Principal Component Regression Model of Extreme Daily Rainfall for Climate-Related Impact Studies

Chavalit Chaleeraktrakoon and Ukrit Khwanket
Faculty of Engineering, Thammasat University, Pathum Thani 12120.

Correspondence:
Chavalit Chaleeraktrakoon
Department of Civil Engineering,
Faculty of Engineering, Thammasat
University, Klong Luang,
Pathum Thani 12120
E-mail: cchava@engr.tu.ac.th

Abstract
This paper proposes a statistical downscaling model to directly describe the linkage between large-scale climate variables and annual maximum daily rainfalls (AMDR) at a local site. The proposed downscaling model is based on principal component analysis (PCA) of global climate variables using singular value decomposition (SVD). The SVD technique is robust and efficient for calculating the standardized principal components (SPC) of the climate variables. The model has been tested with two popular general circulation models, GCMs, (HadCM3 and CGCM3) and available 41-year (1961 - 2001) AMDR data at 6 sites in the Chi and Mul River Basins (Thailand). The PCA of all possible climate predictors has shown that surface divergence is the most important one for HadCM3 while the airflow movement, humidity, and pressure variables of CGCM3 are equally significant. Further, the tested results have indicated that the proposed method can links the simulated climate predictors given by the GCMs with the local AMDR indices because it adequately reproduces the observed frequencies of the AMDRs in calibration and validation periods. In addition, the scenarios (HadCM3A2, HadCM3B2, and CGCM3A2) of the 50-yr AMDRs have demonstrated that the largest values of AMDRs of most stations are approximately unchanged.

Keywords: Climate variation and change, statistical downscaling, principal component analysis (PCA), singular value decomposition (SVD)

1. Introduction

Annual maximum daily rainfall (AMDR) at a single site is basic hydrologic information available for the analysis of flood risk. The larger the amount of the AMDR, the higher the level of the flood risk. Moreover, extreme precipitation is also fundamental to the design of highway and urban drainage structures, especially when the observed maximum rainfall at critical duration is unavailable. In this case, the AMDR of design return period is hence transformed into that at the desired duration using a simple-or multi-scaling concept [1-2].

Climate variation and change usually have important impact on the local AMDR. Downscaling technique is necessary to be developed for assessing the climate impact for the referred engineering applications because the resolution (200 - 300 km) of the general circulation model (GCM) which, is accepted in practice is too coarse. Several downscaling techniques were thus proposed for the impact assessments [3-4]. They generally link the basic climate variables of the GCM at global scale, usually referred to as NCEP variables, to a station's daily precipitation, which is subsequently extracted for the considered AMDR.

The existing downscaling techniques can be categorized into dynamic downscaling (DD) and statistical downscaling (SD). In general, the DD technique models the physical dynamics of atmospheric variables at a grid spacing of 20 - 50 km [5-7]. However, the explicit solutions of the atmospheric dynamics cause the DD technique to be costly. Moreover, its domain size, number of experiments, and duration of simulation are often subjective, and usually lead to poor results. For the SD technique, it is mainly based on the relationship or transfer function between the NCEP climate predictors and local predictand (e.g., daily precipitation) [8-11]. The SD technique is advantageous over DD in that (1) it is flexible for problem adaptation, (2) its computational resources are inexpensive, and (3) it is possible for the analysis of uncertainties and risks [12].

Alternatively, this paper proposes an SD technique that directly describes local AMDR using NCEP variables. The proposed SD technique applies principal component analysis (PCA) to reduce the large number of the global climate variables into their fewer significantly standardized principal components (SSPC). Note that this PCA application is different from a recent work by [4] in that our PCA is used for relating to the station AMDR while the PCA of the previous work is applied to the occurrence and magnitude of daily rainfall. In this study, the SSPC are calculated based on the technique of singular value decomposition (SVD) because it is computationally efficient and robust [13-14]. The linkages between the SSPC predictors and observed at–site AMDRs for 6 stations in the Chi and Mul river basins have shown that their AMDR probability distributions adequately describe the appropriate empirical ones for both calibration and validation periods. In addition, the projections of the 50–yr AMDRs under HadCM3A2, HadCM3B2, and CGCM3A2 scenarios have been performed for several future periods using the proposed model.

