Theor Appl Climatol
DOI 10.1007/s00704-015-1580-1

ORIGINAL PAPER

Statistical downscaling rainfall using artificial neural network: significantly wetter Bangkok?

Minh Tue Vu[1] · Thannob Aribarg[2,3] · Siriporn Supratid[3] · Srivatsan V Raghavan[1] · Shie-Yui Liong[1,4]

Received: 1 April 2014 / Accepted: 20 July 2015

Abstract Artificial neural network (ANN) is an established technique with a flexible mathematical structure that is capable of identifying complex nonlinear relationships between input and output data. The present study utilizes ANN as a method of statistically downscaling global climate models (GCMs) during the rainy season at meteorological site locations in Bangkok, Thailand. The study illustrates the applications of the feed forward back propagation using large-scale predictor variables derived from both the ERA-Interim reanalyses data and present day/future GCM data. The predictors are first selected over different grid boxes surrounding Bangkok region and then screened by using principal component analysis (PCA) to filter the best correlated predictors for ANN training. The reanalyses downscaled results of the present day climate show good agreement against station precipitation with a correlation coefficient of 0.8 and a Nash-Sutcliffe efficiency of 0.65. The final downscaled results for four GCMs show an increasing trend of precipitation for rainy season over Bangkok by the end of the twenty-first century. The extreme values of precipitation determined using statistical indices show strong increases of wetness. These findings will be useful for policy makers in pondering adaptation measures due to flooding such as whether the current drainage network system is sufficient to meet the changing climate and to plan for a range of related adaptation/mitigation measures.

✉ Minh Tue Vu
tue@nus.edu.sg

[1] Tropical Marine Science Institute, National University of Singapore, 18 Kent Ridge Road, Singapore 119227, Singapore

[2] Climate Change and Disaster Centre, Rangsit University, Rangsit, Thailand

[3] Department of Information Technology, Rangsit University, Rangsit, Thailand

[4] Willis Research Network, Willis Re Inc., London, UK

1 Introduction

There is now a strong evidence that global warming is occurring rapidly. The Intergovernmental Panel on Climate Change (IPCC) has mentioned in its Fifth Assessment Report (AR5), published in 2013, that it is certain that global mean surface temperature has increased since the late nineteenth century and the decade of the 2000s has been the warmest. It is also reported that it is virtually certain that maximum and minimum temperatures over land have increased on a global scale since 1950. The IPCC reports that there is also a medium confidence in a significant human influence on global scale changes in precipitation patterns, including increases in northern hemisphere mid to high latitudes. It is likely that human influence has affected the global water cycle since 1960. It is reported that future increases in precipitation extremes related to the monsoon is very likely in South America, Africa, East Asia, South Asia, Australia and Southeast Asia. Temperatures may see increases by 3 °C and rainfall increases up to 40 % by the end of the century, over Southeast Asia (IPCC 2013). Overall, many of the indicated climatic changes and projections have been found to be consistent with its earlier report, AR4.

This emphasizes the severity of climate change and its impacts and the strong involvement of human-induced changes to that effect. Because of the coarse resolutions of the global climate models (GCMs), there exist limitations in quantifying climate and its changes over regions or specific locations that are many times smaller than what is resolved by large scale models. The usual resolution of GCMs range from 150 to

Published online: 09 August 2015

400 km whilst the subregional/local information is required at much smaller scales, may be 10–50 km or even to a station location over one grid point. In a need for such local scale information, we apply an empirical downscaling to study future climate changes during the rainy season over specified locations in Bangkok, Thailand, a city very vulnerable to climate change (Bigio 2003; Senga (2009). With the availability of full observational records for station data in Bangkok, we apply a statistical downscaling method to downscale future climate under different scenarios. In this study, we analyse three of the GCMs under the IPCC's A1B scenario of the Coupled Model Intercomparison Project 3 (CMIP3) and one GCM under the Representative Concentration Pathways (RCP) 4.5 (van Vuuren et al. 2011), to assess future changes in rainfall over Bangkok.

The main assumptions for statistical downscaling are that (i) high quality large scale and local data will be available for a sufficiently long period to establish robust relationships of the current climate and (ii) relationships which are derived from recent climate will be relevant in a future climate. Many papers have dealt with statistical downscaling concepts, their prospects and their limitations (von Storch 1995; Hewitson and Crane 1996; Wilby and Wigley 1997; Zorita and von Storch 1997; Gyalistras et al. 1994; Murphy 1999, 2000; Widmann and Bretherton 2000). The advantages of using this technique are that they are computationally inexpensive and can easily be applied to analyse the output data from different GCM experiments. The applications of this downscaling technique vary widely with respect to regions, spatial and temporal scales, type of predictors (those climate variables which are used to predict) and predictands (those climate variables which are predicted) and their climate statistics (Jones et al. 2004). There are three main categories for statistical downscaling are (1) transfer function (Wilby et al. 1998, 2002), (2) weather typing (von Storch et al. 1993) and (3) stochastic weather generator (Kilsby et al. 2007). The most typical transfer function approach with linear relationship between predictors and predictand that has been widely used is by the Statistical Downscaling Model (SDSM) (Wilby et al. 2002). However, SDSM is good for downscaling temperature fields but not good for precipitation (Chen et al. 2012b).

Some recent studies have used the artificial neural network (ANN) to statistically downscale climate variables. Fistikoglu and Okkan (2011) and Okkan and Fistikoglu (2013) applied Levenberge-Marquardt algorithm-based feed forward neural networks (LM-FFNN) to downscale GCM CGCM3 for present day and future climates for precipitation and temperature over the Thatali River basin in Turkey. After downscaling, bias correction was applied to these variables and evaluated at watershed scale by means of a parametric hydrological model GR2M to observe the probable impacts of temperature and precipitation changes on runoff for near-future periods. Their findings indicated that the near-future runoff was likely to decrease by 32 % leading to a reduction in water supply. Mendes and Marengo (2010) compared ANN with an autocorrelation statistical downscaling model to downscale five coupled GCMs for a region in the Amazon Basin. Their results indicated that the ANN model significantly outperformed the other statistical models for downscaling of daily precipitation. Hu et al. (2013) compared three statistical downscaling methods: using SDSM, Generalized Linear model for daily Climate and Nonhomogeneous Hidden Markov Model for downscaling summer precipitation at a network of 14 stations over the Yellow river region using the National Centre for Environmental Prediction (NCEP) reanalysis data and GCMs CGCM3 and ECHAM5. The paper concluded that, under the future climate scenario, all areas of the study region are likely to experience increases in both rainfall totals and extremes. Their results also highlighted that climate projections based on results from only one downscaling method should be interpreted with caution. Similar studies for other regions in China were also done by Liu et al. (2013), Chen et al. (2012a), Hu et al. (2012) and Liu et al. (2011).

There have been some but not many studies on climate change over Thailand. Sharma and Babel (2013) applied the bias correction and disaggregation techniques to downscale precipitation from GCM ECHAM4/OPYC for three future climate 5-year periods. The downscaled precipitation was used as input to the rainfall runoff model HEC-HMS for the upper Ping river basin of Thailand. Their result suggested a decrease of 13–19 % in the annual stream flow over the future compared to the baseline period and a shift in the seasonal stream flow with peak flow occurring in October–November rather than the usual September as observed in the baseline period. Kang et al. (2007) applied the correlation analysis and singular value decomposition analysis (SVDA) to downscale precipitation for Bangkok region. Sea level pressure (SLP) was chosen as the sole predictor from the APEC Climate Center Multi-model Ensemble because the SLP in the western North Pacific had a stronger correlation with rainfall than other climate predictor variables. Their findings indicated that the downscaling can predict variability of precipitation for all the stations because the parent model was good in simulating the SLP patterns.

In such a context, we study Bangkok rainfall during the rainy season to assess whether Bangkok will experience more intense rainfall under a changing climate, using the Artificial Neural Network technique.

2 Methodology: study region, model and data

In this study, we apply a nonlinear relationship using artificial neural network (ANN) (Priddy and Keller 2005) with input data from the ERA-Interim reanalysis data for training/testing. We use three GCMs under the Special Report on Emission Scenario (SRES) A1B of the IPCC and one GCM under the latest RCP 4.5 scenario, to statistically downscale the precipitation over Bangkok area for present (1980–2000) and future climates (2080–2100). The details of these GCMs are provided in Section 2.2.3.

