

Long title:

# **Statistical Downscaling of General Circulation Model Outputs to Precipitation**

## **Part 1: Calibration and Validation**

Short title:

# **Downscaling of GCM Outputs to Precipitation Calibration and Validation**

**D.A. Sachindra<sup>a,\*</sup>, F. Huang<sup>a</sup>, A. Barton<sup>a,b</sup>, and B.J.C. Perera<sup>a</sup>**

*<sup>a</sup>Victoria University, P.O. Box 14428, Melbourne, Victoria 8001, Australia.*

*<sup>b</sup>University of Ballarat, PO Box 663, Ballarat, Victoria 3353, Australia.*

\*Correspondence to: D. A. Sachindra, College of Engineering and Science, Footscray Park Campus, Victoria University, P.O. Box 14428, Melbourne, Victoria 8001, Australia. E-mail: [sachindra.dhanapalaarachchige@live.vu.edu.au](mailto:sachindra.dhanapalaarachchige@live.vu.edu.au). Telephone: +61 03 9919 4907.

The authors acknowledge the financial assistance provided by the Australian Research Council Linkage Grant scheme and the Grampians Wimmera Mallee Water Corporation for this project.

## **Abstract**

This paper is the first of two companion papers providing details of the development of two separate models for statistically downscaling monthly precipitation. The first model was developed with NCEP/NCAR reanalysis outputs and the second model was built using the outputs of Hadley Centre Coupled Model version 3 GCM (HadCM3). Both models were based on the multi-linear regression (MLR) technique and were built for a precipitation station located in Victoria, Australia. Probable predictors were selected based on the past literature and hydrology. Potential predictors were selected for each calendar month separately from the NCEP/NCAR reanalysis data, considering the correlations that they maintained with observed precipitation. Based on the strength of the correlations, these potential predictors were introduced to the downscaling model until its performance in validation, in terms of Nash-Sutcliffe Efficiency (NSE), was maximised. In this manner, for each calendar month, the final sets of potential predictors and the best downscaling models with NCEP/NCAR reanalysis data were identified. The HadCM3 20<sup>th</sup> century climate experiment data corresponding to these final sets of potential predictors were used to calibrate and validate the second model. In calibration and validation, the model developed with NCEP/NCAR reanalysis data displayed NSEs of 0.74 and 0.70, respectively. The model built with HadCM3 outputs showed NSEs of 0.44 and 0.17 during the calibration and validation periods, respectively. Both models tended to under-predict high precipitation values and over-predict near-zero precipitation values, during both calibration and validation. However, this prediction characteristic was more pronounced by the model developed with HadCM3 outputs. A graphical comparison of observed precipitation, the precipitation reproduced by the two downscaling models and the raw precipitation output of HadCM3, showed that there is large bias in the precipitation output of HadCM3. This indicated the need of a bias-correction, which is detailed in the second companion paper.

**Keywords:** *Statistical downscaling, Precipitation, General Circulation Model*

## **1. INTRODUCTION**

Changes in the global climate since the 20<sup>th</sup> century (notably rises in the global temperature), were mostly attributed to anthropogenic greenhouse gas (GHG) emissions, rather than natural variability in climate (Crowley, 2000). Furthermore, as stated in IPCC (2007), the rise in global and continental temperatures during the 20<sup>th</sup> century can be credibly reproduced with climate models, only if both natural and anthropogenic forces were considered. Sea level rise, reduction of snow coverage, extreme precipitation events, heat waves and rise in the frequencies of hot events and tropical cyclones are considered to be some of the impacts of climate change (Alavian et al., 2009).

Over the period 1997-2008, the average precipitation over the southern part of southeast Australia declined by about 11% from the long term average, leading to a reduction in runoff of approximately 35% (Chiew et al., 2010). The Australian state of Victoria suffered a severe drought (referred to as “the Millennium drought”) from 1997, until the torrential rainfalls in late 2010 and early 2011. During 1998-2007, annual average precipitation in Victoria decreased by about 13% from the long term average and the highest decline in rainfall of 28%, occurred over the autumn months. The average rainfall in autumn and early winter dropped well below the long term average, while the rainfall in summer remained as it was (Timbal and Jones, 2008). This drought forced the introduction of severe water restrictions in many regions of Victoria. The region southwest of Western Australia is experiencing a drought which has been in effect since late 1960s (Smith et al., 2000). Unlike the Millennium drought, which has now ended, the drought in southwest of Western Australia has not shown any signs of ending and is considered to have experienced a step change in climate (Government of Western Australia Department of Water, 2009). The changes in the climate described in the above examples are believed to be the possible impacts of anthropogenic climate change and natural variability of the climate.

Precipitation is regarded as the predominant factor in determining the availability of water resources in a catchment. The food supply of humans and animals, irrigation, hydropower generation and recreational purposes are just some of the major sectors directly under the influence of precipitation. Hence it is

understood that the reliable prediction of future precipitation, especially under a changing climate, is of great importance in assessing future water availability.

General circulation models (GCMs) are considered the most reliable tools in studying climate change (Maraun et al., 2010). They have proven their potential in reproducing the past observed climatic changes, considering the GHG concentrations in the atmosphere (Goyal et al., 2011). However, GCMs produce their projections at relatively coarse spatial scales and they are unable to resolve sub-grid scale features such as topography, clouds and land use. Since GCMs generate outputs at coarse grid scales in the order of a few hundred kilometres, their outputs cannot be directly used in catchment scale climate impact studies, which usually need hydroclimatic data at fine spatial resolutions. The scale mismatch between the GCM outputs and the hydroclimatic information needed at the catchment level is a major obstacle in climate impact assessment studies of hydrology and water resources (Willems and Vrac, 2011).

As a solution to the scale mismatch between the GCMs outputs and the hydroclimatic information required at catchment scale, downscaling techniques have been developed. Downscaling techniques are classified into two broad classes; dynamic downscaling and statistical downscaling. In dynamic downscaling, outputs of GCMs are fed into regional climate models (RCMs) as boundary conditions to enable the prediction of the regional climate at the spatial scale of 5-50 kilometres (Yang et al., 2012). This procedure is based on the complex physics of atmospheric processes and involves high computational costs. In dynamic downscaling techniques, it is assumed that the parameterisation schemes selected for the past climate are also valid for the climate in future. In addition, dynamic downscaling techniques are highly dependant on the boundary conditions provided by the GCMs. However, dynamic downscaling could produce spatially distributed hydroclimatic predictions over the catchment of interest (Maurer et al., 2008).

Statistical downscaling relies on the empirical relationships derived between the GCM outputs (predictors of downscaling models) and the catchment scale hydroclimatic variables (predictands of downscaling models) such as precipitation, streamflow, and evaporation (Hay and Clark, 2003). Unlike dynamic downscaling, statistical downscaling does not involve complex atmospheric physics and hence is

computationally less expensive (Sachindra et al., 2012). In statistical downscaling, for the establishment of relationships between the GCM outputs and the catchment scale hydroclimatic variables, preferably long records of observed hydroclimatic data are required (Sachindra et al., 2013). This is because a long record of observations could possibly contain the full variability of the observed climate and hence allow the downscaling models to better model the changes in the climate. However this can limit the effective use of statistical downscaling in data scarce regions. Statistical downscaling techniques are based on the major assumption that the relationships derived between the GCM outputs and the catchment scale hydroclimatic variables for the past observed climate are equally valid for the future, under changing climate (von Storch et al., 2000). Also similar to dynamic downscaling, statistical downscaling techniques are highly dependent on the outputs of the GCMs which are used as inputs to the downscaling model.

Statistical downscaling techniques are grouped under three categories; weather classification, regression models and weather generators (Wilby et al., 2004). In weather classification methods, large scale weather patterns are grouped under a finite number of discrete states (Anandhi, 2010). Then the links between the catchment scale weather at certain times and the large scale weather patterns are identified. Hence, by considering the large scale weather patterns at any given time, the corresponding catchment scale weather can be deduced. The method of meteorological analogues (Timbal et al., 2009, Charles et al., 2013; Shao and Li, 2013) and recursive partitioning (Schnur and Lettenmaier, 1998) are examples for the weather classification techniques. Regression techniques develop either linear or non-linear regression equations between the GCM outputs and the catchment scale hydroclimatic variables. Regression based downscaling methods are regarded as the most widely used statistical downscaling techniques (Nasseri et al., 2013). This is mainly due to their simplicity and effectiveness. Chu et al. (2010) used Multi-Linear Regression (MLR) for downscaling GCM outputs to daily mean temperature, pan evaporation and precipitation. Tisseuil et al. (2010) used Artificial Neural Networks (ANN), Generalized Additive Models (GAM), Generalized Linear Models (GLM), and Aggregated Boosted Trees (ABT) for downscaling GCM outputs to daily streamflows. Gene Expression Programming (GEP) and MLR techniques were employed by Hashmi et al. (2011) for downscaling GCM outputs to daily precipitation. The Least Square Support Vector Machine Regression (LS-SVM-R) was used by Anandhi et al. (2012) and Sachindra et al. (2013) for downscaling GCM outputs to daily relative humidity and monthly streamflows respectively.

