscaling. Equation (1) explains the method used to calculate PRP, with $P \cdot r$ being the partial correlation coefficient and $R1$ the correlation coefficient between the selected predictor and local rainfall:

$$PRP = \left(\frac{P \cdot r - R1}{R1} \right) \quad (1)$$

Table 2 | Variables in the CanESM2 model

Variables	Definition	Variables	Definition
msslgl	Mean sea level pressure	p5zhgl	Divergence at 500 hPa
p1_fgl	Geostrophic airflow velocity at the surface	p8_fgl	Vorticity at 850 hPa
p1_ugl	Zonal velocity component at the surface	p8_ugl	Zonal velocity component at 850 hPa
p1_vgl	Meridional velocity component at the surface	p8_vgl	Meridional velocity component at 850 hPa
p1_zgl	Vorticity at surface	p8_zgl	Geostrophic airflow velocity at 850 hPa
p1thgl	Wind direction at the surface	p850gl	850 hPa geopotential height
p1zhgl	Divergence at the surface	p8thgl	Wind direction at 850 hPa
p5_fgl	Vorticity at 500 hPa	p8zhgl	Divergence at 850 hPa
p5_ugl	Zonal velocity component at 500 hPa	prepgl	Total precipitation
p5_vgl	Meridional velocity component at 500 hPa	s500gl	Relative humidity at 500 hPa height
p5_zgl	Geostrophic airflow velocity at 500 hPa	s850gl	Relative humidity at 850 hPa height
p500gl	500 hPa geopotential height	shumgl	Specific humidity
p5thgl	Wind direction at 500 hPa	temp	Mean temperature at 2 m

- Stepwise regression

In the SR method, a predictor variable (climate variable) showing the greatest correlation with the dependent variable (rainfall data) was selected as the first term to be calculated in the regression equation. The variables excluded from the equations were then examined to identify which predictors fit the equation and should be added. Likewise, the variables in the equation were examined to determine which predictor variables in the equation should be eliminated. The confidence level of 0.01 was used in this study. By evaluating its determination coefficient (R^2) value, if the predictor did not result in a statistically significant increase in R^2 , the variable was eliminated from the equation. The variables were selected at every step until none could be inserted or eliminated from the equation (Joshi *et al.* 2015; Osman & Abdellati 2016).

3.2.2. Constructing relationships between selected variables and observed rainfall

The selected variables from previous steps were forced to two statistical models, namely SDSM and the PCA, to seek relationships with local daily rainfall. Details of these two methods are as follows:

- **Statistical Downscaling Model**

The SDSM is a decision support tool, combining multiple linear regression and the stochastic weather generator developed by Wilby *et al.* (2002). The model requires two types of datasets, namely local variables (e.g., temperature and precipitation) and large-scale circulation variables (predictors) located in the grid closest to the study area. Empirical relationships between selected large-scale indices of the NCEP dataset and local variables were generated using multiple linear regression and their regression parameters produced during the process of calibration. The stochastic weather generator and the regression parameters were applied with GCM predictors to simulate up to 100 daily time-series to seek a suitable correlation with local data (Hussain *et al.* 2015). Finally, future climates were then predicted using the scenario generator operator in the SDSM with 20 ensemble members and the mean value of the ensemble dataset then represented as future climate-change data (Mekonnen & Disse 2018).

In this study, the SDSM model was set up following the instructions of Wilby & Dawson (2007); this application was also applied in the works of Mekonnen & Disse (2018), Molina & Bernhofer (2019), and Hassan & Hashim (2020). After screening variables using the methods presented above (Section 3.2.1), the process of calibration was continued using the selected variables of NCEP and daily rainfall observations during 1985–1995 and regression parameters were generated. The parameters together with the NCEP variables and measured daily rainfall during 1996–2005 were then employed in the weather generator function for the validation step. The scenario generator operation was then applied using the selected predictors supplied by CanESM2 for either historical or future climate to generate daily simulated rainfalls. This study particularly focused on three scenarios of climate change situation (RCP2.6, RCP4.5, and RCP8.5) with three different future periods (2020–2040, 2041–2070, and 2071–2100).

- **Principal Component Analysis**

PCA is a technique for establishing the correlation between datasets and grouping potentially correlated datasets and is widely used in multivariate statistical analysis (Wilks 1995). Constructing a correlation between variables using PCA helps to minimize the dimensionality of a dataset and to overcome relationship multicollinearity (Tootle *et al.* 2007). General statistics and the linear algebra concept consisting of covariance, standard deviation, eigenvectors, and eigenvalues are employed in PCA to develop the principal components (PCs). The PCs are similar in number to the independent variables (predictors) forced to PCA and the highest explained variances of the predictors shown in PC1 while decreasing in the following PCs. Equations (2) and (3) present the relationship between the three variables $[S]$, e' , λ indicating the variance-covariance matrix, eigenvector, and eigenvalue, respectively. PCs ($U_m; m = 1, 2 \dots m$) are then computed as Equation (4):

$$[S]e' = \lambda e' \quad (2)$$

$$[S] - \lambda[I]e' = 0 \quad (3)$$

$$U_m = e_m^T x' \quad (4)$$

The significant variables of NCEP were initially used in PCA and the principal component coefficients and PCs of the NCEP data were then computed at the first step of the model construction in this study. The regression technique was applied

using the PCs responsible for the first three high explained variances of the selected predictors of NCEP and daily measured rainfall, and regression coefficients were then generated for 1985–1995 in the calibration process. The coefficients were then used in the validation process between the selected NCEP predictors and daily rainfall for 1996–2005. The selected predictors of CanESM2 in the historical period and future period under three scenarios were finally employed in PCA using the parameters developed in the calibration process. Similar to the use of the SDSM application, only three future periods and three scenarios were discussed.

In summarize, statistical downscaling was used to select variables and establish relationships between them and can be summarized into four methods (Figure 2): M1: SP/PCA, M2: SP/SDSM, M3: SR/PCA, and M4: SR/SDSM. The most effective method for generating daily downscale rainfall data was then applied for generating future rainfall under climate change scenarios.

3.2.3. Projection of future rainfall under climate change scenarios

This research used outputs from the CanESM2 model under climate change scenario AR5 (RCP2.6, RCP4.5, and RCP8.5), forming the precipitation projections under all these three situations by station and spatial average for three time periods: near-future (2020–2040), mid-future (2041–2070), and far-future (2071–2100). The amount of future precipitation was employed to assess changes in rainfall and analyze its characteristics. Under the climate change scenario, model-based rainfall data were utilized for statistical analysis to determine the future rainfall trend. The amount of rainfall per station was assessed on an annual basis. To obtain a clearer view of the future rainfall trend across the basin, the spatial average rainfall was evaluated both monthly and annually.