2. Statistical Downscaling Technique

Let $X=[x_{ij}]$ be the $m \times n$ matrix of the average annual NCEP data x_{ij} (m = the considered period in years and n = the number of the NCEP variables), $\tilde{X}$ be its standardized matrix of X with zero-means and unit-variances [i.e., $\tilde{X} = (X-\bar{X})S^{-1}$ in which $\bar{X}$is the $m \times n$ matrix of mean values and $\boldsymbol{S}$ is the $n \times n$ diagonal matrix of standard deviations of X], and $\boldsymbol{P}$ be the $m \times r$ (r = the rank of $\tilde{X}$) matrix of standardized principal component (SPC). The annual NCEP data is adopted in this study because this NCEP data base is clearly defined, modeling of AMDR occurrence is not necessary for scenario construction, and their predictive ability is comparable to the others (average NCEP over rainy season and NCEP on the date of AMDR occurrence), as shown in [15]. The SPC matrix $\boldsymbol{P}$ is orthogonal by columns. That is, the $r \times r$ product matrix of $1/(m-1)]\ \boldsymbol{P}^T\boldsymbol{P}$ where T = transpose operator, is equal to the identity matrix $\boldsymbol{I}(r)$ of order r.

The SPC matrix $\boldsymbol{P}$ is best computed by inverting the SVD technique as [16-18]

$$\boldsymbol{P} = \tilde{X}\boldsymbol{E}\boldsymbol{D}^{-1/2} \tag{1}$$

in which D is the $r \times r$ diagonal matrix of positive eigenvalues arranged in descending order ($d_{11} > d_{22} > d_{33} \ldots > d_{rr}$ where d_{jj} = element of D); and E is the $n \times r$ matrix of associated eigenvectors with orthonomal columns [$E^T E = I\ (r)$]and rows [$EE^T = I\ (m)$]. Notice in Eq. (1) that the SPC P is the weighted sum of every standardized NCEP variable where the weights are the product of $\underline{E}$ and $D^{-1/2}$. Moreover, all SPCs Pj (j = 1, 2, …, r) are arranged in descending order of significance based on the amount of data variances D that they take into account. That is, P_1 is the most significant; P_2 is the second most, and so on because $d_{11} > d_{22} > d_{33} \ldots > d_{rr}$. Note also that the product of $ED^{-½}$ is the correlation matrix Q between the SPC matrix P and the NCEP one $\tilde{X}$. The maximum q_{ij} of all correlations in the column vector Q_j indicates that the SSPC P_j is actually the standardized NCEP variable $\tilde{X}_i$.

If the matrix $\tilde{X}$ is nonsingular ($r \leq n$), the eigenvalue matrix D and corresponding eigenvector one E will be estimated from a characteristic equation as

$$RE = ED \tag{2}$$

where R is the $n \times n$ correlation matrix expressed by

$$R = \frac{1}{m\text{-}1}\tilde{X}^T\,\tilde{X} \tag{3}$$

otherwise ($r \leq m$). Calculate first the eigenvalue matrix D and the $m \times r$ associated eigenvectors $\acute{E}$ of $\acute{R}$, in which the $m \times m$ row product matrix $\acute{R}$=[1/(m-1)] $\tilde{X}\ \tilde{X}^T$, using a characteristic equation ($\acute{R}\acute{E} = \acute{G}\,D$). The eigenvector matrix G of R is then computed as

$$E = \frac{1}{\sqrt{m\text{-}1}}\tilde{X}^T\acute{E}\,D^{-(½)} \tag{4}$$

To find out the SSPC sub–matrix P^* of P, the cut–off level is to be specified for partitioning the first k column vectors of P as P^*. The remaining ones are considered as noises $\underline{P}$. In general, cumulative variances explained about 70 – 90% of the total variances and the eigenvalue $d_{jj} \geq 1$ are used as empirical cut–off rules [19]. Among these criteria, the lower limit of the explained variances is adopted in this study because its description is adequately accurate in AMDR downscaling.

These significant predictors P^* are then related to the $m \times 1$ column vector Z of normalized AMDR where

$$Z = \text{In}\ Y \tag{5}$$

and Y is observed AMDR. The multiple linear regression form is chosen because it is simpler and easier to understand the linkage between the predictors and the predictand, as compared with non–linear downscaling methods [20]. The linear relationship can be written as

$$Z = A + P^*B + \Psi \tag{6}$$

where A and B are the column vectors of regression coefficients of dimensions $m \times 1$ and $k \times 1$ respectively, and Ψ is the $m \times 1$ column vector of residual errors with zero means and σ^2 variances. The coefficients (A and B) and residual variances (σ^2) are estimated by the method of least squares [21].