Model Output Statistics (MOS) is a statistical downscaling technique used in post-processing the outputs of climate or weather models (Maraun et al., 2010), by relating them with catchment scale observation using a linear regression technique (Marzban et al. 2006). This enables the reduction of systematic bias in the predictions of the model. Weather generators produce weather data for the future by scaling their parameters according to the corresponding changes characterised in the GCM outputs for the future. These techniques possess the advantage of generating series of climatic data of any desired length of time with similar statistical properties as observations used in the weather generator (Khalili et al., 2009). The combination of Markov chains and two parameter Gamma distribution is an example of a weather generator (Richardson, 1981), in which Markov chains are used to predict the occurrences of a climatic variable and the Gamma distribution is used to determine the corresponding amounts. The applications of weather generators in statistical downscaling are found in the studies of Semenov and Stratonovitch (2010), Iizumi et al. (2012), Khazaei et al. (2013).

In general, any statistical downscaling model is calibrated and validated (developed) using the reanalysis outputs (e.g. NCEP/NCAR) and observations, corresponding to the past climate. For producing the future projections, outputs of a GCM pertaining to a certain GHG emission scenario are introduced to this downscaling model. This procedure does not provide a smooth transition from the model development phase (calibration and validation) to the future projection phase, as the former and latter steps are performed with the outputs of two different sources which have different levels of accuracy. In other words, the inputs used in the development phase and the future projection phase of a conventional downscaling model are not homogeneous. As a solution to this issue, a downscaling model calibrated and validated with GCM outputs can be used in producing future projections with the outputs of the same GCM, pertaining to a future GHG emission scenario. Since the outputs of the same GCM are used for the model development and future projections, there is homogeneity in the modelling process. However, in the published literature there was no evidence of past studies which attempted the use of a downscaling model developed with GCM outputs.

This paper, which is the first of a series of two companion papers, discusses the calibration and validation of two statistical downscaling models based on Multi-Linear Regression (MLR) technique. The two

statistical downscaling models were developed separately, for downscaling monthly outputs of (1) National Centers for Environmental Prediction / National Center for Atmospheric Research (NCEP/NCAR) reanalysis and (2) Hadley Centre Coupled Model version 3 GCM (HadCM3), to monthly precipitation. As the case study, a precipitation station located within the Grampians water supply system in north-western Victoria in Australia was selected. A performance comparison between two downscaling models for the calibration and validation phases was also performed.

Downscaling GCM outputs to precipitation at monthly temporal scale does not permit capturing the variations of precipitation within a month (e.g. wet and dry days, precipitation intensity and extremes of precipitation). However, still monthly precipitation projections produced using downscaling models could aid in the management of water resources which include operations such as water allocation for crops, domestic and industrial needs, and also environmental flows, especially in the planning stage of a water resources project.

The remainder of this paper was structured as follows. The study area and the data used in the study were briefly described in Section 2, followed by the generic methodology in Section 3. Thereafter, in Section 4, the application of this methodology to the precipitation station considered was provided along with a discussion on the model results. A summary on the model development process and results, along with the conclusions drawn from the study were provided in Section 5. In the second paper the bias-correction and future precipitation projections are detailed.

## **2. STUDY AREA AND DATA**

The Grampians water supply system in north-western Victoria is a large multi reservoir system owned and operated by the Grampians Wimmera Mallee Water Cooperation (GWMWater) ([www.gwmwater.org.au](http://www.gwmwater.org.au)). For this study, a precipitation station at Halls Gap post office (Lat. -37.14°, Lon. 142.52°, elevation from mean sea level about 236m), located in the Grampians system was selected. At this station, the annual average precipitation over the period 1950-2010 was about 950mm. In this region, winter and summer are the wettest and the driest seasons respectively. Observed daily

precipitation record from 1950 to 2010 was obtained from the SILO database (<http://www.longpaddock.qld.gov.au/silo/>) of Queensland Climate Change Centre of Excellence and aggregated to monthly precipitation, for the calibration and validation of downscaling models. In that observed daily precipitation record 31.2% of the data were missing and those missing data were filled by the Queensland Climate Change Centre of Excellence in the SILO database using the spatial interpolation method detailed in Jeffrey et al. (2001). In order to provide the inputs for the calibration and validation of the first downscaling model, NCEP/NCAR monthly reanalysis data for the period 1950-2010 were downloaded from <http://www.esrl.noaa.gov/psd/>. Monthly precipitation outputs produced by the HadCM 3 GCM for the 20<sup>th</sup> century climate experiment were extracted from the Programme for Climate Model Diagnosis and Inter-comparison (PCMDI) (<https://esgcet.llnl.gov:8443/index.jsp>) for the period 1950-1999, for developing the second downscaling model.

### **3. GENERIC METHODOLOGY**

The first step of the downscaling exercise was to define an adequately large atmospheric domain above the precipitation station. It was considered that an adequately large atmospheric domain would enable sufficient atmospheric influence on the climate at the points of interest (e.g. a precipitation station) within the catchment.

A set of probable predictor variables was identified based on a review of past literature on downscaling GCM outputs to precipitation and hydrology. These probable variables are the most likely candidates to influence precipitation at the catchment scale. In selecting predictors in the past studies (e.g. Timbal et al. (2009), Anandhi (2008), Kannan and Ghosh (2013)), factors such as (1) availability in GCM and reanalysis data sets, (2) reliable simulation by GCMs (3) usage in similar studies, (4) fundamentals of hydrology, (4) correlations with the predictand etc were considered. Potential predictors are subsets of the set of probable predictor variables. These sets of potential predictors are the most influential variables on precipitation at the stations considered. The predictor-predictand relationships vary from season to season and also from (geographic) region to region, following the spatio-temporal variations of the atmospheric circulations (Karl et al., 1990). Therefore the sets of potential predictors also vary spatio-temporally. In



this study, in order to better model the precipitation, considering the seasonal variations of the atmospheric circulations, potential predictors were identified for each calendar month, and downscaling models were developed separately for each of the 12 calendar months. Sachindra et al. (2013) found that both Least Square SVM (a complex non-linear downscaling technique) and MLR (a relatively simple linear downscaling technique) have comparable capabilities in directly downscaling GCM outputs to catchment scale streamflows. Hence, in this study MLR technique was used to downscale GCM outputs to catchment scale precipitation.

Following the methodology proposed by Sachindra et al. (2013), the probable predictors obtained from a reanalysis database were split into 20 year time slices, in the chronological order. The Pearson correlation coefficients (Pearson, 1895) between these probable predictors and the observed monthly precipitation were calculated for each 20 year time slice and the whole period, at all grid points in the atmospheric domain, for each calendar month. Thereafter, the probable variables which exhibited the best statistically significant correlations (at 95 % confidence level,  $p = 0.05$ ) with observed precipitation, over all 20 year time slices and the whole period consistently, were extracted as the potential predictors. The consistently correlated variables refer to the predictors which maintained correlations without any sign variations (e.g. positive to negative or vice versa) and large variation in magnitudes over the time slices and the whole period of the study. Once the selection of potential predictors was completed, two downscaling models were developed (calibrated/validated) separately, the first using the reanalysis outputs and the second with the corresponding 20<sup>th</sup> century climate experiment outputs of the GCM. The development of two separate downscaling models, one with reanalysis outputs and the other with GCM outputs, enabled the determination of how accurately the model developed with GCM outputs could reproduce the past precipitation observations, in comparison to its counterpart model. Furthermore, this process allows for understanding the potential of the downscaling model developed with GCM outputs, for its use in producing the precipitation projections into future. Reanalysis data are accepted to be more accurate than GCM outputs, owing to the rigorous quality control and corrective measures to which they are subjected to (e.g. NCEP/NCAR reanalysis - Kalnay et al., 1996). Since the reanalysis outputs are more accurate than the GCM outputs, the downscaling model built with reanalysis outputs should better perform in the calibration and validation periods. If the downscaling model developed with GCM outputs was capable of

reproducing the past precipitation observations adequately, it enables the use of this same model for the future projections of precipitation. In this case, a homogeneous set of data produced by the same GCM is used for the calibration, validation and future projection. Therefore, this can be regarded as a better option, than using the GCM outputs pertaining to future on the downscaling model developed with reanalysis outputs to project the precipitation at the station of interest into future.

For the calibration phase of the downscaling model developed with reanalysis data, the first two thirds of these reanalysis (corresponding to potential predictors) and observed precipitation data (predictand) were used, while the rest of the data were used for the model validation. The potential predictors for both calibration and validation were standardised with the means and the standard deviations of reanalysis data corresponding to the calibration phase (Sachindra et al., 2013). In model calibration, initially, the three potential predictors which have shown the best correlations with precipitation over the whole period of the study were introduced to the downscaling model. The parameters (coefficients and constants in the MLR equations) of the downscaling model were optimised in calibration, by minimising the sum of the squares of the errors. Then the model validation was performed with the calibrated model. The performance of the model during calibration and validation in reproducing the observed precipitation was assessed using the Nash-Sutcliffe efficiency (NSE) (Nash and Sutcliffe, 1970). Thereafter, the next potential predictors which showed the best correlation with precipitation were introduced to the previously added predictors of the downscaling model, one at a time. This process of stepwise addition of potential predictors was practised until the model performance in terms of NSE in validation reaches a maximum. This process allowed finding the best set of potential predictors and the best downscaling model for a calendar month. The downscaling model calibration and validation were performed for each calendar month separately.