The standardized anomaly index (SAI) was also utilized in this study to investigate rainfall variance in the Phetchaburi River Basin and the frequency of rainy and dry spells in the region under the climate change scenario. This index is widely applied to explain the characteristics of rainfall projections under climate change scenarios, for example, Blanka *et al.* (2013), Ademe & Eshetu (2021), Kwawuvi *et al.* (2022). The projected rainfalls for the entire future period were calculated and the results were reported for the near-, mid-, and far-term. Historical SAI values were calculated using rainfall observations. The SAI values can be determined using Equation (5), where positive and negative SAI values indicate the occurrence of heavy rainfall and drought, according to the consideration criteria presented in Table 3. P_i is the rainfall value in year i , μ is the average value of annual rainfall, and δ the standard deviation of rainfall in the study period:

$$\text{SAI} = \frac{P_i - \mu}{\delta} \quad (5)$$

3.3. Model performance estimation

To evaluate the performance of each statistical model, three objective functions, which were coefficient of determination (R^2), Nash–Sutcliffe efficiency (NSE) and root mean square error (RMSE), were applied in this study. NSE was initially considered for the model performance. Moriasi *et al.* (2007) indicated that NSE can identify the relative magnitude of the residual variance compared with the variance of observations. R^2 and RMSE, which can explain the linear relationship and the average of the different amounts between observed and simulated rainfalls, respectively, were then examined.

- Coefficient of determination (R^2)

R^2 is an index that shows the relationship of data from linear regression analysis. The value is between 0 and 1. If the R^2 value approaches 1, this indicates that the independent and dependent variables are highly correlated, whereas if R^2 approaches 0, the independent and dependent variables will be less correlated. It can be calculated using Equation (6):

$$R^2 = \left(\frac{\sum_{i=1}^n (O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^n (O_i - \bar{O})^2} \sqrt{\sum_{i=1}^n (P_i - \bar{P})^2}} \right)^2 \quad (6)$$

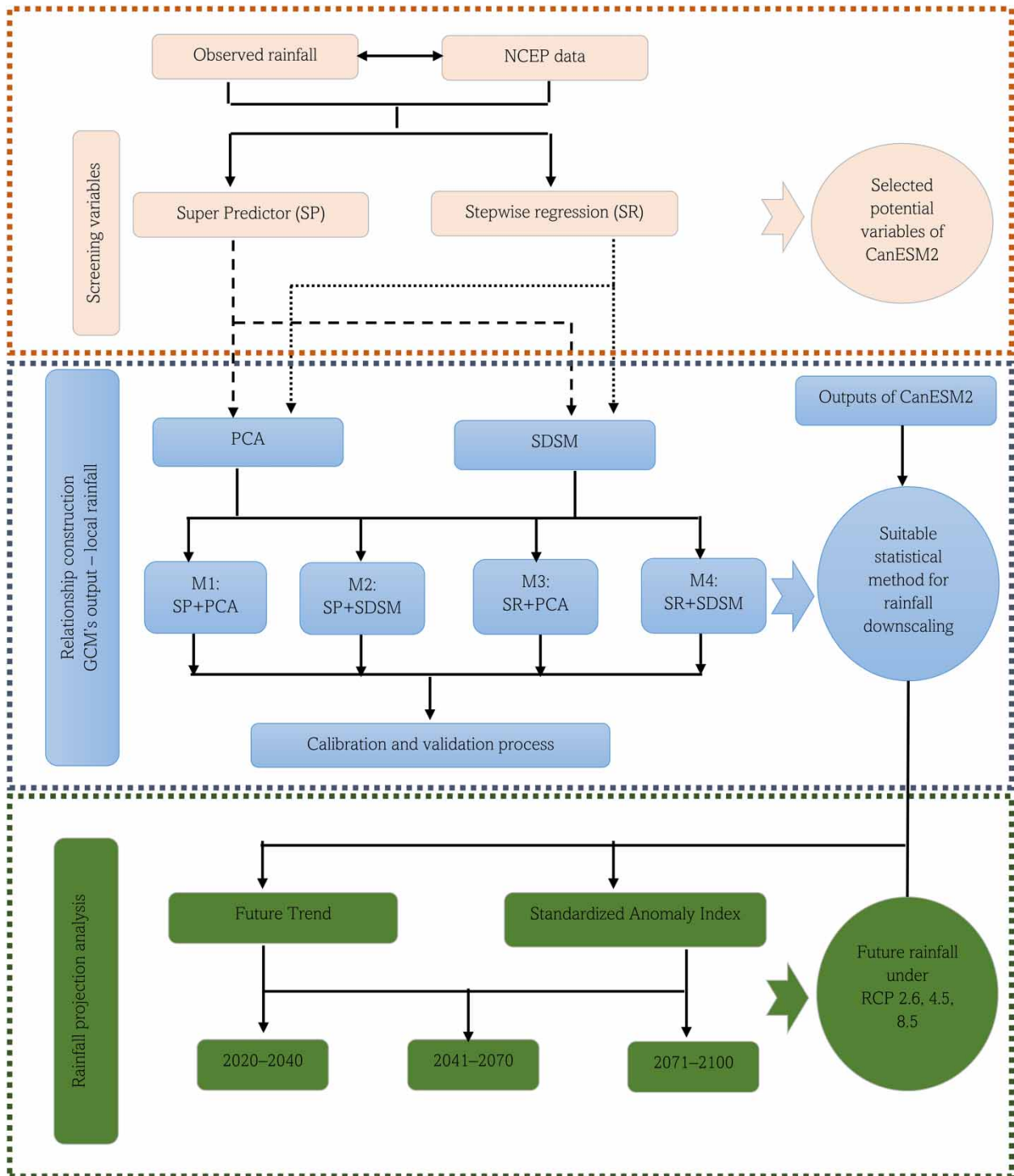


Figure 2 | Study procedures.

- Nash–Sutcliffe efficiency

NSE is used to display the model’s validity, illustrating the relationship between the measured rainfall variance and the simulated rainfall variance, ranging from $-\infty$ to 1. If the NSE is close to 1, it indicates that the model-based values are close to the measured values, and if equal to 1, the model can predict data without error, as computed by Equation (7):

Table 3 | Criteria for determining the severity of SAI

SAI value	Level of severity
≥ 2	Extremely wet
1.50–1.99	Very wet
1.00–1.49	Moderately wet
–0.99 to 0.99	Near normal
–1.00 to –1.49	Moderately dry
–1.50 to –1.99	Severely dry
≤ -2	Extremely dry

$$\text{NSE} = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \bar{P}_i)^2} \quad (7)$$

- Root Mean Square Error

RMSE illustrates the difference between model-based and observed rainfall data. If the RMSE value is low, it implies a small deviation in the simulated rainfall data from the measured rainfall and accurately imitates its statistical characteristics. It can be calculated using Equation (8):

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - O_i)^2} \quad (8)$$

O is the measured rainfall volume, P is the model-based rainfall volume, \bar{O} is the average measured rainfall volume, and \bar{P} is the average model-based rainfall volume.

4. RESULTS

4.1. Predictor selection

The SP variable selection results indicated that two to four variables were selected for each station out of a total of 26 variables. As expected, the rainfall variable (prcpgl) was considered as a potential predictor showing the most frequent correlation with the measured rainfall for all stations, followed by geostrophic airflow velocity at 850 hPa (p8_zgl) demonstrating a correlation with eight stations. In addition, a divergence of 850 hPa (p8zhgl) also showed a relationship with the rainfall data obtained from seven rain gauge stations (Table 4).