3. Numerical Applications

Six series of 47-year (1961-2007) observed AMDR in the Chi and Mul river basins (Thailand) and the corresponding NCEP data of GCMs (HadCM3 and CGCM3) during 1961-2001 were considered for illustrating the development of the statistical downscaling technique. The characteristics of NCEP variables and outputs from HadCM3 and CGCM3 can be seen in their characteristics in [12]

Fig. 1. presents the locations of the measured rainfall sequences. In the model development, the NCEP and the maximum rainfall data of both 1961-1975 and 1991-2001 periods were used for model calibration. Note that the total length of the 26-year data was different from that (15 years, 1961-1975) of most climate change impact studies because we would like to have the AMDR series as long as possible. The remaining data of between the time intervals were later applied for model validation (1976-1990).

Table 1. shows the significant NCEP variables of HadCM3 for every considered station during the calibration periods. The table demonstrates that all NCEP predictors relate to the magnitude and direction of various winds. Among these climate variables, surface divergence is found to be the most important. However, the NCEP predictors of CGCM3 are quite different because they are a combination of the airflow movement, humidity, and pressure variables (see Table 2). Notice that the CGCM3 predictors are equally important in modeling the AMDR under climate variation and change for the study area. Also, note that the SSPC between the HadCM3 and CGCM3 will be different because their grid sizes are not the same (HadCM3 2.5° × 3.75° and CGCM3 3.75° × 3.75°). Moreover, for each model, its significant predictors are distinguished against its grid location (e.g., Khon Kaen and Roi Et).

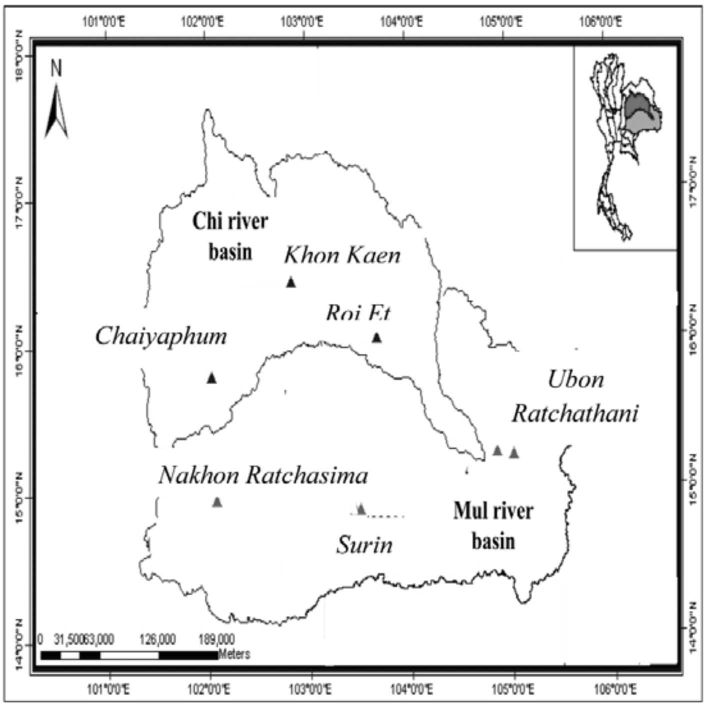

Fig. 1. Locations of selected rainfall gauging stations in the Chi and Mul River Basins.

Table 1. Significant NCEP variables of HadCM3 for every considered station during 1961-1975 and 1991-2001.

Station	NCEP Predictors
Khon Kaen	surface wind direction, surface divergence, 500 hPa vorticity
Chaiyaphum, Nakhon Ratchasima	surface zonal velocity, surface meridional velocity, 500 hPa airflow strength
Roi Et, Ubon Ratchathani, Surin	surface divergence, 500 hPa zonal velocity, 850 hPa wind direction

Table 2. Significant NCEP variables of CGCM3 for every considered station during 1961−1975 and 1991−2001.