If the stepwise development was not employed in the development of the model based on the reanalysis outputs, all potential predictors could have been introduced into the downscaling model at once. This could have introduced data redundancy errors due to the inter-dependency or cross-correlations between the predictors leading, to over-fitting in calibration and under-fitting in validation. The stepwise model

development and selection of the model which showed the best performance in validation guaranteed the avoidance of selection of models which showed over-fitting in calibration and under-fitting in validation.

As mentioned earlier in this paper, the second downscaling model (with sub-models for each calendar month) was developed (calibrated/validated) with the GCM outputs corresponding to the climate of the 20<sup>th</sup> century. In the calibration and validation of this downscaling model, observed precipitation at the station of interest was used as the predictand. The same calibration period used for first model was also used for this model. The rest of the GCM data were used for the validation. Inputs for both the calibration and validation phases of this model were standardised with the means and the standard deviations of the GCM outputs pertaining to the calibration period. The best potential predictors identified in the calibration and validation processes of the downscaling model developed with reanalysis outputs were also used in the development of this model, assuming the validity of these potential predictors for both downscaling models. The calibration of the second model was performed for each calendar month by introducing the 20<sup>th</sup> century climate outputs of the GCM pertaining to the best potential predictors. The optimum parameters of the MLR based downscaling models were determined by minimising the sum of the squared errors between the model predicted precipitation and the observed precipitation. These MLR models with the same parameters determined in the calibration phase were used in the validation. Unlike in the development of the model which was driven with reanalysis outputs, stepwise development process was not adopted in building the model driven with GCM outputs, as the best potential variables were already identified.

Graphical and numerical comparisons between the observed precipitation and precipitation outputs of the above described two statistical downscaling models were performed. Both graphical and numerical assessments were employed, as numerical assessments alone may not be robust enough in the evaluation of model performances. The graphical comparison of precipitation included the time series and scatter plots of the model reproduced precipitation against observations. The numerical assessment of the two downscaling models was done by statistical measures such as average, standard deviation, coefficient of variation, NSE, Seasonally Adjusted NSE (SANS) (Wang, 2006; Sachindra et al., 2013) and the

coefficient of determination ( $R^2$ ). Note that all MLR based downscaling models discussed in this paper were developed using the statistics toolbox in MATLAB (Version - R2008b).

#### **4. APPLICATION**

The generic methodology described in Section 3 was applied to the precipitation station at the Halls Gap post office in the operational area of GWMWater, Victoria, Australia.

##### **4.1. Atmospheric domain for downscaling**

There are no clear guidelines on the selection of the optimum size of the atmospheric domain for a statistical downscaling study. Najafi et al. (2011) successfully used an atmospheric domain with 7 X 4 grid points in the longitudinal and latitudinal directions respectively at a spatial resolution of 2.5° in both directions, for the statistical downscaling of outputs of CGCM3 to monthly precipitation. Their study demonstrated that the atmospheric domain does not necessarily have to be a square in shape. However, if the atmospheric domain is too rectangular in shape, the influences of large scale atmospheric circulations on the point of interest in the catchment are more considered on the wider sides of the domain, and the influences coming from the narrower sides are less considered or neglected. Hence, such domain shape should be avoided in statistical downscaling. A larger atmospheric domain increases the computational cost and time involved in the investigation. However, a larger domain aids in identifying influences of large scale atmospheric circulations over a wider area. When the atmospheric domain is too small, it may not be able to adequately capture the atmospheric circulations responsible for the hydroclimatology in the catchment. Therefore, the atmospheric domain which is an important component of any statistical downscaling study should be of adequate size and of an appropriate shape. In general a domain size of 6 X 6 grid points at a spatial resolution of 2.5° in both longitudinal and latitudinal directions is regarded as an adequate size (Tripathi et al., 2006). An atmospheric domain with spatial dimensions of 7 X 6 grid points at a spatial resolution of 2.5° in both longitudinal and latitudinal directions was selected for the downscaling study described in this paper. The size of this atmospheric domain was determined considering its ability to represent the large scale atmospheric phenomena which influence the precipitation at the point of interest and also the computational cost. The same atmospheric domain over

the same study area was successfully used by Sachindra et al. (2013) for statistically downscaling GCM outputs to catchment streamflows. The spatial resolution of this atmospheric domain was maintained at 2.5° in both longitudinal and latitudinal directions, making it compliant with the spatial resolution of the NCEP/NCAR reanalysis outputs. The atmospheric domain used in this study is shown in Figure 1. The shaded region in Figure 1 depicts the operational area of GWMWater, and the precipitation station considered in this study is located in its southmost region.

Figure 1 Atmospheric domain for downscaling

#### **4.2. Selection of probable and potential predictors for downscaling**

A pool of probable predictors was selected based on hydrology and past studies by Anandhi et al. (2008) and Timbal et al. (2009), on downscaling GCM outputs to precipitation. In the downscaling study by Timbal et al. (2009), predictor variables influential on the generation of precipitation, over the south and south eastern Australia (this includes the present study area) were identified. The probable predictor pool selected for the study described in this paper consisted of geopotential heights at 200hPa, 500hPa, 700hPa, 850hPa and 1000hPa pressure levels; relative humidity at 500hPa, 700hPa, 850hPa and 1000hPa pressure levels; specific humidity at 2m height, 500hPa, 850hPa and 1000hPa pressure levels; air temperatures at 2m height, 500hPa, 850hPa and 1000hPa pressure levels; surface skin temperature, surface pressure, mean sea level pressure, surface precipitation rate, and zonal and meridional wind speeds at 850hpa pressure level. These probable predictors were common for all calendar months. The monthly data for these 23 probable predictors for the 42 grid points shown in Figure 1 were extracted from the NCEP/NCAR reanalysis data archive at <http://www.esrl.noaa.gov/psd/>.

The probable predictors and the observed monthly precipitation totals from 1950 to 2010 were split into three 20 year time slices; 1950-1969, 1970-1989, and 1990-2010. The last time slice was 21 years in length. The Pearson correlation coefficients between the probable predictors and the observed monthly precipitation were calculated for all three time slices and the whole period (1950-2010), at each grid point in the atmospheric domain (see Figure 1). The probable predictors which showed good statistically

significant correlations (at 95 % confidence level,  $p = 0.05$ ) consistently over the three time slices and the whole period were selected as the potential predictors (Sachindra et al., 2013). This process was repeated for all 12 calendar months, yielding 12 sets of potential predictors.

The El Niño - Southern Oscillation (ENSO) and the Indian Ocean Dipole (IOD) are regarded as two large scale atmospheric phenomena influential on the climate of Victoria, Australia. A correlation analysis performed over the period 1950-2010 between the Southern Oscillation Index (SOI) which is representative of ENSO and observed precipitation at the Halls Gap post office indicated that these correlations vary between 0.03 (March) and 0.33 (October). Similarly, the correlations between the Dipole Mode Index (DMI) which is representative of IOD and observed precipitation ranged between -0.01 (February) and -0.46 (August) during the period 1958-2010. Hence, it was realised that the influences of these large scale atmospheric phenomena on the observed precipitation at the Halls Gap post office are weak in nature. Therefore it was understood that the inclusion of such indices in the inputs to the downscaling models will not lead to any improvement to the precipitation predictions. Furthermore, Chiew et al. (1998) detailed the influences of ENSO on the rainfall, drought and streamflows in Australia, using the SOI and Sea Surface Temperature (SST), and concluded that, the correlations between these ENSO indicators and hydroclimatic variables are not sufficiently strong for a consistent prediction.

### **4.3. MLR downscaling model calibration and validation**

#### *4.3.1 Model calibration and validation with NCEP/NCAR data*

The potential predictors selected from the probable pool were separated into two chronological groups; 1950 to 1989 and 1990 to 2010, the former for model calibration and the latter for model validation. The potential predictors were standardised for both calibration and validation periods using the means and the standard deviations pertaining to the period 1950 to 1989 (calibration period). The standardised potential variables were ranked based on the magnitude of their correlations with the observed monthly precipitation, over the whole period of the study (1950-2010). Then these potential variables were introduced to the MLR based downscaling model as described in Section 3. In the manner described in

Section 3, for each calendar month the best set of potential predictors and the best MLR based downscaling model were selected. Table 1 shows the final (or the best) set of potential predictors used in the downscaling model developed with NCEP/NCAR reanalysis outputs for the month of January. Also this table contains the correlations between the observed precipitation and the final set of potential predictors, during the three 20 year time slices and the whole period of the study.

Table 1 Final set of potential predictors used in the January downscaling model and their correlations with observed precipitation in each time slice and whole period.