Predictor selection using SR revealed that the number of variables selected for each station was between eight and ten more than when using SP. The precipitation for all stations was also selected using the SP method. However, there were also a larger number of other potential predictors showing for all ten stations compared with SP, such as the zonal velocity component at 850 hPa (p8_ugl), geostrophic airflow velocity at 850 hPa (p8_zgl), and relative humidity at 850 hPa (s500gl). Other interesting large circulation indices responsible for rainfall events at the majority of stations were geostrophic airflow velocity at the surface (p1_fgl) and the meridional velocity component at 850 hPa (p8_vgl).

Comparing these two predictor selection methods, the SP obtained fewer selected predictors. This was because the SP method considered the partial correlation value coupled with the percentage reduction in the correlation of each variable to eliminate it from the equation, whereas the SR method considered only the value of R^2 after inputting the variable into the equation and eliminating the variable when it did not result in a significant increase in R^2 . Details of the predictor selection are shown in Table 4. The number of predictors used in the statistical downscaling process was usually different depending on the methods performed in the screening step and location of study area. Examples of selected predictors for rainfall downscaling can be seen in Teegavarapu & Goly (2018), González-Roji *et al.* (2019), and Molina & Bernhofer (2019).

Table 4 | Results of variable selection

No.	Variable	Variable selection by station									
		424004 SP SR	465002 SP SR	465004 SP SR	465005 SP SR	465006 SP SR	465010 SP SR	465012 SP SR	465201 SP SR	500202 SP SR	500301 SP SR
1	mslpgl		○	○	● ○		○	○		○	
2	p1_fgl	○			○	○	○	○	○	○	○
3	p1_ugl										○
4	p1_vgl		○								
5	p1_zgl	○	○						○	○	
6	p1thgl			○					○		
7	p1zhgl	○	○							○	
8	p5_fgl		○	○				○			○
9	p5_ugl									●	
10	p5_vgl										
11	p5_zgl	○	○		●	○		○	○		○
12	p5thgl										
13	p5zhgl										
14	p8_fgl	○			○		○	○	○	○	
15	p8_ugl	○	○	○	○	○	○	○	○	○	○
16	p8_vgl		○	○	○	○	○	○	○	○	
17	p8_zgl	● ○	● ○	○	● ○	○	● ○	● ○	● ○	○	● ○
18	p8thgl	●									○
19	p8zhgl	●	●	●		●	●	●		●	●
20	p500gl								●		
21	p850gl		●								
22	prcpgl	● ○	● ○	● ○	● ○	● ○	● ○	● ○	● ○	● ○	● ○
23	s500gl	○	○	○	○	○	○	○	○	● ○	○
24	s850gl										○
25	shumgl					○		○			
26	tempgl										

Note: SP (●) is the super predictor method and SR (○) is the stepwise regression method.

4.2. Model calibration and verification

4.2.1. Calibration and verification results of all four models

The predictors selected from the previous step were then forced into two statistical models (PCA and SDSM). The data were divided into two sections: calibration and verification periods 1985–1995 and 1996–2005, respectively. Three objective functions (R^2 , NSE, and RMSE) were calculated to estimate the model performance by comparing the model results with the rainfall observations for each station (Table 5). The results for all four models are presented in the following sections.

• M1 (SP/PCA)

The M1 model used SP for predictor selection followed by PCA to generate the relationship between selected predictors and local rainfall. During the calibration periods, this model showed moderate skills in all stations with R^2 ranging from 0.33–0.55, NSE 0.33–0.53, and RMSE 56.37–81.87 mm/month. However, a significant difference in the validation period could be observed in many stations with lower performance compared with the calibration period, except for ST-424004,

Table 5 | Results of model calibration and validation

Station code		M1 (SP-PCA)			M2 (SP-SDSM)			M3 (SR-PCA)			M4 (SR-SDSM)		
		R^2	NSE	RMSE	R^2	NSE	RMSE	R^2	NSE	RMSE	R^2	NSE	RMSE
424004	Cal	0.55	0.53	72.66	0.60	0.60	66.86	0.71	0.66	62.09	0.62	0.62	65.27
	Val	0.52	0.51	66.08	0.65	0.65	55.75	0.68	0.65	55.51	0.65	0.65	55.56
465002	Cal	0.50	0.47	78.57	0.63	0.62	66.29	0.64	0.58	69.55	0.65	0.64	65.05
	Val	0.26	0.25	92.18	0.53	0.53	72.89	0.47	0.43	80.02	0.51	0.50	74.81
465004	Cal	0.43	0.41	78.54	0.58	0.57	66.76	0.59	0.55	68.63	0.64	0.63	62.26
	Val	0.28	0.28	80.92	0.53	0.52	65.91	0.40	0.40	74.15	0.48	0.46	70.25
465005	Cal	0.35	0.33	56.37	0.50	0.44	51.33	0.53	0.48	49.35	0.52	0.47	50.00
	Val	0.24	0.12	88.13	0.54	0.53	64.53	0.46	0.31	78.02	0.51	0.53	66.71
465006	Cal	0.44	0.42	67.08	0.54	0.52	61.09	0.58	0.56	58.44	0.57	0.58	58.65
	Val	0.36	0.42	74.66	0.64	0.67	56.05	0.53	0.55	62.34	0.62	0.62	57.50
465010	Cal	0.33	0.33	75.80	0.40	0.40	72.09	0.56	0.50	65.57	0.50	0.50	66.02
	Val	0.20	0.18	82.74	0.55	0.54	62.05	0.38	0.35	74.08	0.51	0.51	64.36
465012	Cal	0.45	0.44	57.75	0.56	0.50	54.16	0.72	0.65	46.05	0.56	0.48	55.29
	Val	0.21	0.19	100.49	0.51	0.50	78.97	0.48	0.35	90.05	0.63	0.60	72.78
465201	Cal	0.50	0.47	77.60	0.56	0.55	71.00	0.59	0.55	71.22	0.60	0.60	67.56
	Val	0.39	0.39	74.26	0.62	0.61	58.91	0.53	0.51	66.24	0.63	0.62	58.19
500202	Cal	0.41	0.40	81.87	0.50	0.49	75.14	0.55	0.51	74.00	0.56	0.55	70.60
	Val	0.21	0.19	93.27	0.42	0.42	79.25	0.45	0.43	78.26	0.42	0.43	78.25
500301	Cal	0.46	0.46	71.77	0.59	0.59	62.56	0.70	0.67	56.00	0.61	0.61	61.10
	Val	0.24	0.23	100.58	0.47	0.47	83.76	0.47	0.46	84.32	0.47	0.47	83.61

which showed a slight difference between these two periods with R^2 , NSE, and RMSE of 0.52, 0.51, and 66.08 mm/month, respectively. The model results for most stations (seven) exhibited low performance with R^2 ranging from 0.20–0.28, 0.12–0.28, and 80.92–100.58 mm/month during the validation period. The other two stations presented moderate skills with R^2 in the range 0.36–0.39, NSE 0.39–0.42, and RMSE 74.26–74.66.