2.1 Study area

Bangkok is the capital and commercial city of Thailand and is one of the highly developed cities in Southeast Asia and as well largely vulnerable to climate change (ADB 2009). The region has a tropical wet and dry type of climate, Koppen type Aw (Peel et al. 2007), with long durations of sunshine, high surface temperatures and high humidity. Thailand experiences a tropical climate and depends on wind field characteristics, particularly on the location and intensity of the intertropical confluence. The Inter-tropical Convergence Zone (ITCZ) refers to the zone where trade winds of the two hemispheres converge. Seasonal movement of the ITCZ induces large-scale reversals of the wind regime. Two large-scale reversals in air motion influence the climate of Thailand: the northeast and southwest monsoon (Food and Agricultural Organization (FAO) 1982). The northeast monsoon occurs from November to February when the ITCZ moves southward. This type of wind brings cool and dry air from the Siberian anticyclone down to the country and creates a dry season for Thailand. The southwest monsoon during summer gathers humid air from the Indian Ocean when ITCZ moves northward. It first creates a heavy downpour along the west coast of Thailand and then spreads over the entire country, forming the rainy season from early May to the end of October (Bachelet et al. 1992; Boochabun et al. 2004; Hung et al. 2009). By analysis of the rainfall time series of Bangkok, the total annual rainfall is about 1600 mm. By seasons, the total rainfall during the rainy season (May–Oct) covers more than 85 % of the annual rainfall whilst the dry season is only about 15 %. This implies that the rainy season strongly influences the total annual rainfall. Therefore, the study focuses on the future climate change over Bangkok area only during the rainy season.

2.2 Data

2.2.1 Station data

The observed daily precipitation data, for the period 1980–2000, was obtained from the Thailand Meteorological Department (TMD) for the Bangkok station, located in central of Bangkok, at coordinates, latitude 13.73° N and longitude 100.56° E and depicted in Fig. 1.

2.2.2 Reanalysis data

The predictors' dataset, which comprises large-scale climate variables, was obtained from the reanalysis data for training/testing process. The reanalysis data used was the European Re-Analysis Interim dataset (Dee et al. 2011), referred to in this study as 'ERAI'. The data have a spatial resolution of 1.5° longitude × latitude. Twenty-five ERAI grid points surrounding Bangkok area have been chosen with all the predictors, which consisted of mean sea level pressure, relative/specific humidity, zonal/meridional wind and temperature, and are displayed in Fig. 1.

2.2.3 Global climate model data

Global climate data were obtained from three GCMs: ECHAM5 (1.875° × 1.875°), from the Max Planck Institute (MPI), Germany; CCSM3.0 (1.4° × 1.4°), from the National Centre for Atmospheric Research (NCAR), USA and MIROC-medium resolution (medres) (2.8° × 2.8°) from the Center for Climate System Research, Japan, all under the CMIP3 scenario A1B. One more GCM data was obtained from the atmospheric model component of the Max Planck Institute's Earth System Model medium resolution (MPI-ESM-MR), whose future scenario is under the CMIP5 scenario RCP 4.5. We consider the RCP 4.5 for comparison against A1B scenario, as RCP 4.5 is the next closest scenario of the SRES scenarios next to B1 as sufficient predictor data were not available from B1 scenarios for many GCMs. A present-day climate period 1980–2000 and a future period 2080–2100 were considered in the study. However, not all predictors that were available from the reanalysis data were available from the GCMs. Such limitations have also been mentioned by Brands et al. (2011). Hence, there was a need to make a very careful screening of the predictor variables, which is described in detail in Section 3. The choice of downscaling as many GCMs and scenarios as possible arose from a suggestion by Wilby and Dawson (2004): 'Users should not restrict themselves to the use of a single GCM or emission scenario for downscaling. By applying multiple forcing scenarios (via different GCMs, ensemble member, time-slices or emission pathways) better insight may be gained into the magnitude of these uncertainties'.

Details on the predictors available from ERAI and other GCM datasets are tabulated in Table 1.

2.3 Artificial neural network and principal component analysis

2.3.1 Artificial neural network

The ANN model used in this study is a feed forward network with three layers: an input layer, a hidden layer and an output layer where each layer is connected by weights. This three-layered feed forward network is commonly used due to its general applicability to a variety of different problems (Hsu et al. 1995). The learning procedure tries to find a set of weights which gives a mapping that fits well with the input

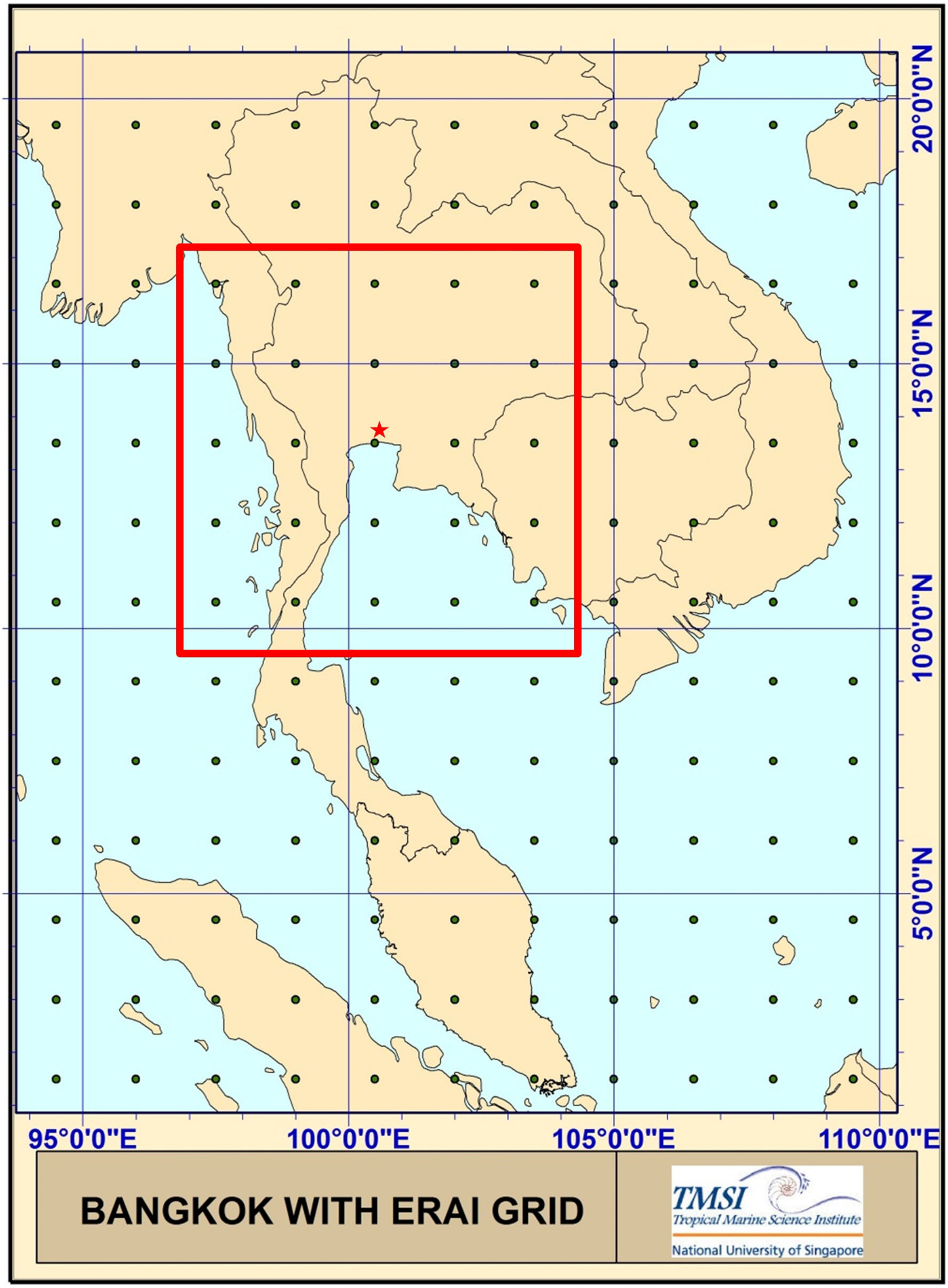


Fig. 1 Distribution of ERA-Interim grid boxes (*dots* indicate 1.5° × 1.5°) and the location of 25 grid boxes surrounding Bangkok (*red star*)

and output. The following function as an arbitrary nonlinear function could be assumed:

$$F(x, w) = y \tag{1}$$

where x is the input vector (predictors) presented to network, w is the weight of the network and y is the corresponding output (predictand) which is approximated or predicted by the network.

Table 1 Daily predictors available for reanalysis and GCM data, Bangkok location

Daily Variables	Code	ERAI	ECHAM5	CCSM3	MIROC-m	MPI-ESM
Mean sea level pressure	mslp	x	x	x	x	x
500 hPa geopotential height	p500	x	x	x	x	x
850 hPa geopotential height	p850	x	x	x	x	x
Near surface zonal velocity	p__u	x	x	x	x	x
Zonal velocity at 500 hPa height	p5_u	x	x	x	x	x
Zonal velocity at 850 hPa height	p8_u	x	x	x	x	x
Near surface meridional velocity	p__v	x	x	x	x	x
meridional velocity at 500 hPa height	p5_v	x	x	x	x	x
meridional velocity at 850 hPa height	p8_v	x	x	x	x	x
Near surface relative humidity	rhum	x	x			x
Relative humidity at 500 hPa height	r500	x	x			x
Relative humidity at 850 hPa height	r850	x	x			x
Near surface specific humidity	shum	x		x	x	x
specific humidity at 500 hPa height	s500	x		x	x	x
specific humidity at 850 hPa height	s850	x		x	x	x
Mean temperature at 2 m	temp	x	x	x	x	x

The standard training algorithm used in this method is the back propagation algorithm. Coulibaly et al. (2000) reviewed the ANN-based modelling in hydrology and reported that 90 % of the experiments extensively make use of the multilayer-feed forward neural network (FNN) trained by the standard backpropagation (BP) algorithm. The full mathematical derivation of this algorithm can be found in several neural networks textbooks (Haykin 1994; Bishop 1995; Martin et al. 1996). Two problems that can arise when using the BP training algorithms are, first, that convergence may be slow and, second, that the final weights may be trapped in local minima over the highly complex error surface (Bishop 1995). Several recent improvements to the back propagation algorithm (variable learning rate and the use of a momentum coefficient) have helped to minimize these problems (Kuligowski and Barros 1998; Tangang et al. 1998). However, an alternative approach has been the introduction of numerically optimized techniques such as the Levenberg-Marquardt (LM) method. This is based on an approximation of the Gauss-Newton's method (Björck 1996). It has the advantage of converging faster and with more reliability than most back propagation schemes (Martin et al. 1996).