Table 2 provides the final sets of potential predictors used in the downscaling models in each calendar month. The final sets of potential predictors used in the downscaling models consisted of: surface precipitation rate; specific humidity, relative humidity and geopotential heights at various pressure levels; mean sea-level pressure; surface pressure; and zonal and meridional wind speeds at 850hPa pressure level. However, surface precipitation rate was identified as the most influential potential predictor on precipitation, appearing in the final sets of potential predictors for all calendar months except July. Surface precipitation rate produced by GCMs is a precipitation flux (precipitation per unit time across unit area at earth surface) which is analogous to precipitation at a point over a specific time period (e.g. daily or monthly precipitation). Therefore the strong influence of surface precipitation rate on monthly precipitation was justified. The highest correlations between the NCEP/NCAR precipitation rate and the observed precipitation over the period 1950-2010 within each calendar month were; June (0.82), August (0.82), October (0.79), September (0.77), December (0.74), May (0.71), January (0.69), April (0.64), February (0.61), March (0.61) and November (0.48). Precipitation outputs of GCMs have been also used in the past downscaling studies. Timbal et al. (2009) used precipitation rate in downscaling daily precipitation and Tisseuil et al. (2010) used precipitation rate for downscaling daily streamflows. Maraun et al. (2013) stated that despite the errors, the precipitation output of a GCM can still contain useful information about the observed precipitation. Hence it was realised that precipitation output of a GCM can be used as an input to a downscaling model.

Specific humidity (mass of water vapour per unit mass of air), and relative humidity (ratio of actual water vapour pressure of the air to the saturation vapour pressure) at various pressure levels are indicators of the atmospheric water vapour content which leads to the formation of clouds (Peixoto and Oort, 1996). Humidity variables (relative or specific humidity) which are indicators of the atmospheric water vapour content were potential predictors in 7 (February, March, May, September, October, November and December) of the 12 calendar months. According to Nazemosadat and Cordery (1997), geopotential heights are influential on the generation of precipitation, as they are representative of large scale atmospheric pressure variations such as the El Niño Southern Oscillation (ENSO). Zonal and meridional wind fields are influential on the evaporation from open surface water bodies and they govern the movement of rain bearing clouds (Bureau of Meteorology, 2010), and hence it was suitable to include wind fields in the final sets of potential predictors. It is noteworthy to mention that, according to Table 2, except in August and November, grid point {4,4} found to be a dominant location for the final sets of potential predictors. The grid point {4,4} is the closest grid point to the precipitation station considered in this study.

In general, humidity variables and precipitation rate are more capable of explaining the precipitation process (refer to Table 2). However as shown in Table 2, in the month of July, the set of potential predictors used in the downscaling models contained only the wind speeds and the geopotential heights at 850hPa. It was realised that these variables are still able to explain the precipitation process with a good degree of accuracy, as the downscaling model developed for July using the NCEP/NCAR reanalysis outputs displayed NSEs of 0.58 and 0.50 in the calibration and validation phases respectively. Furthermore, as these potential variables are selected based on the magnitude and also the consistency of correlations with observed precipitation over time, it is argued that the final sets of potential predictors used in the downscaling models are able to characterise the changes in precipitation at the point of interest, also in the future.

Table 2 Final sets of potential predictors for each calendar month



In Table 2, it could be found that the majority of the potential predictors in the final sets were selected from the grid points surrounding the precipitation station of interest [(3,3), (3,4), (3,5), (4,3), (4,4), (4,5), (5,3), (5,4) and (5,5)]. However some potential predictors in the final sets were selected from the distant grid points of the domain as the precipitation at the station of interest is not only influenced by the atmosphere in close proximity to the station but also by the atmospheric processes that occur far away. The best grid locations of the potential predictors provided in Table 2 were selected not only based on the strength of the correlation between the potential predictors and observed precipitation, but also considering the consistency of the correlation over three time slices and the whole period of the study. Therefore it was assumed that the best grid locations of the final sets of potential predictors used in this study will remain the same in future.

#### *4.3.2 Model calibration and validation with HadCM3 20<sup>th</sup> century climate experiment data*

The 20<sup>th</sup> century climate experiment data of HadCM3 GCM were obtained for the period 1950-1999, corresponding to the final sets of potential predictors shown in Table 2. HadCM3 model has been forced with both natural and anthropogenic forcings to reproduce the climate of the 20<sup>th</sup> century (Knight, 2003). As the natural forcings; sea-surface temperature (SST) and sea-ice anomalies, variations in the total solar irradiance and stratospheric volcanic aerosols etc have been used in HadCM3. As anthropogenic forcings; GHG concentrations in the atmosphere, changes in tropospheric and stratospheric ozone, the effects of atmospheric sulphate aerosols and changes in land surface characteristics have been used in HadCM3. The 20<sup>th</sup> century climate experiment data of HadCM3 were split into two groups; (a) 1950-1989 for model calibration and (b) 1990-1999 for the model validation. HadCM3 data for both the calibration and validation phases were standardised with the means and the standard deviations of HadCM3 data corresponding to 1950-1989 period. In calibration, the standardised sets of data pertaining to the best potential predictors shown in Table 2 were introduced to the MLR based downscaling model. During calibration, the optimum values for the model parameters were determined by minimising the sum of squared errors between the model predicted and observed precipitation values. In validation, the HadCM3 data for the 1990-1999 period were introduced to the calibrated MLR models. The same procedure was repeated for all calendar months. Unlike in the calibration and validation of the downscaling model which

was developed with NCEP/NCAR reanalysis outputs, the stepwise development procedure was not adopted in these models. A correlation coefficient analysis performed between the 20<sup>th</sup> century climate experiment outputs of HadCM3 and NCEP/NCAR reanalysis outputs over the period 1950-1999, revealed that these correlations are quite weak (e.g. 0.2 – 0.4). Hence it was realised that HadCM3 outputs pertaining to the 20<sup>th</sup> century climate experiment contain large bias. Therefore it was understood that whether the final sets of potential predictors are selected using a stepwise procedure or not, they will not change the performance of the model developed with HadCM3 outputs. It was assumed that final sets of potential predictors identified in the development of the model driven with NCEP/NCAR outputs are also applicable for this model. The difference between the statistical downscaling models built with the HadCM3 20<sup>th</sup> century experiment data ( $\text{Model}_{(\text{HadCM3})}$ ) and the models built with the NCEP/NCAR reanalysis data ( $\text{Model}_{(\text{NCEP/NCAR})}$ ) was that these two models had different optimum values for their parameters (coefficients and constants in MLR equations).

#### *4.3.3 Calibration and validation results of the downscaling models*

Figure 2 shows the time series of monthly observed precipitation and monthly precipitation reproduced by the downscaling model developed with NCEP/NCAR data, for the period 1950-2010. According to Figure 2, the monthly precipitation reproduced by this downscaling model, was in close agreement with the observed precipitation during both calibration and validation periods. Although the model validation was performed in a relatively dry period which included the Millennium drought (1997-2010), this downscaling model has been able to capture the monthly precipitation pattern and the magnitude with good accuracy.

Figure 2 Observed and  $\text{Model}_{(\text{NCEP/NCAR})}$  reproduced monthly precipitation (1950 to 2010)

Figure 3 shows the scatter plots of monthly observed precipitation and precipitation reproduced by the downscaling model developed with NCEP/NCAR data, for the calibration (1950-1989) and validation (1990-2010) phases. As seen in Figure 3, during both the calibration and validation periods, near zero monthly precipitation values were over predicted and relatively large precipitation values were under

predicted. However, these scatter plots of the model predictions against the observations further confirmed that, the prediction capabilities of the model developed with NCEP/NCAR data in validation are very much comparable with those during calibration.

Figure 3 Scatter plots of observed and Model<sub>(NCEP/NCAR)</sub> reproduced monthly precipitation for calibration (1950-1989) and validation (1990-2010)

Figure 4 illustrates the time series of monthly observed precipitation and monthly precipitation reproduced by the downscaling model built with HadCM3 data, for the period 1950-1999. It was seen that this model was not able to satisfactorily reproduce the high precipitation values. Furthermore, the agreement between the observed and model reproduced precipitation was much less compared to that of the model developed with NCEP/NCAR reanalysis outputs. However the model developed with HadCM3 outputs properly captured the pattern of the observed precipitation as shown in Figure 4. It should be noted that the validation phase of the model developed with HadCM3 data was confined to the period 1990-1999, due to the unavailability of data beyond year 1999, under the 20<sup>th</sup> century climate experiment.

Figure 4 Observed and Model<sub>(HadCM3)</sub> reproduced monthly precipitation (1950 to 1999)

Figure 5 represents the scatter plots for the calibration (1950-1989) and validation (1990-1999) phases of the downscaling model developed with HadCM3 data. It was seen that in the calibration and validation phases, high precipitation values were largely under-predicted. During both phases, the model displayed a clear trend of over-predicting the majority of low precipitation values. However, these characteristics were also seen in the predictions of the model developed with NCEP/NCAR data, but with less intensity.

Statistical downscaling models in general fail to capture the full range of the variance of a predictand such as precipitation (Wilby et al., 2004). This is because, in general the variance in the observations of precipitation is much greater than the variance in the large scale atmospheric variables obtained from the GCM or the reanalysis data. When the downscaling model is run with the GCM or the reanalysis data it tends to explain the mid range of the variance of the observed precipitation better than the low and high

extremes. Therefore statistical downscaling models in general tend to reproduce the average of the precipitation better than the low and high extremes. In other words, this results in an under-estimation of large precipitation values and over-estimations of near zero precipitation values. Tripathi et al. (2006) also commented that even a downscaling model based on support vector machine technique (complex non-linear regression technique) fails to properly reproduce the extremes of precipitation though it captures the average well.