• M2 (SP/SDSM)

This model used SDSM for generating a correlation between large circulation indices selected by SP. Although the M1 and M2 models were forced with the same predictors, significant differences in the results for these two models could be observed. For the M2 model, one station indicated high skills (R^2 of 0.63, NSE of 0.62, RMSE of 66.29 mm/month) whereas the remaining stations showed moderate skills during the calibration period (R^2 0.40–0.63, NSE 0.40–0.62, RMSE 51.33–75.14 mm/month). These results were similar to the M1 model in the calibration period; however, large differences between these two models could be found during the validation period. In the validation period, the performance of the M2 model at five stations was close to that in the calibration period (R^2 0.42–0.53, NSE 0.42–0.53, RMSE 65.91–83.76 mm/month), while the remaining stations presented higher skills than in the calibration period with R^2 and NSE in the range 0.54–0.65 and RMSE 55.75–64.53 mm/month.

• M3 (SR/PCA)

The M3 model used different predictors compared with the two models presented previously. SR together with PCA was applied to this model. Three stations (ST-424004, ST-465012, and ST-500301) presented high agreement between rainfall using this model while local rainfall for R^2 , NSE, and RMSE were in the range 0.70–0.71, 0.65–0.67, and 46.05–62.09 mm/month, respectively, whereas a moderate performance was found for the remaining seven stations (R^2 0.53–0.64, NSE 0.48–0.58, RMSE 49.35–74.00 mm/month) during the calibration period. However, during the validation period, only one station (ST-424004) generated a good simulation corresponding to the observed rainfall (R^2 0.68, NSE 0.65, RMSE 55.51 mm/month). The simulated rainfall for the other nine stations was within the moderate-level R^2 in the range 0.38–0.53, NSE 0.31–0.55, and RMSE 62.34–90.05 mm/month.

• M4 (SR/SDSM)

Selected predictors using SR were forced to the SDSM for rainfall simulation in the M4 model. The rainfall generation for five stations showed high agreement with the local rainfall (R^2 in the range 0.60–0.65, NSE 0.60–0.64, RMSE 61.10–65.27 mm/month) during the calibration period, while rainfall simulation for the other five stations gave moderate agreement with the observations. For the validation period, four stations were determined to have high skill levels for rainfall simulation using this model with R^2 in the range 0.62–0.65, NSE 0.60–0.65, RMSE 55.56–72.78 mm/month. It is worth mentioning that two stations showing rainfall simulations presented better agreement with local rainfall in the validation period compared with that in the calibration period.

4.2.2. Model performance comparison

Although forced with the same predictors, the rainfall simulations for the M1 and M2 models were generated using PCA and SDSM, respectively. The results revealed that SDSM exhibited significantly better performance, with higher values of R^2 and NSE and lower RMSE for the majority of stations compared with PCA during both the calibration and validation periods. The M3 and M4 models produced similar results using the same predictors selected by SR. Higher simulation skills could be seen in the M4 model at most stations compared with the M3 model. The M3 and M4 models showed slight differences in model efficiency, while the M1 and M2 models exhibited larger differences.

It can also be observed that different predictors were used as input variables in PCA (M1 and M3), generating rainfall simulation with greater sensitivity compared with the SDSM (M2 and M4). The M1 and M3 model estimators exhibited significant differences at most stations, in contrast to the M2 and M4 models which showed slight differences in all three objective functions. Considering the agreement between rainfall simulation and observation at all stations, the results of the M4 model for six stations showing higher NSE values (ST-465002, ST-465004, ST-465006, ST-465010, ST-465201, and ST-500202) compared with the other stations during the calibration period and most stations presented slightly lower NSE values during the validation period except for ST-465006 and ST-465201. For ST-465005, insignificant differences of R^2 , NSE, and RMSE between the M3 and M4 models were revealed during the calibration section but the results during the validation part of M3 was significantly lower with R^2 and NSE of 0.46 and 0.31 and high RMSE of 78.02 mm compared with the result of M4. The M4 model exhibited the better performance for seven out of ten stations, so this model was then selected to predict future rainfall under climate change scenarios.

4.3. Rainfall projections under climate change scenarios

Changes in rainfall were measured using the forecast rainfall under climate change scenarios RCP2.6, RCP4.5, and RCP8.5 in three future periods, namely near-future (2020–2040), mid-future (2041–2070), and far-future (2071–2100), and compared with the historical period (1985–2019) as presented in Figure 3 and Table 6. Figure 3 illustrates that large variations of the change of rainfall projections compared with the measured rainfalls for all ten stations could be clearly seen for RCP8.5

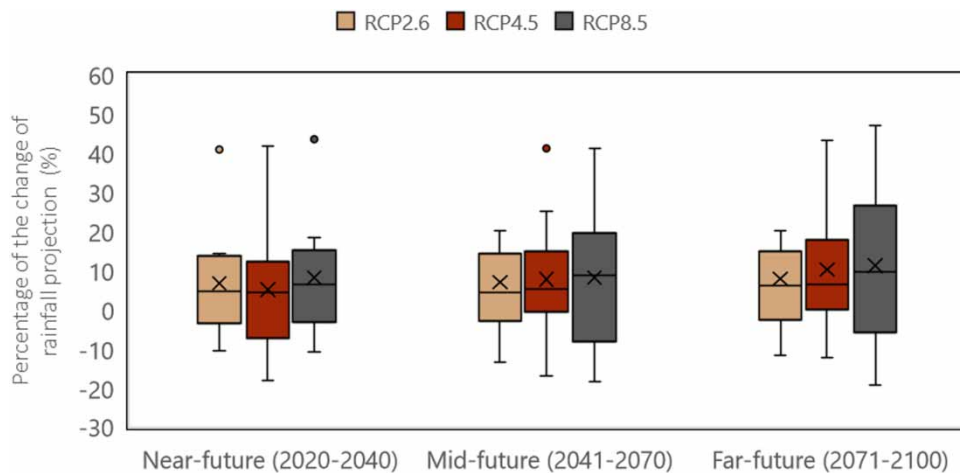


Figure 3 | Rainfall projection changes compared with the historical period (1985–2019) under three climate change scenarios.