In this study, the sigmoid transfer function is applied in the hidden layer and linear transfer functions in the output layers. The number of nodes in the hidden layer is calculated in equation (2)

$$l = \frac{1}{2}(i + o) + \sqrt{N} \tag{2}$$

where l is number of node in the hidden layer, i is the number of input node, o is the number of output node and N is the number of records in the training set.

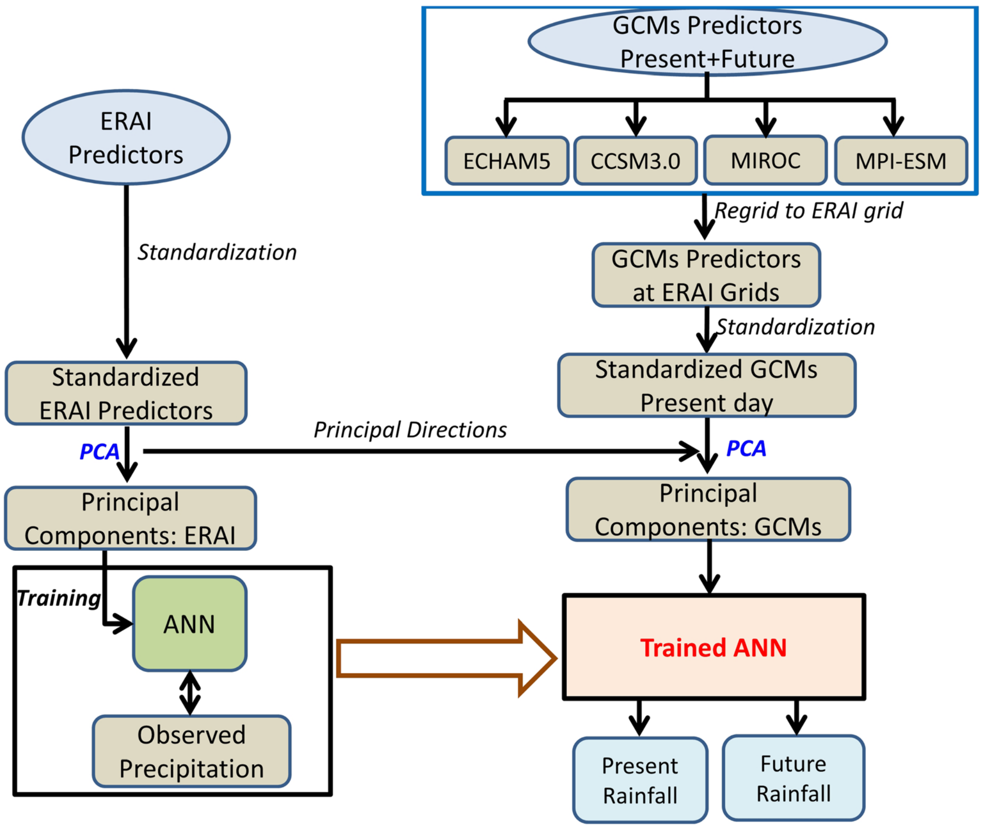
2.3.2 Principal component analysis (PCA)

Principal component analysis, also known as the Karhunen Loeve transform, is one of the most popular techniques for dimensionality reduction (Jolliffe 1986). PCA is applied to transform the set of correlated M-dimensional predictors into another set of M-dimensional uncorrelated vectors called principal components (PCs) by linear combination, such that most of the information content of the original data set is stored in the first few dimensions of the new set. In this study, PCA was performed to reduce the dimensionality of the predictors from 25 grid boxes (Fig. 1). The PCs of the PCA method give 90 % of the information of the original predictors. These new datasets are provided as input to the ANN downscaling model.

2.4 Methodology

In this study, the overall methodology of proposed statistical downscaling process is displayed in Fig. 2. All the GCM predictor grids were interpolated to the same resolution as reanalysis data ERAI, prior to statistical downscaling. Subsequently, the GCM predictors need to pass through standardization (aka normalization) which is a process involving subtracting the mean and divided by the standard deviation of the predictor variable for a predefined baseline period. This is done to remove the systematic biases of the GCM predictors relative to the station rainfall data or reanalysis data (Wilby and Dawson 2004). The standardized predictors were then extracted for 25 grid points surrounding Bangkok which are shown in Fig. 1. There are total 16 predictors (Table 1) for ERAI, 16 for MPI-ESM and 13 each for ECHAM5, CCSM3, and MIROC-medres. Hence, the maximum dimension of

Fig. 2 Schematization of methodology



probable predictors associated with 25 grid points are 25 × 16 as inputs for each ANN model run (for ERAI and MPI-ESM) and 25 × 13 inputs for other three GCMs. However, within 25 grid points for the same predictor (mean sea level pressule (mslp), for example), the predictor value at one grid point is highly correlated with its neighbour. Thus, it might lead to multi-collinearity. Therefore, the PCA is performed to reduce the dimensionality of the predictor variables. The outputs of the statistical downscaling of the GCMs include both the present-day and future datasets for scenarios A1B and RCP4.5.

The development of downscaling models for each of GCM started with the selection of predictors, principal component analysis, then training/testing ERAI predictors using the neural network. The trained ANN model is applied to obtain projections of precipitation from the GCMs mentioned above (see Table 2).

3 Choice of predictors

One of the most important steps in applying an empirical downscaling is to screen for the most appropriate predictor variables, as the downscaling results are strongly sensitive to the choice of the predictors (Hewitson and Crane 1996). For example, when precipitation is the predictand, the humidity in multiple vertical atmospheric levels has a high correlation and hence is sensitive to the choice predictors (Charles et al. 1999). In addition, a large number of predictor sets from GCMs are not available as one could obtain from the reanalysis data. The choice of predictors could vary from region to region depending on the characteristics of large-scale

Table 2 Choices of predictors for ANN training

Code	ECHAM5	CCSM 3	MIROC-m	MPI-ESM
mslp	x	x	x	x
p500				
p850	x	x		x
p__u	x	x	x	x
p5_u				
p8_u	x	x	x	x
p__v	x	x		x
p5_v				
p8_v				
rhum	x			x
r500				x
r850				
shum		x	x	x
s500			x	x
s850				
temp	x		x	x

predictors and predictand as well as the direction of monsoon season. Wilby and Wigley (2000) suggested that an ideal set of predictors should have the following characteristics: (1) physically and conceptually sensible with respect to the predictand, (2) readily available from archives of observed data and GCM output and (3) accurately modelled by GCMs. It was also recommended that the candidate predictor suite contains variables describing atmospheric circulation, thickness, stability and moisture content. Because of this constraint, a simple procedure such as a partial correlation analysis may be used to screen for the most appropriate predictor variables (Wilby et al. 2003). Scatter plots and spatial correlation are often used to select predictors (Dibike and Coulibaly 2005).

In this study, the spatial correlation between rainfall station data with large scale predictors from ERAI reanalyses is described in Section 3.1. Inter-seasonal plots for the southwest monsoon rainfall predictors between ERAI and GCM predictors are provided in Section 3.2. The above steps are useful to verify if the predictors are realistically simulated by the GCMs and if the predictors from ERAI and station data are well correlated. Similar procedures are found in Anandhi et al. (2009) and Ghosh (2010).

3.1 *Spatial correlations between observations and reanalysis*

Identifying empirical relationships between gridded predictors and single site predictands is central to all statistical downscaling methods (Wilby et al. 2002). In this procedure, spatial correlations were calculated in monthly scales for the present-day period using rainfall data from Bangkok station against a set of 16 predictors (Table 1) of the ERAI reanalysis during the rainy season, May, June, July, August, September and October (MJJASO). The correlation analysis was computed using long term monthly data for 20 years from 1980 to 2000 using ERAI and Bangkok rainfall station data. The results of the spatial correlation coefficients are shown in Fig. 3. The spatial correlations are shown as contour lines across the map. The high correlation index is the value closer to 1 (positive correlation) or −1 (negative correlation). From Fig. 3, the ERAI predictors, which have good correlation with southwest monsoon rainfall at Bangkok are a set of 12 predictors: mslp, p__u, p8_u, p5_u, p850, p500, p__v, rhum, r850, shum, s500 and temp. The highest correlation is for zonal wind at 850 and 500 mb (p8_u and p5_u) ($R \sim 0.45$) as the southwest monsoon wind is the main driving force to flow in humid air from the Indian Ocean towards Thailand. The meridional wind on the other hand has negative correlation with Bangkok rainfall with correlation of about 0.35 (p__v, p8_v). The relative humidity at surface and at 850 mb also showed a correlation of $R = 0.4$, next to zonal wind's correlation with rainfall. The predictors zonal wind and humidity at different levels normally showed good correlations against local rainfall as in many other studies by Hu et al. (2013); Mahmood and Babel (2013); Liu et al. (2013) and Duhan and Pandey (2014).