Figure 5 Scatter plots of observed and Model<sub>(HadCM3)</sub> reproduced monthly precipitation for calibration (1950-1989) and validation (1990-1999)

The performances of the two downscaling models, during the calibration and validation phases were numerically assessed by comparing the mean, the standard deviation and the coefficient of variation of the model predictions with those of observations, and these results are shown in Table 3. It can be seen that both downscaling models developed with NCEP/NCAR and HadCM3 outputs reproduced the observed averages of the precipitation with good accuracy, in both calibration and validation phases. This finding was quite consistent with that of Sachindra et al. (2013), in which MLR and LS-SVM techniques were employed for downscaling NCEP/NCAR outputs to streamflows. However, in the present study, neither of the two models properly captured the standard deviation and the coefficient of variation of the observed precipitation, during both the calibration and validation phases. This characteristic was more noticeable in the outputs of the downscaling model developed with HadCM3 data. It indicated that, in particular, the model developed with HadCM3 data could not reproduce the entire variance of the observed precipitation. In Figure 4, the same characteristic was seen in the time series plots. This characteristic was seen with less severity in the outputs of the model developed with NCEP/NCAR reanalysis data.

Table 3 Performances of downscaling models in calibration and validation

The model performances in calibration and validation were further quantified with the NSE, the SANS and the coefficient of determination ( $R^2$ ). The SANS considers the seasonal means of precipitation in

measuring the model performances, unlike the original NSE, which considers only the mean of precipitation for the whole period. During calibration, the statistical downscaling model developed with NCEP/NCAR reanalysis data displayed NSE, SANS and  $R^2$  of 0.74, 0.66 and 0.74 respectively. However for the same period, the downscaling model developed with HadCM3 outputs, produced NSE, SANS and  $R^2$  of 0.44, 0.26 and 0.44 respectively. In the validation phase, the model developed with NCEP/NCAR outputs produced NSE, SANS and  $R^2$  of 0.70, 0.61 and 0.72. During the validation period, the model developed with HadCM3 outputs, produced NSE, SANS and  $R^2$  of 0.17, -0.20 and 0.22 respectively. These findings indicated that both downscaling models have performed relatively better during the calibration period than in the validation period. However, it was seen that the downscaling model developed with NCEP/NCAR data performed well in the calibration and validation phases, compared to its counterpart model which was built with HadCM3 outputs. This statement was further supported by the findings of scatter plots shown in Figures 3 and 5.

Figure 6 depicts the agreement between the precipitation reproduced by the model developed with NCEP/NCAR outputs and the observed precipitation, during the calibration (1950-1989) and validation (1990-2010) periods, on a seasonal basis. As shown in Figure 6, it was determined that this model demonstrates good capabilities in reproducing the observations in calibration and validation, in all four seasons, despite the tendencies of under-predicting high precipitation values and over-predicting near zero precipitation values which were evident in all four seasons. The four seasons are defined as summer (December-February), autumn (March-May), winter (June-August) and spring (September-November).

Figure 6 Seasonal scatter plots of observed and Model<sub>(NCEP/NCAR)</sub> reproduced monthly precipitation for calibration (1950-1989) and validation (1990-2010)

Figure 7 displays the seasonal scatter plots for the calibration (1950-1989) and validation (1990-1999) periods of the model developed with HadCM3 outputs. Large under-predictions of precipitation were seen in all four seasons during both the calibration and validation phases of this model. During all four seasons in the validation period, a relatively poor agreement between the observed and model reproduced precipitation was seen. This characteristic was more intense in autumn, winter and spring than in summer.

Figure 7 Seasonal scatter plots of observed and Model<sub>(HadCM3)</sub> reproduced monthly precipitation for calibration (1950-1989) and validation (1990-1999)

Table 4 shows the seasonal statistics of the observed precipitation and the precipitation reproduced by the models developed with NCEP/NCAR reanalysis and HadCM3 data, for the calibration and validation periods. In the calibration phase, during all four seasons, averages of the observed precipitation were near perfectly reproduced by both downscaling models. In the validation phase, although not as good as in calibration, both models were capable in reproducing the averages of observed precipitation in all four seasons with some under and over-predictions. During all four seasons in the validation period, both downscaling models tended to over-predict the average of the observed precipitation. This was due to the fact that the calibration was performed over a wetter period and the validation was done during a relatively dryer period. However, according to Figures 2 and 4 both downscaling models were able to adequately capture the precipitation pattern seen in the observations, throughout the calibration and validation periods. The under-estimation of the standard deviation and the coefficient of variation was seen in all four seasons of both models, during the calibration and validation periods. This characteristic was more severe in the case of the model developed with HadCM3 outputs. Since there is a large scale gap between the GCM outputs and the catchment scale, not all the variance in observations of a predictand (at a point in the catchment) can be explained by the GCM. Therefore, regression based statistical downscaling techniques are capable of capturing only the part of the variance (deterministic component of the variance) of a predictand which is conditioned by the GCM (Hewitson et al., 2013). The local scale random variance of the predictand (stochastic component of the variance) is not simulated by the regression based downscaling models, as it is not explicitly explained by the GCM. At the catchment scale, capturing the full variance of a predictand is important. This can be achieved by the application of a suitable bias-correction method for post processing the outputs of the downscaling model (Maraun, 2013). Techniques such as randomization may also help in capturing the full variance of a predictand (von Storch, 1999).

In the model developed with NCEP/NCAR data, the best performances in calibration in terms of NSE and  $R^2$  were seen during winter while the lowest performances were observed in summer. For this model, in validation, autumn produced the best performance. The model developed with HadCM3 outputs showed relatively low NSE and  $R^2$  in all four seasons of the calibration period. The negative NSEs were seen in autumn, winter and spring during the validation period, which indicated the limited performances of this downscaling model.

As mentioned in Section 1, the largest drop in precipitation over Victoria during the Millennium drought was observed in autumn. The decline in the average of the observed precipitation in autumn, during the Millennium drought (1997-2010), at the station considered in this study, was 27.5%, from the long-term average (1950-1989). The downscaling model developed with NCEP/NCAR reanalysis outputs was able to successfully reproduce this large drop in the average as 22.4%.

Table 4 Seasonal performances of downscaling models

According to the findings discussed previously, it was realised that the downscaling model developed with NCEP/NCAR reanalysis data has better potential in downscaling precipitation, in comparison with its counterpart model built with HadCM3 outputs. This was due to the better quality of NCEP/NCAR reanalysis outputs characterised by better synchronicity with observed precipitation, high precipitation simulation etc in comparison to those of HadCM3 outputs. Furthermore, it was seen that MLR has the potential for modelling the relationship between the predictors and the monthly precipitation adequately. As shown in Tables 3 and 4, and Figure 3 with the final sets of potential variables given in Table 2, the downscaling model developed with NCEP/NACR reanalysis outputs reproduced the observed precipitation with good degree of accuracy. Therefore it was realised that the final sets of potential variables used in the downscaling models are capable of capturing the precipitation process to a good degree.

Figure 8 shows the exceedance probability curve for the observed precipitation, precipitation reproduced by the downscaling models with NCEP/NCAR and HadCM3 outputs, and the raw precipitation output of

HadCM3 model for the 20<sup>th</sup> century climate experiment at grid point {4,4} (see Figure 1 for location), over the period 1950-1999. Since point {4,4} is the closest grid point to the precipitation station, HadCM3 20<sup>th</sup> century climate experiment outputs at this point was considered to be representative of the precipitation station considered in this study. Note that the precipitation rate (which was the observed precipitation equivalent output of HadCM3) was converted to monthly precipitation, for plotting the corresponding exceedance curve in Figure 8.

Figure 8 Precipitation probability exceedance curves (1950 to 1999)

According to Figure 8 it was seen that there is a large mismatch between the raw precipitation output at grid point {4,4} of HadCM3 model and the observed precipitation, during the period 1950-1999. The large bias in the precipitation output of HadCM3 indicated that its regional precipitation simulation is less reliable. Larger differences between the observations and raw HadCM3 precipitation outputs were seen for precipitations with low probability of exceedance, such as extremely high precipitations. Furthermore, relatively small anomalies were seen for precipitation values with low magnitudes. For the majority of exceedance probabilities, this mismatch was seen as a large under-prediction in HadCM3 precipitation outputs. The mismatch between the observations and the raw HadCM3 precipitation output was mainly due to the bias present in HadCM3 outputs. As defined by Salvi et al. (2011), bias is the difference between the GCM outputs and the pertaining observations. GCM bias is a result of the limited knowledge of the atmospheric processes and the simplified representation of the complex climate system in GCMs (Li et al., 2010). The other possible factor contributing to the poor agreement between observations and HadCM3 outputs is, that grid point {4,4} may not exactly represent the precipitation at the station considered in this study. Furthermore, in case of the precipitation gauge located at the Halls Gap post office, topographical reasons also have possibly contributed to the bias in the GCM outputs, as Halls Gap is located in a valley surrounded by a mountain range.