Table 6 | Annual rainfall projections under the RCP2.6, RCP4.5, and RCP8.5 scenarios

Predicted rainfall	Percentage of rainfall change (%)									
	424004	465002	465004	465005	465006	465010	465012	465201	500202	500301
Measured rainfall (mm)										
1985–2019	1,078	893	896	778	1,012	828	916	1,022	976	1,150
Forecast rainfall (mm)										
Near-future										
RCP2.6	7.8	41.0	14.3	−10.5	7.4	−3.0	13.7	−4.8	−0.5	1.8
RCP4.5	6.4	41.8	15.4	−18.0	5.4	−6.3	11.4	−10.1	3.7	1.9
RCP8.5	8.1	43.6	18.4	−10.8	7.9	−2.2	14.0	−5.8	3.7	5.1
Mid-future										
RCP2.6	7.8	42.4	20.3	−13.5	6.6	−0.8	12.2	−5.2	2.1	−2.0
RCP4.5	6.7	41.3	25.1	−16.9	7.9	3.6	11.5	−7.3	3.7	1.9
RCP8.5	5.9	41.1	36.8	−18.3	11.5	12.9	13.7	−16.2	0.9	−5.6
Far-future										
RCP2.6	7.1	41.7	20.2	−11.7	9.2	−1.9	13.0	−4.1	5.3	0.6
RCP4.5	3.8	43.2	34.1	−12.1	11.3	8.0	12.4	−5.1	5.2	1.9
RCP8.5	5.2	47.2	39.5	−19.1	22.3	12.8	11.1	−9.0	8.1	−4.7

during mid- and far-future. A higher percentage of the average rainfall projection change under RCP8.5 was presented while the slight difference of the projections compared with observations under RCP2.6 and RCP4.5 for all three periods was exhibited. Table 6 presents the slight decrease of annual average rainfall at three stations by around 2.2%–18.0% for all scenarios in the near-future but the majority of stations presented rainfall increments of 1.8%–18.4%. The highest increase in annual rainfall was predicted for ST-465002, ranging from 41.0% to 43.6%. Annual rainfall in the mid- and far-future periods was similar to that in the near-future except for only two stations, which presented decreasing annual rainfall. Future rainfall in these two future periods showed a slightly higher decrease compared with that in the near-future, with the largest reduction occurring in RCP8.5 of 19.1%. Eight stations generally exhibited rainfall increments during the mid- and far-future with RCP8.5 indicating the largest percentage of rainfall increase, especially for the far-future. The same station showed the highest rainfall increase for the mid- and far-future as for the near-future with RCP8.5 (47.2%).

According to the spatial estimation of average precipitation observations at all ten stations in the Phetchaburi River Basin, during 1985–2019, the average precipitation was around 960 mm/year with the distribution occurring from May to October. Rainfall starts to increase in May and decreases slightly in the following month while increasing to the highest level of rainfall in October. The model produced a pattern of rain distribution similar to the measured rainfall when monthly precipitations were forecast under the RCP2.6, RCP4.5, and RCP8.5 scenarios for the three future periods. The monthly rainfall amount was not significantly different from the measured rainfall, as shown in Figure 4. In the far-future, the amount of precipitation increased to its maximum (276 mm) in October under the RCP8.5 scenario compared with the other scenarios and other periods and decreased to its minimum in January. When comparing the average and measured annual rainfall, it was found that in the far-future, under the RCP8.5 scenario, annual average rainfall increased to the highest level of 6%, as shown in Figure 4.

4.4. Analysis of wet and dry years under climate change scenarios

Rainfall intensity and drought were analyzed using the SAI as shown in Figure 5. Only the years with SAI severity values greater than 0.99 and less than −0.99 are illustrated in the figure. Slightly larger variations of SAI of rainfall projections compared with observations could be seen in the near-future with the RCP2.6 scenario and incidents of severely to extremely dry occurred more often for all scenarios in the future except for the RCP2.6 scenario in the near-term and the RCP8.5 scenario in the far-term (Figure 5(a)). Figure 5(b)–5(e) illustrates wet and dry occurrences in each year. In the historical period, moderately to extremely wet years could be observed more frequently compared with moderately to extremely dry years. The highest

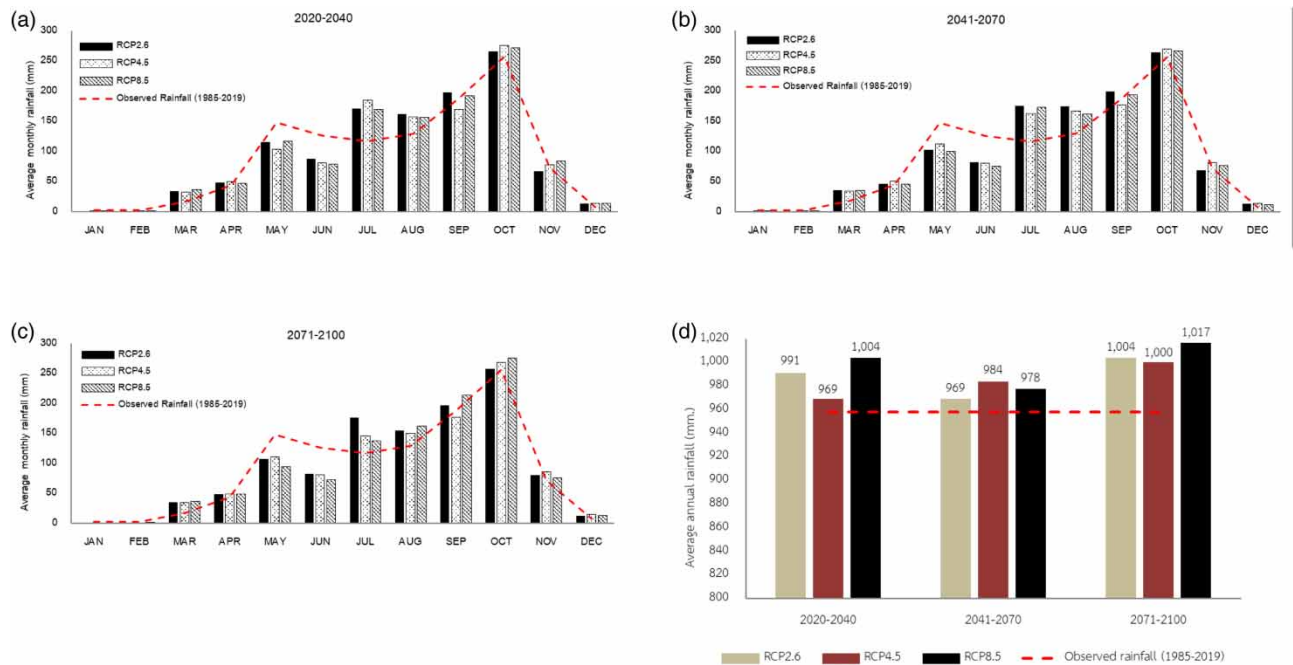


Figure 4 | Comparison of the monthly (a, b, c) and annual (d) spatial average rainfall in the three future periods under various scenarios.

SAI value occurred (extremely wet) in 2007 followed by 2017, which was a very wet year. Two years of severely dry weather occurred in 1990 and 2015 (Figure 5(b)).

Figures 5(c)–5(e) show the projected rainfall under the three climate change scenarios for three periods. In the near-future, under the RCP2.6 scenario, extremely and moderately wet events were presented in 2032 and 2040, respectively, while moderately dry was predicted to occur once in 2036. Moderately to severely dry events under the RCP4.5 scenario were predicted to happen the most frequently for four years. For RCP8.5, two severely dry weather events were shown to occur in 2030 and 2036 while three events of moderately to very wet were expected in 2028, 2029, and 2033. It is worth mentioning that both moderately and severely dry weather in the RCP2.6 and the RCP8.5 scenarios were produced to occur in the same year, namely 2036.