Station	NCEP Predictors
Khon Kaen, Chaiyaphum, Nakhon Ratchasima	mean sea level pressure, 850 hPa wind direction, specific humidity at 850 hPa
Roi Et, Ubon Ratchathani, Surin	surface vorticity, surface divergence, specific humidity at 500 hPa

The AMDR results of the calibrated PCA technique presented in the form of probability plots were then compared with the corresponding observed ones. The downscaling technique would be inferred to adequately describe the historical maximum rainfalls if the assessment intervals are between 2.5−97.5 percentiles of the modeled AMDRs, containing the considered precipitation. The inference by assessment interval is more meaningful than that by point evaluation (e.g., mean or median). The model accuracy will be accepted if the bounds contain the observed AMDR. Fig. 2 displays the historical AMDRs and the downscaled ones of HadCM3 and CGCM3 for various return periods at the study sites. The figure indicates that the proposed downscaling technique reproduces the observed AMDRs sufficiently. The downscaling technique has been also demonstrated its capability of satisfactorily preserving the local AMDR at every site in the validation period (see Fig. 3.).

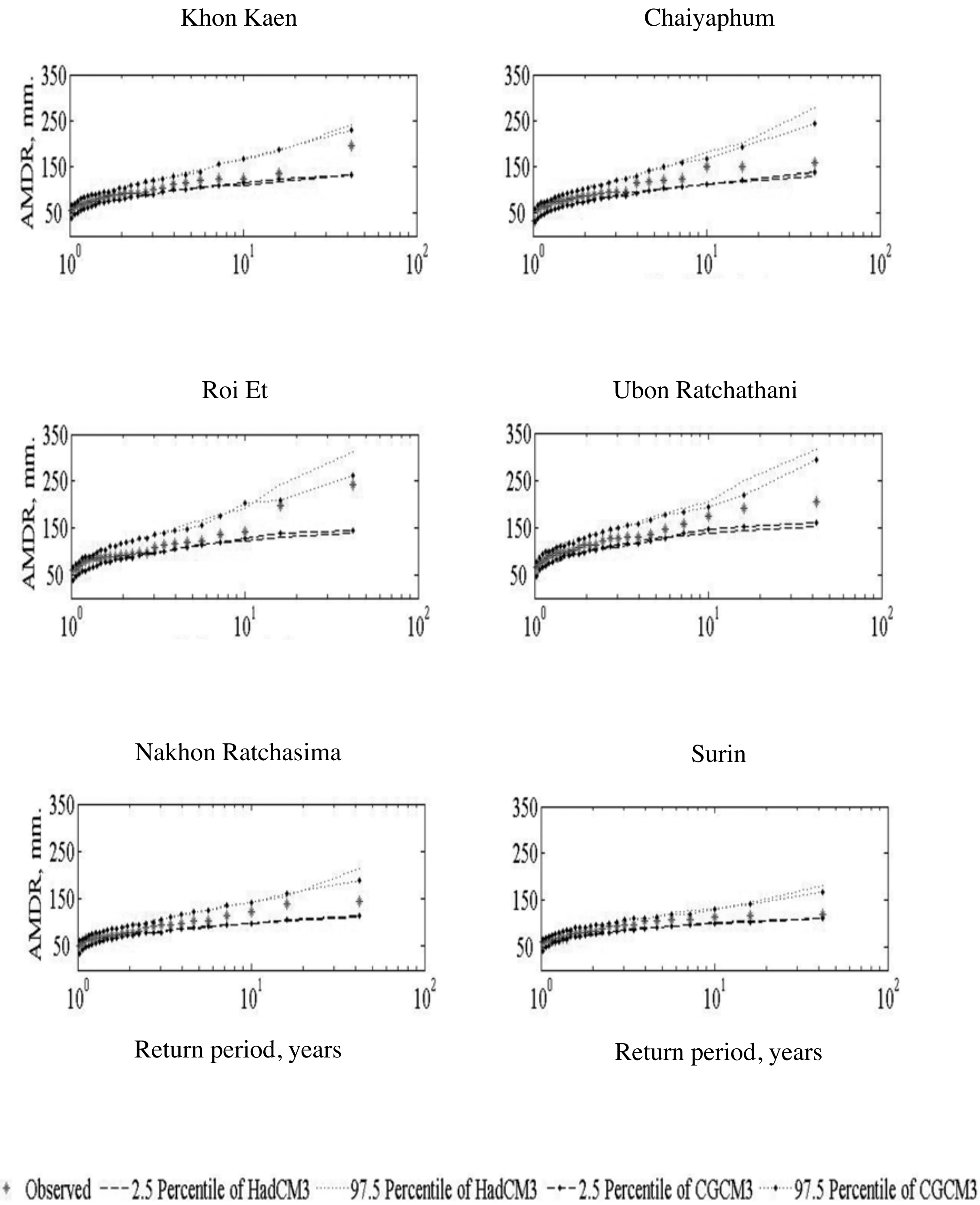

Fig. 2. The probability plots of AMDR and observed ones for the calibration periods (1961-1975 and 1991-2001).