It was noted that the mismatch between the observations and the precipitation downscaled with HadCM3 outputs was less in comparison with that between the observations and the raw precipitation outputs of HadCM3 at grid point {4,4}. This indicated that when the raw outputs of HadCM3 are statistically



downscaled to monthly precipitation, the impact of bias in these raw HadCM3 outputs, on downscaled precipitation was less evident. However, this reduction in bias was not adequate as still there was considerable mismatch between the observed and downscaled precipitation (refer to Figure 8). Therefore, it could be argued that a correction to the bias that is present in HadCM3 outputs is needed in producing precipitation projections into future. It was seen that the precipitation exceedance curve of raw precipitation output of HadCM3 at grid point {4,4} had deviated largely from the precipitation exceedance curve of observations. However, the exceedance curves of precipitation reproduced by the downscaling models developed with NCEP/NCAR reanalysis outputs and HadCM3 outputs were in relatively better agreement with the precipitation exceedance curve of observed precipitation. This led to the conclusion that, the precipitation outputs of the downscaling models developed with NCEP/NCAR reanalysis outputs and HadCM3 outputs are much better than the raw precipitation output of HadCM3 at grid point {4,4}. Furthermore, considering the limited agreement seen between the precipitation downscaled with the NCEP/NCAR and HadCM3 outputs, it was realised that there is a quality mismatch between the data of these two sources. The second paper of this series of two companion papers, describes the bias correction and the precipitation projections produced into future in detail.

## **5. SUMMARY AND CONCLUSIONS**

This paper, which is the first of a series of two companion papers, discussed the development (calibration and validation) of two precipitation downscaling models, employing the multi-linear regression (MLR) technique. The first statistical downscaling model was developed with the NCEP/NCAR reanalysis outputs and the second downscaling model was developed with the HadCM3 outputs. The precipitation station at the Halls Gap post office which is located in the north western part of Victoria, Australia was selected for the demonstration of the development process of the two downscaling models.

It is the general practice to calibrate and validate the downscaling model with some form of reanalysis data (e.g. NCEP/NCAR) for the past climate, and use the outputs of a GCM pertaining to future on the same downscaling model for the projection of climate into future. The major disadvantage of this

procedure is that, for the model development and future projections, data from two entirely different sources are used. This study investigated the potential of using a downscaling model calibrated and validated with GCM outputs, which does not have the above issue.

The selection of probable predictors for these downscaling models was based on the past statistical downscaling studies and hydrology. Potential predictors were extracted for each calendar month from the set of probable predictors considering the Pearson correlations between the probable predictors and observed precipitation, under three 20 year time slices (1950-1969, 1970-1989, and 1990-2010) and the entire period of the study (1950-2010). Potential predictors obtained from the NCEP/NCAR reanalysis outputs were introduced to the MLR based downscaling model, sequentially, based on the magnitude of the correlation between observed precipitation and predictors, over the whole period of the study. This process was continued until the model performances in validation in terms of NSE was maximised. In this manner, the final sets of potential predictors for each calendar month were identified, and downscaling models for each calendar month were developed separately. The HadCM3 outputs corresponding to the final sets of potential predictors identified previously were used for the development of the second downscaling model. It was assumed that these final sets of potential predictors are valid for both downscaling models, developed with NCEP/NCAR and HadCM3 outputs.

The MLR based downscaling model developed with NCEP/NCAR reanalysis outputs proved capable in reproducing the observed monthly precipitation during both calibration (1950-1989) and validation (1990-2010) phases. The performances of this model in calibration were slightly better than those in validation. This model was also able to capture the precipitation drop occurred during the Millennium drought (1997-2010) satisfactorily. However, it displayed tendencies of over-predicting low precipitation values and under-predicting high precipitation values during both the calibration and validation periods.

On the other hand, the MLR based downscaling model developed with HadCM3 outputs displayed limited performances with respect to the model developed with NCEP/NCAR reanalysis outputs during both calibration and validation stages. This model performed better during calibration (1950-1989) than in validation (1990-1999). Similar to the model developed with NCEP/NCAR reanalysis outputs, this

downscaling model also displayed tendencies of over-predicting and under-predicting low and high precipitation values respectively. However, the over and under-predictions associated with the model developed with HadCM3 outputs were much severe than those for its counterpart downscaling model. Due to the termination of HadCM3 outputs at 1999 for the 20<sup>th</sup> century climate experiment, the validation phase of this downscaling model was confined to 1990-1999. Therefore it was not possible to see how this downscaling model will reproduce the precipitation during the Millennium drought.

The conclusions drawn from this study are:

- 1) The precipitation rate which is the precipitation equivalent output of a GCM was found as the most influential predictor on precipitation at the station of interest, over the entire year, except in July;
- 2) Humidity, geopotential heights, mean sea level and surface pressure, and wind speeds also showed good correlations with observed precipitation consistently over time;
- 3) The downscaling model developed with NCEP/NCAR reanalysis outputs performed well in both calibration and validation, while the performances of the model developed with HadCM3 outputs were limited;
- 4) There was a quality mismatch between the NCEP/NCAR reanalysis and HadCM3 outputs, over the period 1950-1999; and
- 5) A bias-correction should be applied in projecting the precipitation into future at the station of interest.

The application of the bias-correction and the projections of precipitation into future are presented in the second companion paper of this series of papers.

## ACKNOWLEDGEMENTS

The authors acknowledge the financial assistance provided by the Australian Research Council Linkage Grant scheme and the Grampians Wimmera Mallee Water Corporation for this project. The authors also wish to thank the editor and the two anonymous reviewers for their useful comments, which have improved the quality of this paper.

## REFERENCES

Alavian V, Qaddumi HM, Dickson E, Diez SM, Danilenko AV, Hirji RF, Puz G, Pizarro C, Jacobsen M, Blankespoor B. 2009. Water and climate change: understanding the risks and making climate-smart investment decisions. Washington D.C. - The Worldbank. Available online at <http://documents.worldbank.org/curated/en/2009/11/11717870/water-climate-change-understanding-risks-making-climate-smart-investment-decisions>. (Accessed on 19<sup>th</sup> July 2013).

Anandhi A, Srinivas VV, Nanjundiah RS, Kumar DN. 2008. Downscaling precipitation to river basin in India for IPCC SRES scenarios using support vector machine. *International Journal of Climatology* **28**: 401-420. DOI:10.1002/joc.1529.

Anandhi, A. 2010. Assessing impact of climate change on season length in Karnataka for IPCC SRES scenarios. *Journal of Earth System Science* **119**: 447-460. DOI:10.1007/s12040-010-0034-5.

Anandhi A, Srinivas VV, Nanjundiah RS, Kumar DN. 2012. Daily relative humidity projections in an Indian river basin for IPCC SRES scenarios. *Theoretical and Applied Climatology* **108**: 85-104. DOI:10.1007/s00704-011-0511-z.

Bureau of Meteorology. 2010. Australian climate influences. Available at <http://www.bom.gov.au/watl/about-weather-and-climate/australian-climate-influences.shtml> (Accessed 25<sup>th</sup> September 2012).

Charles A, Timbal B, Fernandez E, Hendon H. 2013. Analog downscaling of seasonal rainfall forecasts in the Murray darling basin. *Monthly Weather Review* **141**: 1099-1117. DOI:10.1175/MWR-D-12-00098.1.

Chiew FHS, Piechota TC, Dracup JA, McMahon TA. 1998. El Nino/Southern Oscillation and Australian rainfall, streamflow and drought: Links and potential for forecasting. *Journal of Hydrology* **204**: 138-149. DOI:10.1016/S0022-1694(97)00121-2.

Chiew FHS, Young WJ, Cai W, Teng J. 2010. Current drought and future hydroclimate projections in southeast Australia and implications for water resources management. *Stochastic Environmental Research and Risk Assessment* **25**: 602-612. DOI:10.1007/s00477-010-0424-x.

Chu JT, Xia J, Xu CY, Singh VP. 2010. Statistical downscaling of daily mean temperature, pan evaporation and precipitation for climate change scenarios in Haihe River, China. *Theoretical and Applied Climatology* **99**: 149-161. DOI:10.1007/s00704-009-0129-6.

Crowley TJ. 2000. Causes of climate change over the past 1000 years. *Science* **289**: 270-277. DOI:10.1126/science.289.5477.270.

Government of Western Australia Department of Water. 2009. Streamflow trends in south-west Western Australia. Surface water hydrology series. Report no. HY32. Available at <http://www.water.wa.gov.au/PublicationStore/first/87846.pdf>. (Accessed 1<sup>st</sup> November 2012).

Goyal MK, Burn DH, Ojha CSP. 2012. Evaluation of machine learning tools as a statistical downscaling tool: temperatures projections for multi-stations for Thames river basin, Canada. *Theoretical and Applied Climatology* **108**: 519-534. DOI:10.1007/s00704-011-0546-1.

Hashmi MZ, Shamseldin AY, Melville BW. 2011. Statistical downscaling of watershed precipitation using Gene Expression Programming (GEP). *Environmental Modeling and Software* **26**: 1639-1646. DOI:10.1016/j.envsoft.2011.07.007.