This section has described the choice of predictors based on spatial correlation with local rainfall. In the next section, the inter-seasonal southwest monsoon rainfall predictors are plotted with the Pearson correlation coefficient to select the predictors that agree well between the ERAI reanalysis and GCMs. This analysis allows a determination of which predictors are most strongly correlated, which in turn helps in making a good selection of predictors for downscaling. These results can also assist in choosing a set of quasi-independent predictor variables that will reduce the problem of multi-collinearity in developing regression equations (Wilby and Wigley 2000).

3.2 Inter-seasonal southwest monsoon predictors between reanalysis and GCMs

According to Wilby and Wigley (2000), the chosen predictors for training in the reanalysis variables ought to be similar to those from the GCMs. Fig. 4 displays the inter-seasonal southwest monsoon averaged over a 25-grid box domain for all 16 predictors between ERAI and GCM MPI-ESM, in monthly scales. We used the MPI-ESM for this comparison because a full set of predictors, as shown in Table 1, was available (and not available from the other GCMs). There is no unit in this chart because all the predictors were standardized. The blue colour represents ERAI whilst the red represents MPI-ESM (the Pearson correlation coefficient R was computed for the two datasets of the same predictor and added to the title, next to predictor's name as seen in each subplot of Fig. 4). The best set of predictors, which have good inter-seasonal scale and good correlation coefficient are p__v, p8_u, p__u and temp with $R = 0.86$, 0.8, 0.79 and 0.78, respectively. The worst set of predictors consists of p500, p5_v, s850 and r850 in that order. From Fig. 4, we find the set of another 12 predictors which have good inter-seasonal scale and correlations value, which are mslp, p__u, p5_u, p8_u, p850, p__v, p8_v, rhum, r500, shum, s500 and temp.

The other inter-seasonal southwest monsoon plots for predictors between ERAI and CCSM3, ECHAM5 and MIROC-m were also calculated (not shown here) and the same trends (magnitude and distribution) as that of MPI-ESM were found.

3.3 Selection of predictors

As mentioned earlier, the results from the above sections (spatial correlations and inter-seasonal plot) were used in the selection of predictors which are inputs to the ANN method of downscaling. With the position of Bangkok station in Fig. 1, the number of input predictors was selected based on the contour lines distribution of the correlation value surrounding Bangkok region. The 12 predictors, as cited earlier in Section 3.1, for ERAI training dataset with high spatial

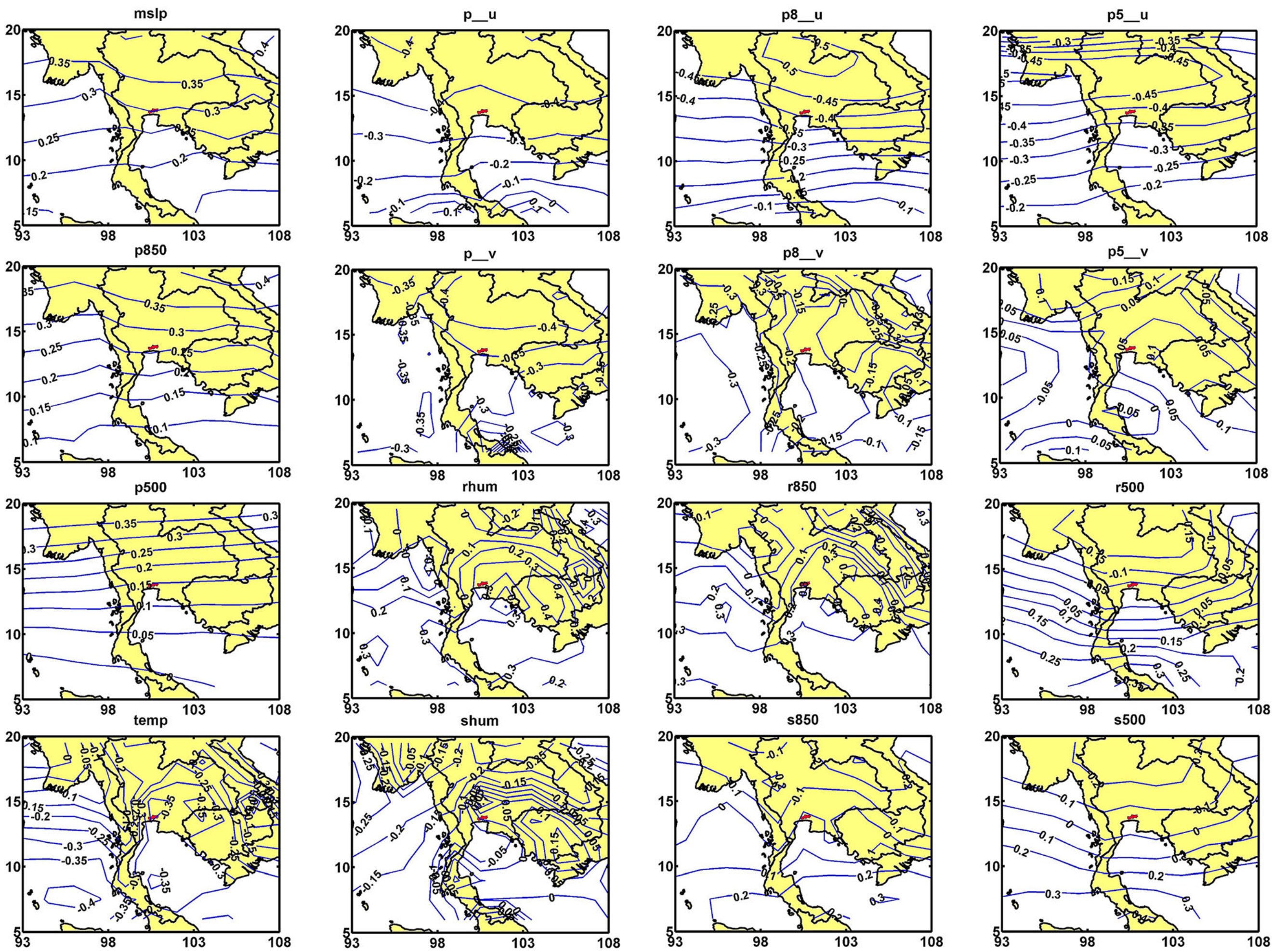


Fig. 3 Spatial correlation coefficient between ERAI predictors and Bangkok rainfall

correlations against Bangkok rainfall station data are mslp, p__u, p8_u, p5_u, p850, p500, p__v, rhum, r850, shum, s500 and temp (Fig. 3). However, based on the inter-seasonal plots for southwest monsoon predictors between ERAI and MPI-ESM (Fig. 4), the two predictors: p500 and r850 were ignored. Thus, only 10 variables: mslp, p__u, p8_u, p850, p__v, rhum, r500, shum, s500 and temp were selected as predictors for MPI-ESM model, which are also common in the ERAI predictor set. The final sets of predictors, derived thus, for different training/testing sets using ERAI and all other GCMs, are shown in Table 2.

These selected ten predictors, from the ERAI reanalyses, are used as input for model training. As displayed in Fig. 1, there are 25 grid boxes used in the computation process. Hence, there are a total of 10 × 25 = 250 predictors as input to the ANN training process. This large number of predictors and in addition, the similar predictors at neighbouring grid boxes would have a very high correlation value. Therefore, a direct use of the predictor variables, in statistical regression, may lead to multi-collinearity and may be computationally expensive. Therefore, to reduce the input dimension to the downscaling process and to choose the best set of predictors that contain all essential climate information, the PCA procedure was applied. The principal components were selected on the basis of the percentage of variance of original data. Statistical downscaling model mapped the variability of climate variables to the variability of rainfall using regression. This criterion was also used by Hughes and Guttorp (1994), Tripathi et al. (2006) and Ghosh (2010).