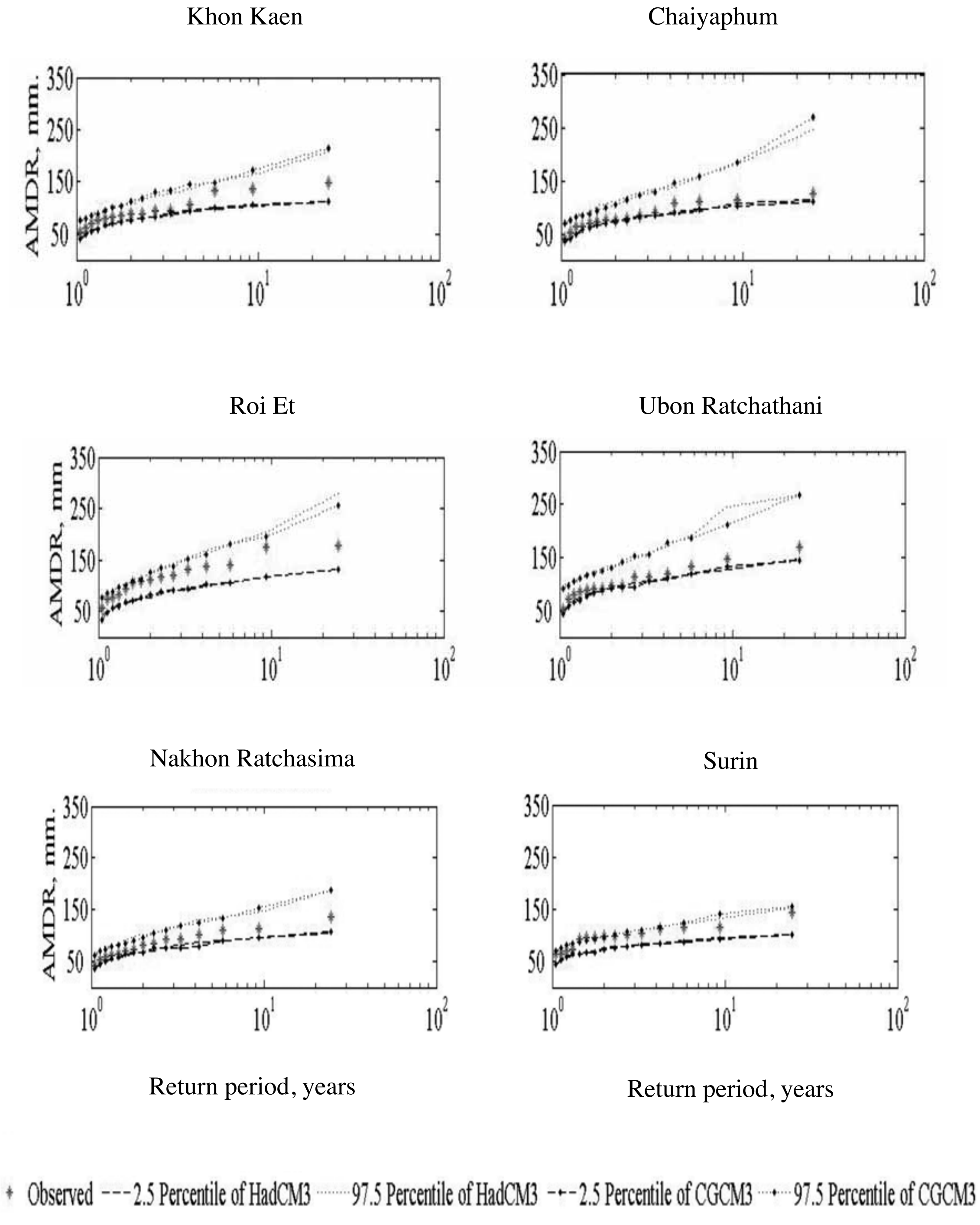

Fig. 3. The probability plots of AMDR and observed ones for the validation periods (1976-1990).