Hay LE, Clark MP. 2003. Use of statistically and dynamically downscaled atmospheric model output for hydrologic simulations in three mountainous basins in the western United States. *Journal of Hydrology* **282**: 56-75. DOI:10.1016/S0022-1694(03)00252-X.

Hewitson B, Jack C, Coop L. 2013. Addressing deterministic and stochastic variance in statistical downscaling. *European Geophysical Union General Assembly 2013*. 7-12 April 2013, Vienna, Austria.

Iizumi T, Takayabu I, Dairaku K, Kusaka H, Nishimori M, Sakurai G, Ishizak, NN, Adachi SA, Semenov MA. 2012. Future change of daily precipitation indices in Japan: A stochastic weather generator-based bootstrap approach to provide probabilistic climate information. *Journal of Geophysical Research D: Atmospheres* **117**: D11114. DOI:10.1029/2011JD017197.

IPCC. 2007. IPCC Fourth assessment: synthesis report - Summary for policymakers. 6-9. Available at [http://www.ipcc.ch/pdf/assessment-report/ar4/syr/ar4\\_syr\\_spm.pdf](http://www.ipcc.ch/pdf/assessment-report/ar4/syr/ar4_syr_spm.pdf). (Accessed on 8<sup>th</sup> November 2012).

Jeffrey SJ, Carter JO, Moodie KB, Beswick AR. 2001. Using spatial interpolation to construct a comprehensive archive of Australian climate data. *Environmental Modelling and Software* **16**: 309-330. DOI:10.1016/S1364-8152(01)00008-1.

Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, Iredell M, Saha S, White G, Woollen J, Zhu Y, Chelliah M, Ebisuzaki W, Higgins W, Janowiak J, Mo KC, Ropelewski C, Wang J, Leetmaa A, Reynolds R, Jenne R, Joseph D. 1996. The NCEP/NCAR reanalysis project. *Bulletin of the American Meteorological Society* **77**: 437-471. DOI:10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

Kannan S, Ghosh S. 2013. A nonparametric kernel regression model for downscaling multisite daily precipitation in the Mahanadi basin. *Water Resources Research* **49**: 1360-1385. DOI:10.1002/wrcr.20118.

Karl TR, Wang WC, Schlesinger ME, Knight RW, Portman D. 1990. A method of relating general circulation model simulated climate to the observed local climate Part I: seasonal statistics. *Journal of Climate* **3**: 1053-1079. DOI:10.1175/1520-0442(1990)003<1053:AMORGC>2.0.CO;2.

Khalili M, Brissette F, Leconte R. 2009. Stochastic multi-site generation of daily weather data. *Stochastic Environmental Research and Risk Assessment* **23**: 837-849. DOI:10.1007/s00477-008-0275-x.

Khazaei MR, Ahmadi S, Saghafian B, Zahabiyoun B. 2013. A new daily weather generator to preserve extremes and low-frequency variability. *Climatic Change*. (Article in press) DOI:10.1007/s10584-013-0740-5.

Knight J. 2003. Report on forcings for the C20C and EMULATE HadAM3 experiments. Available at <http://hadc20c.metoffice.com/forcings.pdf> (Accessed on 30<sup>th</sup> May 2013).

Li H, Sheffield J, Wood EF. 2010. Bias correction of monthly precipitation and temperature fields from Intergovernmental Panel on Climate Change AR4 models using equidistant quantile matching. *Journal of Geophysical Research D: Atmospheres* **115**: 1-20. DOI:10.1029/2009JD012882.

Maraun D, Wetterhall F, Ireson, AM, Chandler RE, Kendon EJ, Widmann M, Brienen S, Rust HW, Sauter T, Themel M, Venema VKC, Chun KP, Goodess CM, Jones RG, Onof C, Vrac M, Thiele-Eich I. 2010. Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Reviews of Geophysics* **48**: DOI:10.1029/2009RG000314.

Maraun D. 2013. Bias correction, quantile mapping, and downscaling: Revisiting the inflation issue. *Journal of Climate* **26**: 2137–2143. DOI:10.1175/JCLI-D-12-00821.1.

Marzban C, Sandgathe S, Kalnay E. 2006. MOS, perfect prog, and reanalysis. *Monthly Weather Review* **134**: 657-663. DOI:10.1175/MWR3088.1.

Maurer EP, Hidalgo HG. 2008. Utility of daily vs. monthly large-scale climate data: an intercomparison of two statistical downscaling methods. *Hydrology and Earth System Science* **12**: 551-563. DOI:10.5194/hess-12-551-2008.

Najafi M, Moradkhani H, Wherry S. 2011. Statistical downscaling of precipitation using machine learning with optimal predictor selection. *Journal of Hydrologic Engineering* **16**: 650-664. DOI:10.1061/(ASCE)HE.1943-5584.0000355.

Nash JE, Sutcliffe JV. 1970. River flow forecasting through conceptual models, part 1 - A discussion of principles. *Journal of Hydrology* **10**: 282-290. DOI:10.1016/0022-1694(70)90255-6.

Nasseri M, Tavakol-Davani H, Zahraie B. 2013. Performance assessment of different data mining methods in statistical downscaling of daily precipitation. *Journal of Hydrology* **492**: 1-14. DOI:10.1016/j.jhydrol.2013.04.017.

Nazemosadat MJ, Cordery I. 1997. The influence of geopotential heights on New South Wales rainfall. *Meteorology and Atmospheric Physics* **63**: 179-193. DOI:10.1007/BF01027384.

Pearson K. 1895. Mathematical contributions to the theory of evolution. iii. regression heredity and panmixia. *Philosophical Transactions of the Royal Society of London Series A* **187**: 253-318. DOI:10.1098/rsta.1896.0007.

Peixoto JP, Oort AH. 1996. The climatology of relative humidity in the atmosphere. *Journal of Climate* **9**: 3443-3463. DOI:10.1175/1520-0442(1996)009<3443:TCORHI>2.0.CO;2.

Richardson CW. 1981. Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water Resources Research* **17**: 182-190. DOI:10.1029/WR017i001p00182.



Sachindra DA, Huang F, Barton AF, Perera BJC. 2012. Issues associated with statistical downscaling of general circulation model outputs: a discussion. *Proceedings of Practical Responses to Climate Change National Conference*. 1-3 May 2012, Canberra, Australia.

Sachindra DA, Huang F, Barton AF, Perera BJC. 2013. Least square support vector and multi-linear regression for statistically downscaling general circulation model outputs to catchment streamflows. *International Journal of Climatology* **33**: 1087-1106. DOI:10.1002/joc.3493.

Salvi K, Kannan S, Ghosh S. 2011. Statistical downscaling and bias-correction for projections of Indian rainfall and temperature in climate change studies. *4<sup>th</sup> International Conference on Environmental and Computer Science*. 16 - 18 September 2011, Singapore, 7-11.

Schnur R, Lettenmaier DP. 1998. A case study of statistical downscaling in Australia using weather classification by recursive partitioning. *Journal of Hydrology* **213**: 362-379. DOI:10.1016/S0022-1694(98)00217-0.

Semenov MA, Stratonovitch P. 2010. Use of multi-model ensembles from global climate models for assessment of climate change impacts. *Climate Research* **41**: 1-14. DOI:10.3354/cr00836.