4 Results and discussion

4.1 Analyses of the present-day climate using ANN

The analyses of the present-day climate were done using the ERAI reanalyses data for the MJJASO season. The PCA-filtered predictors were given as input to the ANN. The training period for the ANN was fixed for a period of 15 years from 1986 to 2000 and the testing period is fixed from 1980 to 1985. The performance of the training and testing set between downscaled ERAI (using predictors for MPI-ESM) and

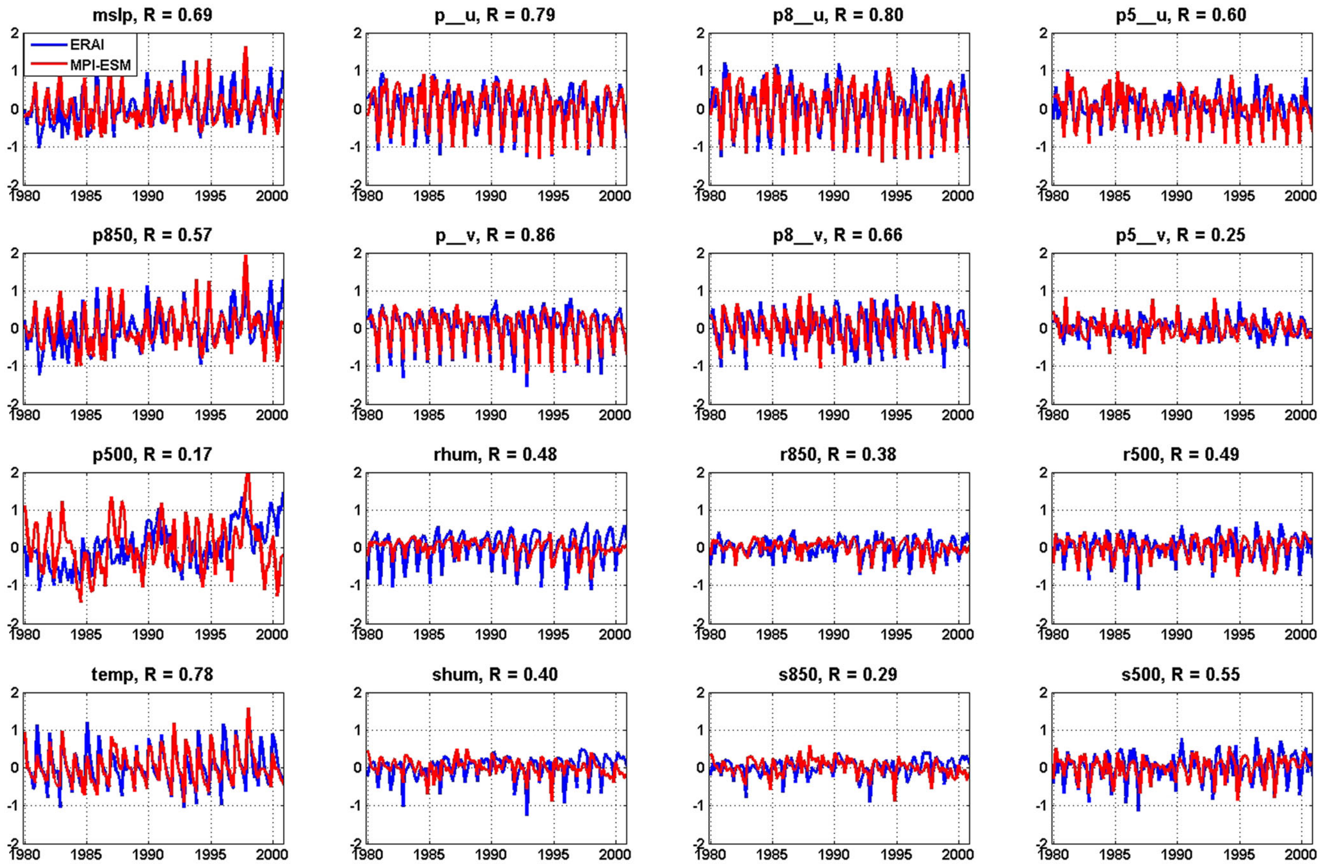


Fig. 4 Inter-seasonal southwest monsoon standardized predictors between ERAI and GCM MPI-ESM for baseline climate 1980–2000

station data is evaluated using the Pearson correlation coefficient (CC), mean square error (MSE), root mean square error (RMSE) and Nash-Sutcliffe efficiency (NSE) (Nash and Sutcliffe 1970) formulated as equations (3), (4) and (5):

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(O_i - S_i)^2} \quad (3)$$

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(O_i - S_i)^2 \quad (4)$$

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{N}(O_i - S_i)^2}{\sum_{i=1}^{N}\left(O_i - \bar{O}\right)^2} \quad (5)$$

where, O and S are the observed and simulated precipitation and N, the period of time steps.

The statistical indices for the above performances are shown in Table 3. The statistical indices (CC, RMSE, MSE and NSE) are relatively better for the training dataset than the testing dataset when compared to station rainfall. The correlation coefficient for training is about 0.8 which is reasonably good and also the testing's CC is about 0.7. The interannual monsoon season MJJASO was plotted against station rainfall data and downscaled ERAI dataset in Fig. 5 for both training and testing periods in monthly scale. From Fig. 5, it is seen that the downscaled ERAI (blue dotted line) underestimates the peak and overestimates the trough of station rainfall (continuous black line). Overall, these results indicate the chosen ANN method proved to be a good tool in validating the downscaled reanalyses.

It is to be noted here that there were four sets of predictors, each, ERAI against four GCMs. The above ANN results are based on the set of ERAI and GCM MPI-ESM using 250 predictors selected from Section 3. The other sets of training/testing to downscale [175 predictors for ECHAM5, 150 predictors for CCSM3 and 150 predictors for MIROC-medres] specified in Table 2 are not shown here. These four sets were used to downscale the present-day 1980–2000 and future 2080–2100 scenario A1B and RCP4.5 for the four respective GCMs, as described in Section 2.2.3.

Table 3 Statistics between station data and downscaled reanalysis ERAI for rainy season: training and testing set

Statistics	Training	Testing
CC	0.81	0.71
RMSE	2.43	2.87
MAE	1.93	2.05
NSE	0.65	0.50

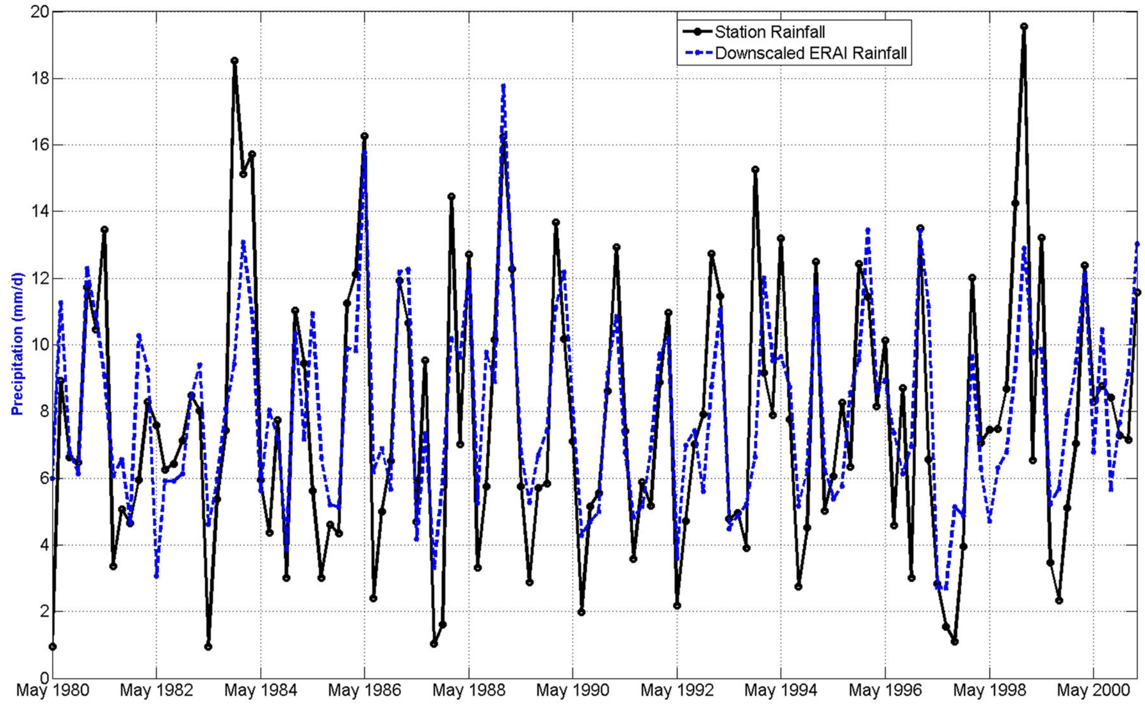


Fig. 5 Inter-seasonal southwest monsoon rainfall with station observation data and downscaled ERAI

4.2 Future rainfall over Bangkok

In order to assess the future rainfall, downscaled projection GCMs were used in the study (see Section 2.2.3). These GCMs form an ensemble dataset, which could give a range of uncertainties of the future climate. Initially, the percentage change in precipitation (with respect to the present-day climate) was assessed, followed