Based on the calibration and validation results, the technique has been thus applied to construct the AMDR scenarios (i.e., HadCM3A2, HadCM3B2, and CGCM3A2) for the base period (1961 - 1990) and future intervals (2020s, 2050s, and 2080s) at all locations using the climate simulation outputs of the referred scenarios. Note that the CGCM3 does not provide the CGCM3B2 scenario. Table 3-5 presents the AMDR anomalies of the 50-yr return period for the 2020s, 2050s, and 2080s intervals (i.e., the largest value in such periods), as compared with the base one. The table indicates that the HadCM3A2, HadCM3B2, and CGCM3A2 scenarios of the maximum rainfall index in the 2020s, 2050s, and 2080s for most stations are approximately unchanged. The range of the AMDR anomalies is usually less than 5 mm. This implies that floods in the future remain the same, provided that land uses and drainage structures are unchanged.

Table 3. AMDR anomalies of HadCM3A2, HadCM3B2, and CGCM3A2 for current period (2020s) compared with base period (1961-1990).

station	AMDR in the 2020s at 50-yr (mm.)		
	HadCM3A2	HadCM3B2	CGCM3A2
Khon Kaen	3.6	6.8	5.7
Chaiyaphum	-7.4	-2.4	-1.5
Roi Et	4.9	-2.9	-1.8
Ubon Ratchathani	3.2	0.3	-0.6
Nakhon Ratchasima	6.0	-5.9	-2.9
Surin	-3.3	-3.4	-1.5

Table 4. AMDR anomalies of HadCM3A2, HadCM3B2, and CGCM3A2 for current period (2050s) compared with base period (1961-1990).

station	AMDR in the 2050s at 50-yr (mm.)		
	HadCM3A2	HadCM3B2	CGCM3A2
Khon Kaen	1.9	5.4	3.2
Chaiyaphum	-5.5	-3.7	2.1
Roi Et	-5.3	7.9	2.3
Ubon Ratchathani	3.2	-4.8	3.6
Nakhon Ratchasima	14.0	-1.7	-1.6
Surin	-0.8	0.3	1.2

Table 5. AMDR anomalies of HadCM3A2, HadCM3B2, and CGCM3A2 for current period (2080s) compared with base period (1961-1990).

station	AMDR in the 2080s at 50-yr (mm.)		
	HadCM3A2	HadCM3B2	CGCM3A2
Khon Kaen	3.7	-0.2	2.0
Chaiyaphum	4.5	-3.2	0.1
Roi Et	-2.7	1.5	-0.7
Ubon Ratchathani	4.8	-1.2	-1.3
Nakhon Ratchasima	9.8	-8.5	-3.1
Surin	2.8	-1.9	-1.4

4. Conclusions

A statistical downscaling technique is proposed in the present study to describe the linkage between NCEP variables and AMDR indices at a local site. The proposal is based mainly on PCA of average annual NCEP data. A combination of the robust and efficient technique of SVD and the 70% of cumulative variances of the NCEP data are used to determine their SSPCs, which are further applied as the significant predictors for the AMDR at the appropriate site.

The model has been developed for relating the SSPCs of the NCEPs (HadCM3 and CGCM3) to the AMDR indices at 6 stations in the Chi and Mul River Basins. The total length of available AMDR data for the chosen sites is 47 years (1961-2007). The model developments use the AMDR data during 1961-1975 and 1991-2001 for calibration. The intermediate data (i.e., 1976-1990) are applied for validation. The screening of all NCEP variables for HadCM3 shows that surface divergence is the most important predictor among all significant wind magnitudes and directions. The airflow movement, humidity, and pressure variables of CGCM3 are equally significant. The calibration and validation results have indicated that the models are feasible for downscaling the AMDRs because they adequately describe the frequencies of observed AMDR at all sites. Moreover, this technique had been shown to be more accurate than the existing method (i.e., AMDR extraction in downscaled daily rainfalls) in downscaling AMDR [22]

Consequently, the downscaling approaches have been thus used to construct the HadCM3A2, HadCM3B2, and CGCM3A2 scenarios of the 50–yr AMDRs for the base period (1961-1990) and future intervals (2020s, 2050, and 2080s). The developed SD technique of AMDR at site was investigated on the ability to project its AMDR scenarios. The HadCM3A2, HadCM3B2, and CGCM3A2 scenarios of the proposed technique in 2020s, 2050s, and 2080s have demonstrated that the largest values of AMDRs of most stations are approximately unchanged.