For the mid-period, wet and dry severe events were remarkably expected to occur more frequently than in the near-period. According to the SAI, an extremely wet event was shown in 2051 while extremely dry events were likely to occur the most frequently for four years under the RCP2.6 scenario. Very wet and severely dry incidents were predicted in 2047 and 2060 under RCP4.5, respectively. Under the RCP8.5 scenario, six years were to be in moderately and severely dry categories while no result was presented for moderately to extremely wet. The year 2059 was predicted to be moderately dry for both RCP4.5 and RCP8.5 scenarios whereas the year 2041 was expected to be extremely dry under the RCP2.6 scenario and moderately dry under the RCP8.5 scenario.

In the far-future, the years showing in moderately to extremely wet categories were noticeably seen more frequently compared with the near- and mid-term. Extremely wet events were likely for two years under the RCP4.5 scenario, and three years under the RCP8.5 scenario, while the most extremely dry condition was predicted for 2080 (RCP4.5). Five years were expected to be in the very wet category; two under the RCP2.6 scenario (2080 and 2092) and three under the RCP4.5 scenario (2095–2097). Severely dry events were also expected in three years under the RCP2.6 scenario for 2093, and RCP4.5 and RCP8.5 scenarios for 2098 and 2083, respectively. It should be emphasized that SAI values under the three scenarios falling within the moderately to extremely wet categories occurred in 2096.

5. DISCUSSION

Four statistical downscaling approaches were developed in this study, consisting of two main steps: predictor selection and correlation establishment. The 26 predictors generated by CanESM2 were selected in the first step using two statistical

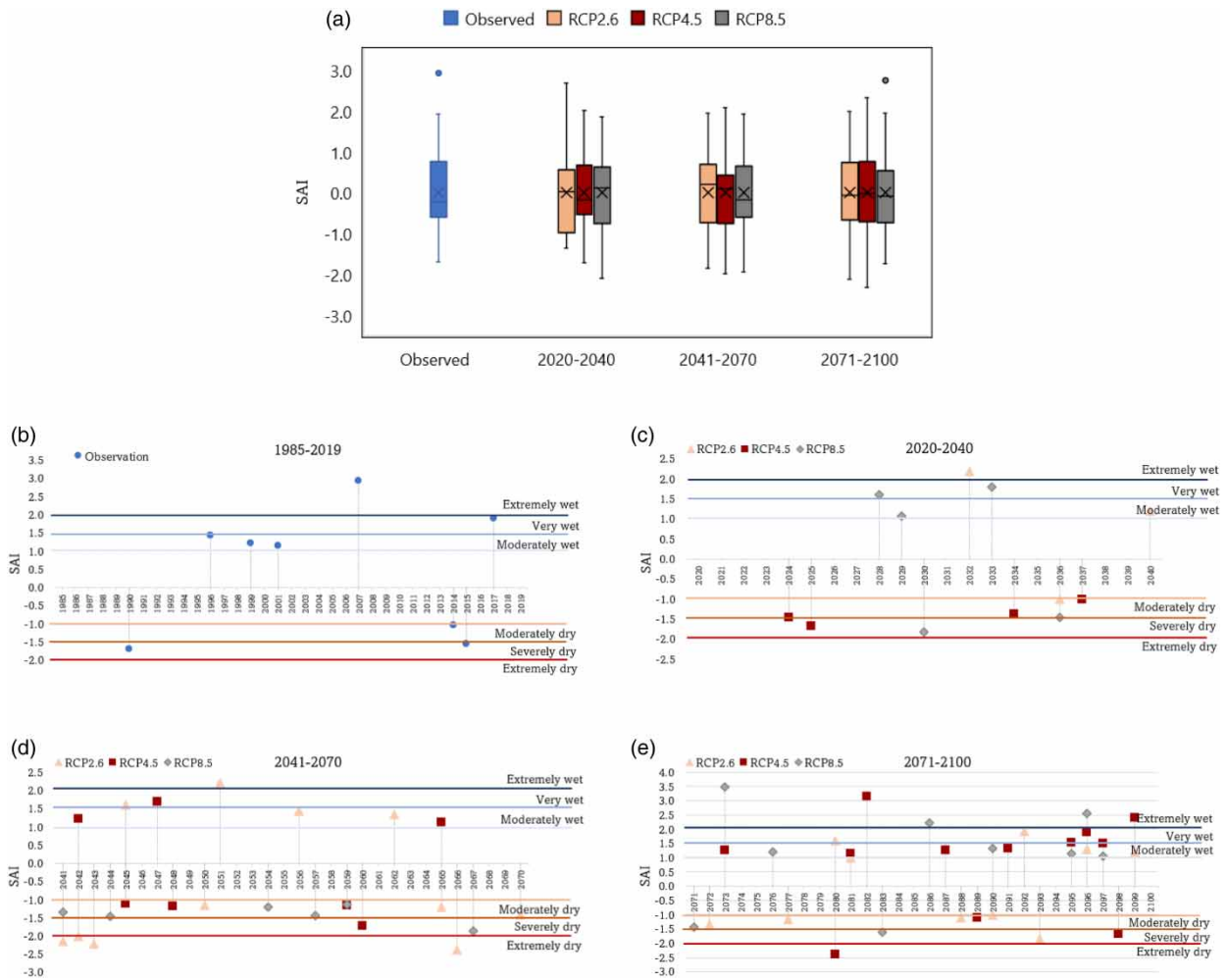


Figure 5 | SAI of rainfall projections under three climate change scenarios.

methods: SP and SR. The results revealed that SR was able to select more predictors for all ten rain gauge stations. However, agreement could be observed in the potential predictors of rainfall and geostrophic airflow velocity at 850 hPa for both selection methods. The SP also indicated divergence at 850 hPa, while the geostrophic airflow velocity at the surface, zonal velocity component at 850 hPa, meridional velocity component at 850 hPa, and relative humidity at 850 hPa height were considered as significant predictors by SR. The use of multiple predictors for rainfall downscaling was applied by various research works, for example, *Tahir et al. (2018)*, *González-Roji et al. (2019)*, and *Sadeghfam et al. (2021)* to obtain satisfactory results of downscaled rainfall. These selected predictors were then forced to use the same statistical methods (PCA: M1 and M3; and SDSM: M2 and M4) for correlation construction. The results revealed that the model using selected predictors generated by SR gave a better performance compared with the one employing the SP selection method.