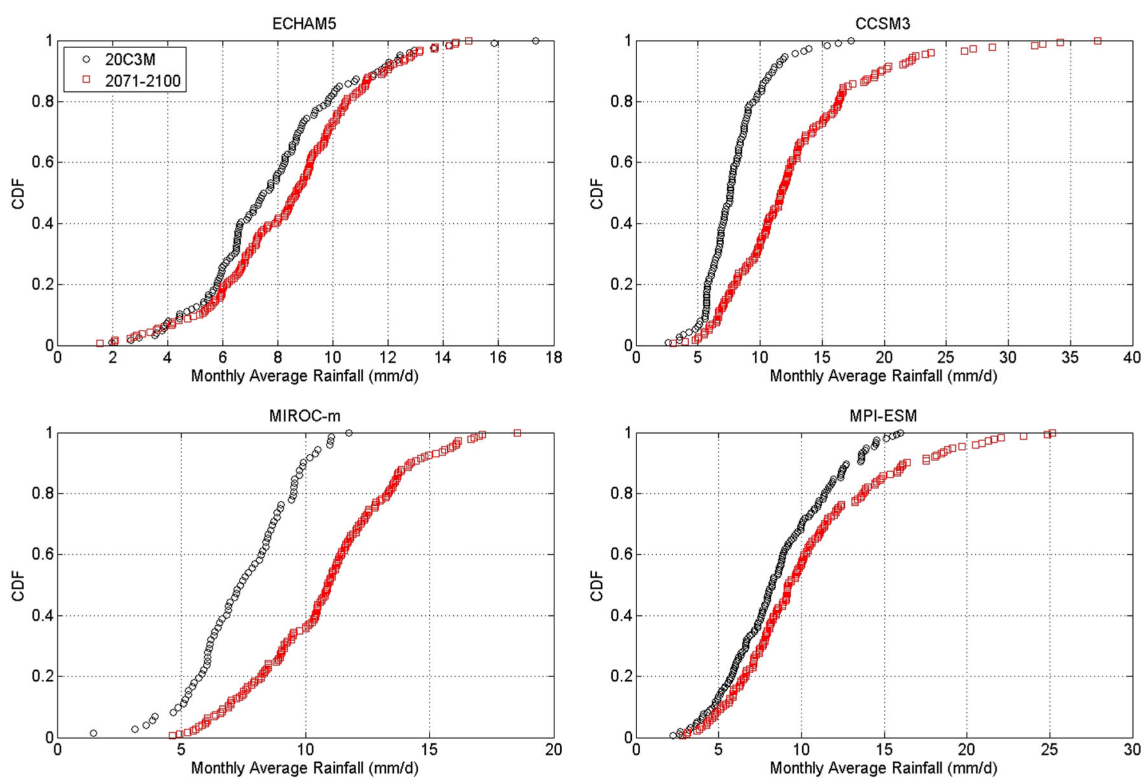


Fig. 6 CDF of downscaled present and projected southwest monsoon rainfall from four GCMs

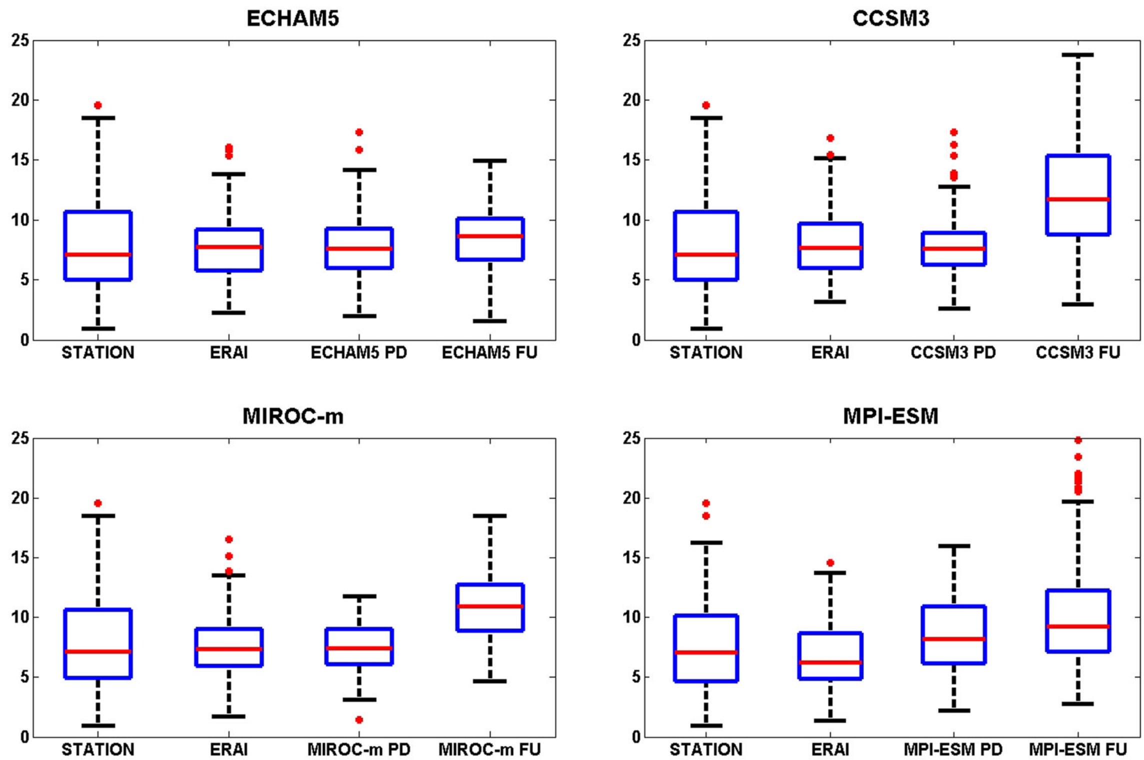


Fig. 7 Boxplot for station rainfall data, trained ANN model with ERAI, downscaled present and projected southwest monsoon from four GCMs

by some statistical evaluations of the changes in rainfall extremes.

4.2.1 Change in magnitude and trend of future precipitation

The results for statistical downscaling of the present-day climate and their future projections are displayed in Fig. 6 as cumulative distribution function (CDF) of both datasets. The CDF graphs of GCM projections are distributed to the right of the present-day climate, indicating that there is an agreement with the increasing trend for all downscaled GCMs in future climate.

The box plots between station data, downscaled reanalysis data and present-day and future climate of the same GCMs are displayed in Fig. 7. This comparison has been done to evaluate the agreement of the GCMs against both station data and reanalyses. At the same time, the median changes of the downscaled GCMs also indicate increases (changes) in rainfall over the future climate compared to the present day. Especially, the boxplots of CCSM3 and MIROC-m clearly show an increasing trend in the future rainfall compared to the present-day conditions.

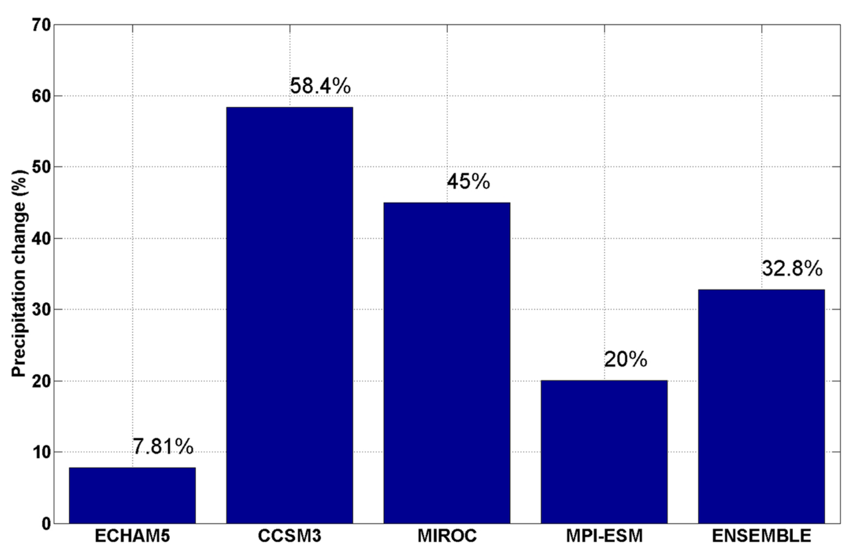


Fig. 8 Delta change (%) between projected southwest monsoon rainfall and present for different downscaled GCMs. The ensemble value is the average of four GCMs's delta change

The change in magnitude of future precipitation was also assessed by using the ‘delta change factor’ approach. This approach has been applied to calculate the different between future and present-day climate using Eq. (6).

$$\Delta(\%) = \frac{(F-P)}{P} \times 100 \quad (6)$$

where F is the future rainfall and P is the present-day rainfall.

Overall, there is an agreement in increasing rainfall during the rainy season, for all four GCMs (Fig. 8). The increase is different for all GCMs during the wet season, from lowest of ECHAM5 A1B (7.8 %) to highest of CCSM3 A1B (58 %). The ensemble results between four GCMs is the average of the four, stands at 32.8 %, which is significant to show that Bangkok might experience wetter conditions during the rainy season in the future.