5. Acknowledgement

The authors are thankful to anonymous reviewers for their valuable comments regarding this paper.

6. Reference

[1] Nguyen, V.T.V., Desramaut , N. and Nguyen, T-D., Estimation of Design Storms in Consideration of Climate Variability and Change, 11th International Conference on Urban Drainage, Edinburgh, Scotland, UK, 2008.

[2] Willems, P., Arnbjerg-Nielsen, K., Olsson, J. and Nguyen, V.T.V., Climate change impact assessment on urban rainfall extremes and urban drainage: Methods and shortcomings, Atmospheric Research, Vol. 103, pp. 106-118, 2012.

[3] Pharasi, S., Development of Statistical Downscaling Methods for the daily precipitation process at a location site, 2006.

[4] Artlert, K., Statistical downscaling of precipitation process for climate-related impact studies, Doctoral thesis, Faculty of Engineering, Thammasat University, 2011.

[5] Denis, B., Laprise, R., Caya, D., and Cote, J., Downscaling ability of one-way nested regional climate model: the Big-Brother experiment, Climate Dynamics, Vol. 18, pp. 627-646, 2002.

[6] Nieto, J.D., and Wilby, R.L., A comparison of statistical downscaling and climate change factor methods: impacts on low flows in the river Thames, United Kingdom, Climate Change, Vol. 69, pp. 245-268, 2005.

[7] Sanchez, E., Gaertner, M.A. and Gallardo, C., Dynamical downscaling of daily precipitation over the Iberian Peninsula: a spatial resolution analysis for present and future climate conditions, Física de la Tierra, Vol. 21, pp. 207-218, 2009.

[8] Wilby, R.L., Charles, S.P., Zorita, E., Timbal, B., Whetton, P., and Mearns, L.O., Guidelines for Use of Climate Scenarios Developed from Statistical Downscaling Methods, 2004.

[9] Choux, M., Development of new climate predictor variables for the statistical downscaling of the daily precipitation process, Master's thesis, Development of civil engineering and applied mechanics, McGill University, 2005.

[10] Wetterhall, F., Bárdossy, A., Chen, D., Halldin, S. and Xu, C., Statistical downscaling of daily precipitation over Sweden using GCM output, Theor Appl Climatol, Vol. 96, pp. 95-103, 2009.

[11] Artleart, K. Chaleeraktrakoon, C. and Nguyen, V.T.V., Modeling and analysis of rainfall processes in the context of climate change for Mekong Chi and Mun River Basins (Thailand), Journal of Hydro-environment Research, Vol. 7, pp. 2-17, 2013.

[12] Wilby, R.L., and Dowson, C.W., SDSM 4.2-A decision support tool for the assessment of regional climate change impacts, User manual, 2007.

[13] Morishima, W., and Mikami, T., Large-scale variations in summer rainfall over South and East Asia and their relations to anomalies of the sea level pressure field in the Tropics, Japanese progress in climatology, pp. 11-16, 1998.

[14] Nam, W., Shin, J.Y., and Heo, J.H., Selection of variables for regional frequency analysis of annual maximum precipitation using multivariate techniques, 2007.

[15] Khwanket, U., Statistical downscaling of Extreme daily rainfalls for climate-related impact studies, Master's degree thesis, Faculty of Engineering, Thammasat University, 2013.

[16] Green, P.E., Analyzing multivariable data, Dryden Press, Hinsdale, III, 1978.

[17] Cavadias, G. S., A multivariate seasonal model for streamflow simulation, La recherché en Hydrologie au Quebec, 1986.

[18] Chaleeraktrakoon, C., Stochastic procedure for generating seasonal flow, J. Hydrol, Eng., Vol. 4(4), pp. 337-347, 1999.

[19] Chaleeraktrakoon, C., Parasimonious SVD/MAR(1) Procedure for Generating Multisite Multiseason Flows, J. Hydrol, Eng., pp. 517-518, 2009.

[20] Wichert, A., and Burge, G., Linear versus nonlinear techniques in downscaling, Climate Research, Vol. 10, pp. 83-93, 1998.

[21] Haan, C.T., Statistical Methods in Hydrology, Iowa State University Press, 1997.

[22] Chaleeraktrakoon, C., and Khwanket, U., A statistical downscaling model for extreme daily rainfalls at a single site, Proceedings of the 2013 World Environmental and Water Resources Congress Cincinnati, OH, 10 pages 2013.