To estimate the ability of the two methods for generating a correlation between predictors and local rainfall, the results of the M1–M2 model and the M3–M4 model were compared. The SDSM presented its ability to better simulate rainfall data corresponding to observations than PCA. The significance of this difference could be observed when the M1 and M2 models were compared, while slight differences were expressed in the M3 and M4 comparison. The performance of SDSM was indicated for accurately developing long-term mean rainfalls in the work of *Mekonnen & Disse (2018)*. Finally, M4 was found to be the best model for achieving rainfall downscaling and was therefore selected for rainfall projection under climate change scenarios in the Phetchaburi River Basin. Less complexity of calculation for all models could be found with insignificant differences due to only statistical approaches applied in the models. However, the SDSM software consists of multi-functions with pre-screening, calibration–validation process statistical analysis and future generation under climate

change scenarios (Wilby *et al.* 2002). This tool is friendly for users, but the preparation process of input data generated by GCMs is required. It should be noted that this study attempted to downscale rainfall data for specific sites in the area, and that additional processes were required for the study of spatial rainfalls under climate change scenarios. In addition, one of the challenging parts of rainfall downscaling is the requirement of complete local rainfall data. Various areas may experience low quality of observed rainfall. Interesting techniques, for example, gridded climate data, could benefit the work of rainfall downscaling in the future (Araghi *et al.* 2023).

Under climate change scenarios RCP2.6, RCP4.5, and RCP8.5 generated using CanESM2, this basin faced higher rainfall in all three scenarios with the highest increment being under the RCP8.5 scenario. The far-future (2071–2100) exhibited the largest increase compared with the other future periods. The rainfall patterns under these three scenarios were not significantly different. However, increasing rainfall was predicted during the dry season (March) and rainy season (September to October). It is worth noting that by considering the rainfall projection of each station, they were sensitive to different predictors applied in the statistical model as it showed significant differences in some stations. The variation of rainfall projections in different climate stations within a basin can also be found in Kwawuvi *et al.* (2022) and Ademe & Eshetu (2021). In addition, extreme rainfall events could be observed for both wet and dry situations. The years exhibiting the highest SAI values calculated using rainfalls under the three climate change scenarios within the three periods (near-future: 2020–2040, mid-future: 2041–2070, and far-future: 2071–2100) received particular attention; for example, 2032 (the near-future), 2051 (the mid-future), and 2073, 2082, 2086, 2096, and 2099 (the far-future) for wet events, and 2041–2043 and 2066 (the mid-future) and 2080 (the far-future) for dry events. However, results of these extreme events were generated under climate change scenarios (RCP2.6, RCP4.5, and RCP8.5) with different assumptions related to various issues such as economic, social, and physical changes, population growth and technological development. Therefore, carefully verifying the rainfalls under these scenarios compared with current situations is strongly suggested. In addition, the outputs of various GCMs are strongly recommended for emphasizing the assessment of climate change impacts in the study area.

6. CONCLUSION

This study proposes a statistical downscaling method to downscale rainfall at rain gauge stations and investigates projected rainfall under climate change scenarios RCP 2.6, 4.5, and 8.5. Observations from the Phetchaburi River Basin, Thailand, were used in this study. Two statistical methods for predictor selection (SP/SR) and two approaches for correlation construction (PCA and statistical downscaling) were combined to downscale and project the rainfall at ten stations. The results revealed that SR exhibited a better performance in selecting predictors than the SP method, while SDSM also showed better skills for establishing a correlation between predictor and predictand. Rainfall and airflow velocity indices were found to be the best predictors out of the 26 variables generated using CanESM2. Rainfall increments were found in the Phetchaburi River Basin under all three scenarios. Calculated using the SAI, the most frequent very to extremely wet years are predicted for the far-future under the three scenarios, and severely to extremely dry years for the mid-future.

ACKNOWLEDGEMENTS

The authors are thankful to the Thai Meteorological Department for providing valuable data used in this study.

DATA AVAILABILITY STATEMENT

All relevant data are available from an online repository or repositories: Output of CanESM2 developed by the Canadian Centre for Climate Modelling and Analysis in Environment and Climate Change Canada can be accessed at <https://climate-scenarios.canada.ca/?page=pred-canesm2>. Rainfall data of the study area can be requested at the Thai Meteorological Department. Readers should contact the corresponding author for details.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

- Ademe, F. & Eshetu, M. 2021 Spatial analysis of climate variability and change in the Great Ethiopian Rift Valley basins. *Journal of Environment and Earth Science* **11** (16), 1–12.