4.2.2 Change in extreme precipitation indices

Assessing the mean changes in the future climate alone is not sufficient to understand the changes in their entirety as the

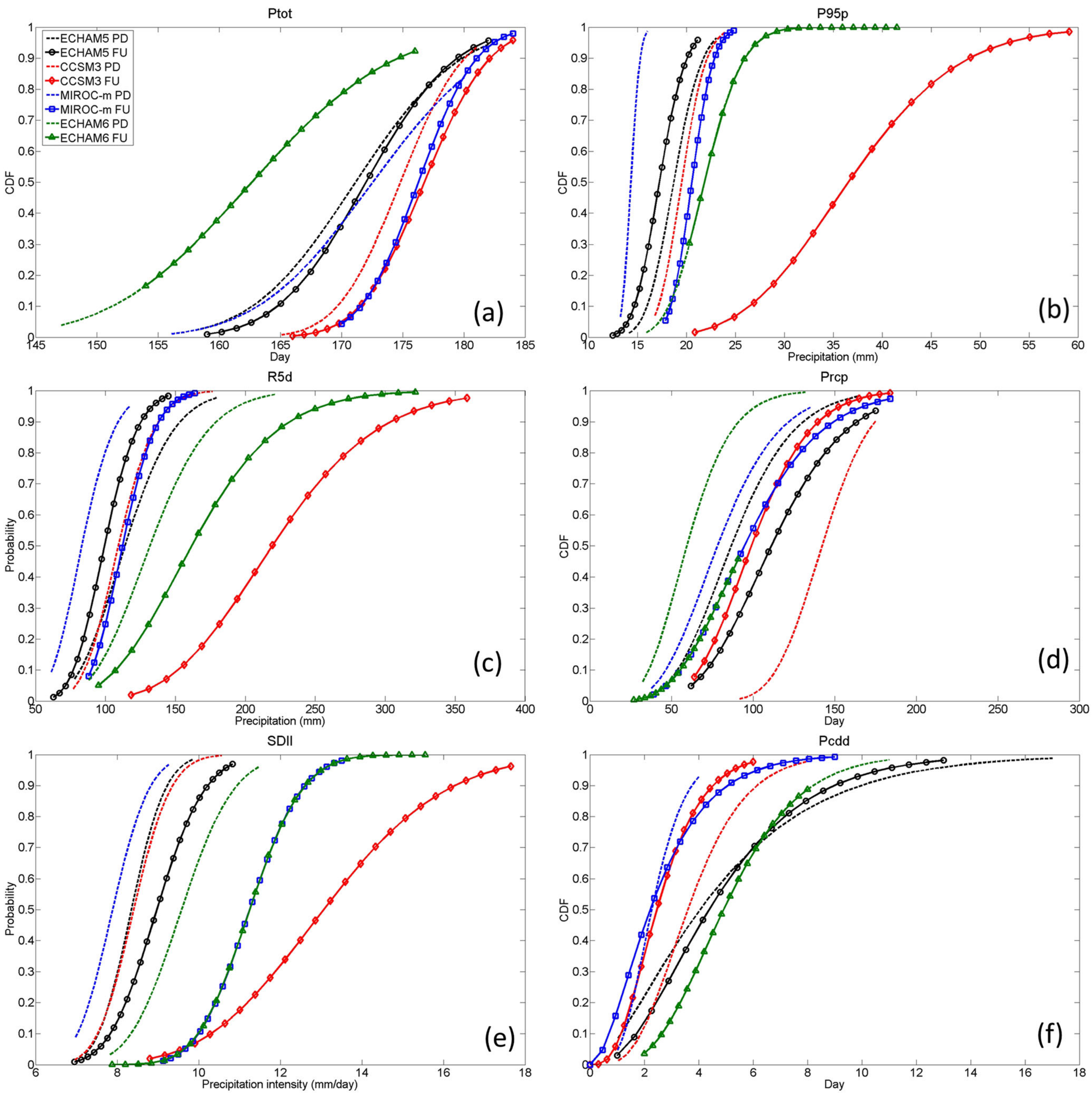


Fig. 9 Cumulative gamma distribution function of statistical indices for downscaled southwest monsoon rainfall, both present (*PD*) and projection (*FU*), from GCMs. **a** Ptot, **b** P95p, **c** R5d, **d** Prcp, **e** SDII and **f** Pcdd

mean values mask the changes in extremes. To this end, we chose a few indices, from the statistical and regional dynamical downscaling of extremes (STARDEX) (The European Commission-funded project) (available at http://www.cru.uea.ac.uk/cru/projects/stardex/), to assess future changes. These indices are the following:

- Prcp Annual maximum wet day frequency (day)
- Ptot Total number of wet days (day)
- SDII Wet day precipitation intensity (mm/day)
- R5d Maximum 5-day consecutive rain (mm)
- P95p 95th percentile of precipitation (mm)
- Pcdd Annual maximum dry day frequency (day)

The relative changes, for these indices, were estimated from future (2080–2100) with respect to the present day (1980–2000), as displayed in Fig. 9. For an easy comparison, the distributions of these six indices have been shown for both for present day and future together, so that the changes in the future distributions are evident. The distributions were calculated for the fitted cumulative gamma probability distribution functions (CDFs) of these indices for baseline period (plotted as dashed line) and future period (continuous line). The present day climate, in the legend, is abbreviated as 'PD', the future 'FU'. Each downscaled GCM is displayed in different colours and line patterns, for clarity. There are eight curves in each statistical index plot to represent both present-day and future climates of the four downscaled GCMs.

Results indicate that the overall trend for the future wet indices (SDII, Prcp, Ptot, R5d, P95p) are on the increase. In contrast, the dry index (Fig. 9f) distribution of the future CDF curve shows mixed trends. These imply that the future climates are likely to be dominated by increases in the precipitation indices during the rainy season. This was the question we bore in mind as the crux of this study, and some initial answers have been made available from these results. These can have strong implications as the increasing extreme rainfall intensity over Bangkok during rainy season may induce large flooding and inundation for the area, which was already vulnerable to one such flooding in 2011. Yet, these results are not final answers such as more detailed work as to incorporating more GCMs and scenarios may be necessary for higher confidence in the results.

5 Conclusion

We performed an empirical downscaling using ANN over Bangkok to assess future changes in precipitation during the rainy season. The reanalysis dataset ERA-Interim was filtered using PCA and downscaled using ANN over the Bangkok region, by choosing a suite of appropriate predictors. Later, we investigated the possible changes in future southwest monsoon rainfall using three different GCMs (CCSM3, ECHAM5 and MIROC-m), under CMIP3 A1B scenario and one CMIP5 GCM MPI-ESM under RCP 4.5. The training and testing using the reanalysis predictors showed the capability of the statistical downscaling by PCA and ANN to reproduce the observed long-term precipitation trend. Changes in the future rainfall indicate an increase in rainfall during MJJASO months. The statistical indices analysis on rainfall extremes shows an increase towards the future for the wet indices. Especially, an increase in extreme wet day rainfall intensity during rainy season is likely that may induce flooding and inundation to the region. These have strong implications for Bangkok region that is highly vulnerable to climate changes and needs adequate mitigation measures. Although this study considered only a few GCMs/scenarios, a larger ensemble of such GCMs/scenarios may expand the range of uncertainties for a better understanding of possible changes. However, we have presented a cost-effective approach in assessing climate changes using the ANN method, in contrast to employing computationally expensive dynamical downscaling. This method can also be used for larger ensembles, something that could prove useful by giving important information for policy makers for adaptation.

Acknowledgments The authors thank the Center for Hazards Research, Department of Civil and Environmental Engineering and CENSAM, SMART, both at the National University of Singapore. The second co-author wishes to thank the Thailand Research Fund through the Royal Golden Jubilee Ph.D. Programme (PHD/0221/2550) funded to Rangsit University for the doctoral programme and the Tropical Marine Science Institute, National University of Singapore, where he spent his internship.

References

ADB, 2009: The economics of climate change in Southeast Asia: a regional review, Manila

Anandhi A, Srinivas VV, Kumara DN, Nanjundiahb RS (2009) Role of predictors in downscaling surface temperature to river basin in India for IPCC SRES scenarios using support vector machine. Int J Climatol 29:583–603. doi:10.1002/joc.1719

Bachelet D, Brown D, Bohm M, Russell P (1992) Climate change in Thailand and its potential impact on rice yield. Clim Chang 21: 347–366

Bigio A (2003) Cities and climate change. In: Kreimer A, Arnold M, Carlin A (eds) Building Safer Cities: The Future of Disaster Risk. World Bank, Washington, pp. 91–100

Bishop CM (1995) Neural networks for pattern recognition. Clarendon Press, Oxford. ISBN-13: 978–0198538646.

Björck A (1996) Numerical methods for least squares problems. SIAM, Philadelphia. ISBN 0-89871-360-9.