- Ahsan, S., Bhat, M. S., Alam, A., Farooq, H. & Shiekh, H. A. 2021 Evaluating the impact of climate change on extreme temperature and precipitation events over the Kashmir Himalaya. *Climate Dynamics* **58**, 1651–1669.
- Araghi, A., Martinez, C. J. & Adamowski, J. F. 2023 Evaluation of TerraClimate gridded data across diverse climates in Iran. *Earth Science Informatics* **16**, 1347–1358.
- Armain, M. Z. S., Hassan, Z. & Harun, S. 2021 Climate change impact under CanESM2 on future rainfall in the state of Kelantan using artificial neural network. *Earth and Environmental Science* **646**, 012033.
- Barokar, Y. J., Saraf, V. R. & Regulwar, D. G. 2019 Simulating maximum temperature for future time series on lower Godavari basin, Maharashtra State, India by using SDSM. In: *Proceedings of the 11th World Congress of EWRA on Water Resources and Environment*, (Garrote, L., Tsakiris, G., Tshirintzis, V. A., Vangelis, H. & Tigkas, D., eds), European Water Resources Association, Athens, Greece, pp. 445–446.
- Benestad, R. E., Chen, D., Mezghani, A., Fan, L. & Parding, K. 2015 On using principal components to represent stations in empirical–statistical downscaling. *Tellus A: Dynamic Meteorology and Oceanography* **67** (1), 28326.
- Blanka, V., Mezösi, G. & Meyer, B. 2013 Projected changes in the drought hazard in Hungary due to climate change. *Időjárás: Quarterly Journal of the Hungarian Meteorological Service* **117** (2), 219–237.
- Gebrechorkos, S. H., Hülsmann, S. & Bernhofer, C. 2019 Statistically downscaled climate dataset for East Africa. *Scientific Data* **6**, 31.
- González-Rojí, S. J., Wilby, R. L., Sáenz, J. & Ibarra-Berastegi, G. 2019 Harmonized evaluation of daily precipitation downscaled using SDSM and WRF + WRFDA models over the Iberian Peninsula. *Climate Dynamics* **53**, 1413–1433.
- Hassan, W. H. & Hashim, F. S. 2020 The effect of climate change on the maximum temperature in Southwest Iraq using HadCM3 and CanESM2 modelling. *SN Applied Sciences* **2**, 1494.
- Hernanz, A., García-Valero, J. A., Domínguez, M., Ramos-Calzado, P., Pastor-Saavedra, M. A. & Rodríguez-Camino, E. 2021 Evaluation of statistical downscaling methods for climate change projections over Spain: present conditions with perfect predictors. *International Journal of Climatology* **42**, 762–776.
- Hua, W., Chen, H., Sun, S. & Zhou, L. 2014 Assessing climatic impacts of future land use and land cover change projected with the CanESM2 model. *International Journal of Climatology* **35**, 3661–3675.
- Huang, J., Zhang, J., Zhang, Z., Xu, C., Wang, B. & Yao, J. 2011 Estimation of future precipitation change in the Yangtze River basin by using statistical downscaling method. *Stochastic Environmental Research and Risk Assessment* **25**, 781–792.
- 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. *WIT Transactions on Ecology and the Environment* **196**, 269–278.
- Hussain, M., Yusof, K. W., Mustafa, M. R., Mahmood, R. & Shaofeng, J. 2017 Projected changes in temperature and precipitation in Sarawak state of Malaysia for selected CMIP5 climate scenarios. *International Journal of Sustainable Development and Planning* **12**, 1299–1311.
- Javaherian, M., Ebrahimi, H. & Aminnejad, B. 2021 Prediction of changes in climatic parameters using CanESM2 model based on RCP scenarios (case study): Lar dam basin. *Ain Shams Engineering Journal* **12**, 445–454.
- Joshi, D., St-Hilaire, A., Ouarda, T. & Daigle, A. 2015 Statistical downscaling of precipitation and temperature using sparse Bayesian learning, multiple linear regression and genetic programming frameworks. *Canadian Water Resources Journal* **40** (4), 392–408.
- Khadka, D. & Pathak, D. 2016 Climate change projection for the Marsyangdi River Basin, Nepal using statistical downscaling of GCM and its implications in geodisasters. *Geoenvironmental Disasters* **3**, 15.
- Kwawuvi, D., Mama, D., Agodzo, S. K., Hartmann, A., Larbi, I., Bessah, E., Limantol, A. M., Dotse, S.-Q. & Yangouliba, G. I. 2022 Spatiotemporal variability and change in rainfall in the Oti River Basin, West Africa. *Journal of Water and Climate Change* **13** (3), 1151–1169.
- Loganathan, P. & Mahindrakar, A. B. 2021 Statistical downscaling using principal component regression for climate change impact assessment at the Cauvery river basin. *Journal of Water and Climate Change* **12** (6), 2314–2324.
- Mahmood, R. & Babel, M. S. 2012 Evaluation of SDSM developed by annual and monthly sub-models for downscaling temperature and precipitation in the Jhelum basin, Pakistan and India. *Theoretical and Applied Climatology* **113** (1–2), 27–44.
- Mekonnen, D. F. & Disse, M. 2018 Analyzing the future climate change of Upper Blue Nile River basin using statistical downscaling techniques. *Hydrology and Earth System Sciences* **22**, 2391–2408.
- Molina, O. D. & Bernhofer, C. 2019 Projected climate changes in four different regions in Colombia. *Environmental Systems Research* **8**, 33.
- Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D. & Veith, T. L. 2007 Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the American Society of Agricultural and Biological Engineers* **50** (3), 885–900.
- O'Connell, E. 2017 Towards adaptation of water resource systems to climatic and socio-economic change. *Water Resources Management* **31**, 2965–2984.
- Osman, Y. Z. & Abdellatif, M. E. 2016 Improving accuracy of downscaling rainfall by combining predictions of different statistical downscale models. *Water Science* **30**, 61–75.
- Padhiary, J., Patra, K. C., Dash, S. S. & Uday Kumar, A. 2020 Climate change impact assessment on hydrological fluxes based on ensemble GCM outputs: a case study in eastern Indian river basin. *Journal of Water and Climate Change* **11** (4), 1676–1694.
- Pathan, A. S. & Waikar, M. L. 2020 Future assessment of precipitation and temperature for developing urban catchment under impact of climate change. *International Journal of Recent Technology and Engineering* **8**, 3395–3404.

- Rahman, G., Rahman, A., Munawar, S., Moazzam, M. F. U., Dawood, M., Miandad, M. & Panezai, S. 2022 Trend analysis of historical and future precipitation projections over a diverse topographic region of Khyber Pakhtunkhwa using SDSM. *Journal of Water and Climate Change* **13** (11), 3792–3811.
- Sadeghfam, S., Khatibi, R., Moradian, T. & Daneshfaraz, R. 2021 Statistical downscaling of precipitation using inclusive multiple modelling (IMM) at two levels. *Journal of Water and Climate Change* **12** (7), 3373–3387.
- Shahriar, S. A., Siddique, M. A. M. & Rahman, S. M. A. 2021 Climate change projection using statistical downscaling model over Chittagong Division, Bangladesh. *Meteorology and Atmospheric Physics* **133**, 1409–1427.
- Shukla, R., Khare, D. & Deo, R. 2015 Statistical downscaling of climate change scenarios of rainfall and temperature over Indira Sagar canal command area in Madhya Pradesh, India. In: *IEEE 14th International Conference on Machine Learning and Applications*, IEEE, Piscataway, NJ, USA, pp. 313–317.
- Singh, D., Jain, S. K. & Gupta, R. D. 2015 Statistical downscaling and projection of future temperature and precipitation change in middle catchment of Sutlej River Basin, India. *Journal of Earth System Science* **124** (4), 843–860.
- Suo, M. Q., Zhang, J., Zhou, Q. & Li, Y. P. 2019 Applicability analysis of SDSM technology to climate simulation in Xingtai city, China. *Earth and Environmental Science* **223**, 012053.
- Tabari, H., Paz, S. M., Buekenhout, D. & Willems, P. 2021 Comparison of statistical downscaling methods for climate change impact analysis on precipitation-driven drought. *Hydrology and Earth System Sciences* **25**, 3493–3517.
- Tahir, T., Hashim, A. M. & Yusof, K. W. 2018 Statistical downscaling of rainfall under transitional climate in Limbang River basin by using SDSM. *Earth and Environmental Science* **140**, 012037.
- Teegavarapu, R. S. V. & Goly, A. 2018 Optimal selection of predictor variables in statistical downscaling models of precipitation. *Water Resources Management* **32**, 1969–1992.
- Tootle, G. A., Singh, A. K., Piechota, T. C. & Farnhan, I. 2007 Long lead time forecasting of US streamflow using partial least squares regression. *Journal of Hydrologic Engineering* **12**, 442–451.
- Wilby, R. L. & Dawson, C. W. 2007 *SDSM 4.2 – A Decision Support Tool for the Assessment of Regional Climate Change Impacts, Version 4.2 User Manual*. Lancaster University, Lancaster and Environment Agency of England and Wales, UK.
- Wilby, R. L., Dawson, C. W. & Barrow, E. M. 2002 *SDSM – a decision support tool for the assessment of regional climate change impacts. Environmental Modelling and Software* **17**, 147–159.
- Wilks, D. S. 1995 *Statistical Methods in the Atmospheric Sciences*, 1st edn. Academic Press, San Diego, CA, USA.
- Yang, D. & Saenko, O. A. 2012 Ocean heat transport and its projected change in CanESM2. *Journal of Climate* **25**, 8148–8163.
- Zhang, R., Cheng, L., Liu, P., Huang, K., Gong, Y., Qin, S. & Liu, D. 2021 Effect of GCM credibility on resource system robustness under climate change based on decision scaling. *Advances in Water Resources* **158**, 104063.

First received 14 December 2022; accepted in revised form 14 July 2023. Available online 31 July 2023