Boochabun K, Tych W, Chappell NA, Carling PA, Lorsirirat K, Pa-Obsaeng S (2004) Statistical modelling of rainfall and river flow in Thailand. J Geo Soc India 64:503–515

Brands S, Herrera S, San-Martin D, Gutierrez JM (2011) Validation of the ENSEMBLES global climate models over southwestern Europe using probability density function, from a downscaling perspective. Clim Res 48:145–161

Charles SP, Bates BC, Hughes JP (1999) A spatio-temporal model for downscaling precipitation occurrence and amounts. J Geophys Res 104:31657–31669. doi:10.1029/1999JD900119

Chen H, Xu CY, Guo S (2012a) Comparison and evaluation of multiple GCMs, statistical downscaling and hydrological models in the study of climate change impacts on runoff. J Hydrol 434-435:36–45

Chen J, Brissette F, Leconte R (2012b) Coupling statistical and dynamical methods for spatial downscaling of precipitation. Clim Chang 114: 509–526. doi:10.1007/s10584-012-0452-2

Coulibaly P, Anctil F, Bobée B (2000) Daily reservoir inflow forecasting using artificial neural networks with stopped training approach. J Hydrol 230:244–257. doi:10.1016/S0022-1694(00)00214-6

Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer P, Bechtold P, Beljaars ACM, van de Berg L, Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M, Geer AJ, Haimberger L, Healy SB, Hersbach H, Hólm EV, Isaksen L, Kållberg P, Köhler M, Matricardi M, McNally AP, Monge-Sanz BM, Morcrette JJ, Park BK, Peubey C, de Rosnay P, Tavolato C, Thépaut JN, Vitart F (2011) The ERA-interim reanalysis: configuration and performance of the data assimilation system. Quar J Royal Met Soc 137:553–597. doi:10.1002/qj.828

Dibike YB, Coulibaly P (2005) Hydrologic impact of climate change in the Saguenay watershed: comparison of downscaling methods and hydrologic models. J Hydrol 307:145–163

Duhan D, Pandey A (2014) Statistical downscaling of temperature using three techniques in the Tons River basin in central India. Theor Appl Climatol. doi:10.1007/s00704-014-1253-5

FAO (1982) A study of the agroclimatology of the humid tropics of Southeast Asia: technical report, FAO/UNESCO/WMO Interagency project on Agroclimatology, pp221

Fistikoglu O, Okkan U (2011) Statistical downscaling of monthly precipitation using NCEP/NCAR reanalysis data for Thatali River basin in Turkey. J Hydrol Eng 16(2):157–164

Ghosh S (2010) SVM-PGSL coupled approach for statistical downscaling to predict rainfall from GCM output. J of Geophys Res 115, D22102. doi:10.1029/2009JD013548.

Gyalistras D, Hv S, Fischlin A, Beniston M (1994) Linking GCM-simulated climatic changes to ecosystem models: case studies of statistical downscaling in the Alps. Clim Res 4:167–189. doi:10.3354/cr004167

Haykin S (1994) Neural networks: a comprehensive foundation. Macmillan College. ISBN 13: 9780023527616

Hewitson BC, Crane RG (1996) Climate downscaling: techniques and application. Clim Res 7:85–96

Hsu KL, Gupta HV, Sorooshian S (1995) Artificial neural network modeling of the rainfall-runoff process. Water Resour Res 31: 2517–2530. doi:10.1029/95WR01955

Hu Y, Maskey S, Uhlenbrook S (2013) Downscaling daily precipitation over the Yellow River source region in China: a comparison of three statistical downscaling methods. Theor Appl Climatol 112(3–4): 447–460. doi:10.1007/s00704-012-0745-4

Hu Y, Maskey S, Uhlenbrook S (2012) Trends in temperature and precipitation extremes in the yellow river source region, China. Clim Chang 110:403–429. doi:10.1007/s10584-011-0056-2

Hughes JP, Guttorp P (1994) A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena. Water Resour Res 30(5):1535–1546

Hung NQ, Babel MS, Weesakul S, Tripathi NK (2009) An artificial neural network model for rainfall forecasting in Bangkok, Thailand. Hydrol Earth Syst Sci 13:1413–1425

IPCC (2013) Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp.

Jolliffe IT (1986) Principal component analysis. Springer New York. ISBN 978–0–387–22440–4

Jones RG, Noguer M, Hassell DC, Hudson D, Wilson SS, Jenkins GJ, Mitchell J (2004) Generating high resolution climate change scenarios using PRECIS. Taylor R (ed.), p44

Kang H, An KH, Park CK, Solis ALS, Stitthichivapak K (2007) Multimodel output statistical downscaling prediction of precipitation in the Philippines and Thailand. Geophys Res Lett 34: L15710. doi: 10.1029/2007GL030730

Kilsby CG, Jones PD, Burton A, Ford AC, Fowler HJ, Harpham C, James P, Smith A, Wilby RL (2007) A daily weather generator for use in climate change studies. Environ Modelling Software 22:1705–1719. doi:10.1016/j.envsoft.2007.02.005

Kuligowski RJ, Barros AP (1998) Experiments in short-term precipitation forecasting using artificial neural networks. Mon Wea Rev 126: 470–482

Liu W, Fu G, Liu C, Charles SP (2013) A comparison of three multi-site statistical downscaling models or daily rainfall in the North China Plain. Theor Appl Climatol 111:585–600. doi:10.1007/S00704-012-0692-0

Liu Z, Xu Z, Charles SP, Fu G, Liu L (2011) Evaluation of two statistical downscaling models for daily precipitation over an arid basin in China. Int J Climatol 31(13):2006–2020. doi:10.1002/joc.2211

Mahmood R, Babel MS (2013) Evaluation of SDSM developed by annual and monthly sub-models for downscaling temperature and precipitation in the Jhelum basin, Pakistan and India. Theor Appl Climatol 113:27–44. doi:10.1007/s00704-012-0765-0

Martin TH, Howard BD, Mark B, Orlando DJ (1996) Neural networks design. PWS Publishing Company. ISBN-13: 978–0–9717321–1–7.

Mendes D, Marengo JA (2010) Temporal downscaling: a comparison between artificial neural network and autocorrelation techniques over the Amazon Basin in present and future climate change scenarios. Theor Appl Climatol 100:413–421. doi:10.1007/s00704-009-0193-y

Murphy J (1999) An evaluation of statistical and dynamical techniques for downscaling local climate. J Clim 12:2256–2284

Murphy J (2000) Predictions of climate change over Europe using statistical and dynamical downscaling techniques. Int J Climatol 20:489–501

Nash JE, Sutcliffe JV (1970) River flow forecasting through conceptual models part I—a discussion of principles. J Hydrol 10(3):282–290

Okkan U, Fistikoglu O (2013) Evaluating climate change effects on runoff by statistical downscaling and hydrological model GR2M. Theor Appl Climatol. doi:10.1007/s00704-013-1005-y

Peel MC, Finlayson BL, McMahon TA (2007) Updated world map of the Koppen-Geiger climate classification. Hydrol Earth Syst Sci 11: 1633–1644

Priddy KL, Keller PE (2005) Artificial neural networks: an introduction. SPIE – The International Society for Optical Engineering, Bellingham

Senga R (2009) Natural or unnatural disasters: the relative vulnerabilities of Southeast Asian megacities to climate change. WWF Report on "Mega-Stress for Mega-Cities".

Sharma D, Babel M (2013) Application of downscaled precipitation for hydrological climate-change impact assessment in the upper Ping River Basin of Thailand. Clim Dyn 41:2589–2602. doi:10.1007/s00382-013-1788-7

Tangang FT, Tang B, Monahan AH, Hsieh WW (1998) Forecasting ENSO events: a neural network–extended EOF approach. J Clim 11:29–41

Tripathi S, Srinivas VV, Nanjundiah RS (2006) Downscaling of precipitation for climate change scenarios: a support vector machine approach. J Hydrol 330:621–640

van Vuuren DP, Edmonds JA, Kainuma M, Riahi K, Thomson AM, Hibbard K, Hurtt GC, Kram T, Krey V, Lamarque J-F, Masui T, Meinshausen M, Nakicenovic N, Smith SJ, Rose S (2011) The representative concentration pathways: an overview. Clim Chang 109: 5–31. doi:10.1007/s10584-011-0148-z

von Storch H (1995) Inconsistencies at the interface of climate impact studies and global climate research. Meteorol Z 4:72–80

von Storch H, Zorita E, Cubasch U (1993) Downscaling of global climate change estimates to regional scales: an application to Iberian rainfall in wintertime. J Clim 6:1161–1171

Widmann M, Bretherton CS (2000) Validation of mesoscale precipitation in the NCEP reanalysis using a new gridcell dataset for the northwestern United States. J Clim 13:1936–1950

Wilby RL, Dawson CW (2004) Using SDSM version 3.1—a decision support tool for the assessment of regional climate change impacts, user manual.

Wilby RL, Dawson CW, Barrow EM (2002) SDSM—a decision support tool for the assessment of regional climate change impacts. Environ Modelling Software 17: 145–157. doi: 10.1016/S1364-8152(01) 00060-3

Wilby RL, Hassan H, Hanaki K (1998) Statistical downscaling of hydrometeorological variables using general circulation model output. J Hydrol 205: 1–19. doi: 10.1016/S0022-1694(97) 00130–3

Wilby RL, Tomlinson OJ, Dawson CW (2003) Multi-site simulation of precipitation by conditional resampling. Clim Res 23: 183–194

Wilby RL, Wigley TML (1997) Downscaling general circulation model output: a review of methods and limitations. Prog Phys Geogr 21: 530–548. doi:10.1177/030913339702100403 21

Wilby RL, Wigley TML (2000) Precipitation predictors for downscaling: observed and general circulation model relationships. Int J Climatol 20:641–661

Zorita E, von Storch H (1997) A survey of statistical downscaling techniques 45p. ISSN 0344-9629