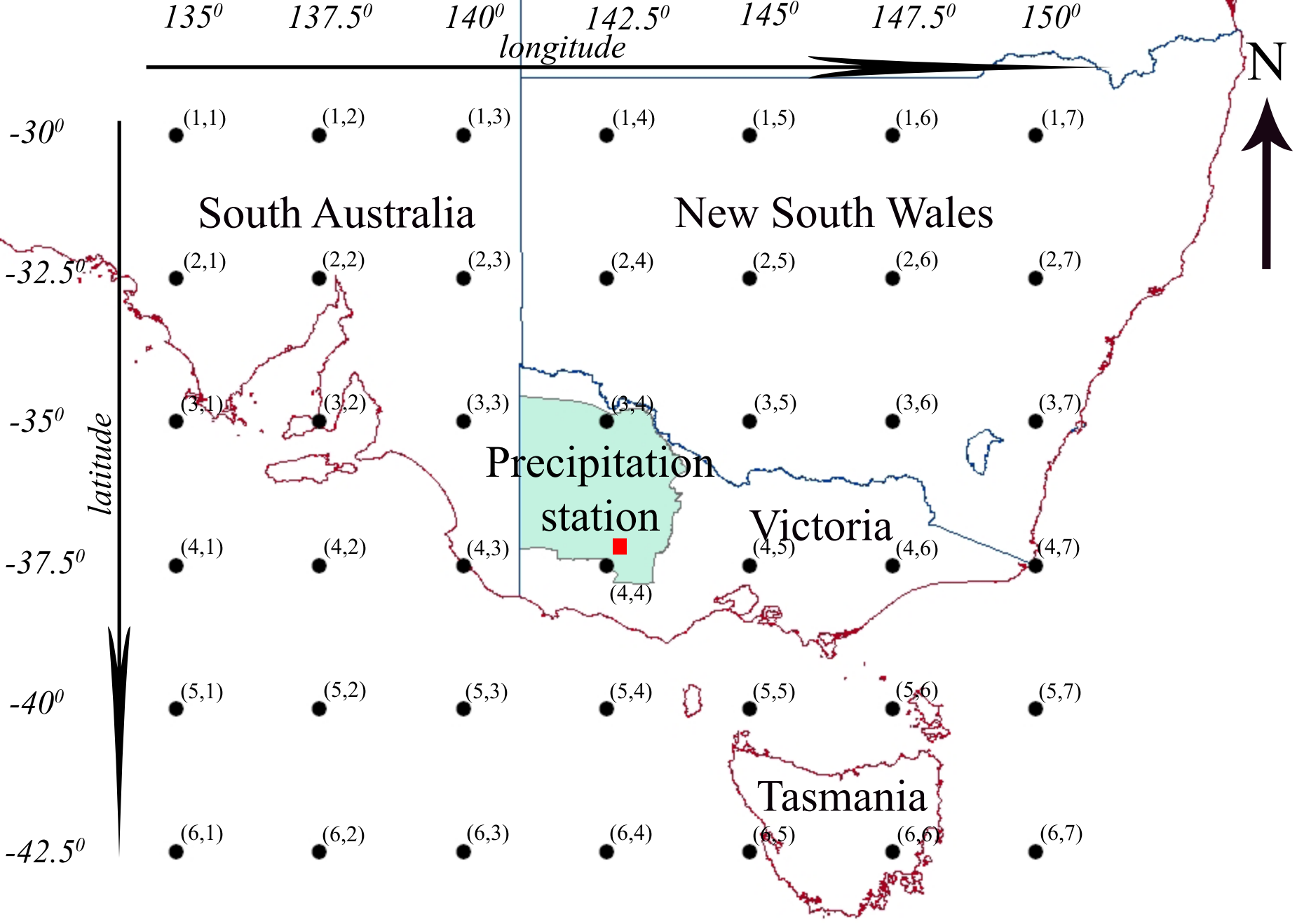
Shao Q, Li M. 2013. An improved statistical analogue downscaling procedure for seasonal precipitation forecast. *Stochastic Environmental Research and Risk Assessment* **27**: 819-830. DOI:10.1007/s00477-012-0610-0.

Smith IN, McIntosh P, Ansell TJ, Reason CJC, McInnes K. 2000. Southwest Western Australian winter rainfall and its association with Indian Ocean climate variability. *International Journal of Climatology* **20**: 1913-1930. DOI:10.1002/1097-0088(200012)20:15<1913::AID-JOC594>3.0.CO;2-J.

- Timbal B, Fernandez E, Li Z. 2009. Generalization of a statistical downscaling model to provide local climate change projections for Australia. *Environmental Modeling and Software* **24**: 341-358. DOI:10.1016/j.envsoft.2008.07.007.
- Timbal B, Jones DA. 2008. Future projections of winter rainfall in southeast Australia using a statistical downscaling technique. *Climatic Change* **86**: 165-187. DOI:10.1007/s10584-007-9279-7.
- Tisseuil C, Vrac M, Lek S, Wade AJ. 2010. Statistical downscaling of river flows. *Journal of Hydrology* **385**: 279-291. DOI:10.1016/j.jhydrol.2010.02.030.
- Tripathi S, Srinivas VV, Nanjundiah RS. 2006. Downscaling of precipitation for climate change scenarios: a support vector machine approach. *Journal of Hydrology* **330**: 621-640. DOI:10.1016/j.jhydrol.2006.04.030.
- von Storch H. 1999. On the use of “inflation” in statistical downscaling. *Journal of Climate* **12**: 3505–3506. DOI:10.1175/1520-0442(1999)012<3505:OTUOII>2.0.CO;2.
- von Storch H, Hewitson B, Mearns L. 2000. Review of empirical downscaling techniques. Proceedings of RegClim spring meeting, 8 - 9 May 2000, Jevnaker, Norway, Available at [http://regclim.met.no/rapport\\_4/Default.htm](http://regclim.met.no/rapport_4/Default.htm). (Accessed on 1<sup>st</sup> October 2012).
- Wang W. 2006. *Stochasticity, nonlinearity and forecasting of streamflow processes*. Deft University Press, Amsterdam. 72-73.
- Wilby RL, Charles SP, Zorita E, Timbal B, Whetton P, Mearns LO. 2004. Guidelines for use of climate scenarios developed from statistical downscaling methods, supporting material to the IPCC. Downloaded from <http://www.ipcc-data.org/>. 3-21.

Willems P, Vrac M. 2011. Statistical precipitation downscaling for small-scale hydrological impact investigations of climate change. *Journal of Hydrology* **402**: 193-205. DOI:10.1016/j.jhydrol.2011.02.030.

Yang T, Li H, Wang W, Xu CY, Yu Z. 2012. Statistical downscaling of extreme daily precipitation, evaporation, and temperature and construction of future scenarios. *Hydrological Processes* **26**: 3510–3523. DOI:10.1002/hyp.8427.



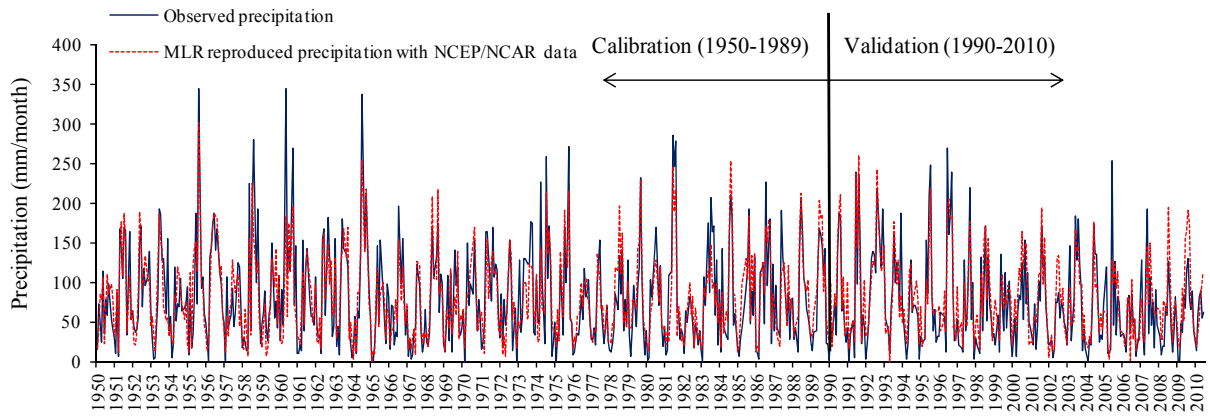
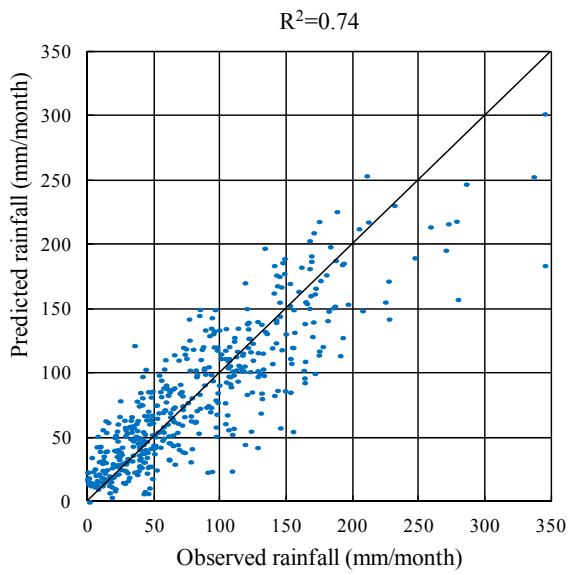
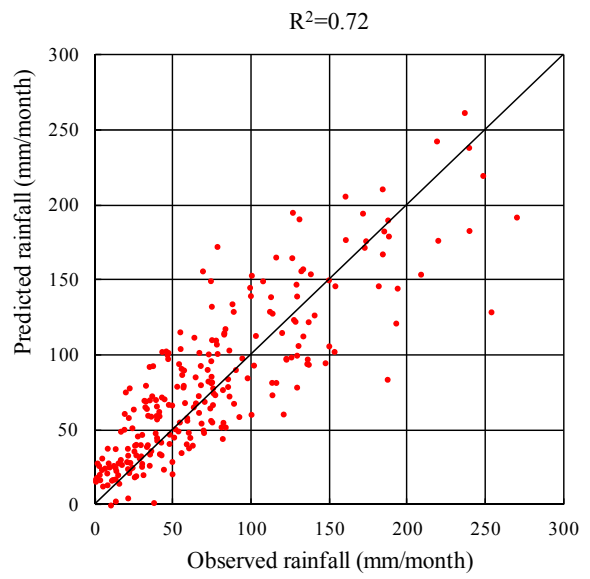


Figure 2 Observed and Model<sub>(NCEP/NCAR)</sub> reproduced monthly precipitation (1950 to 2010)



(a) Calibration



(b) Validation

Figure 3 Scatter plots of observed and Model<sub>(NCEP/NCAR)</sub> reproduced monthly precipitation for calibration (1950-1989) and validation (1990-2010)

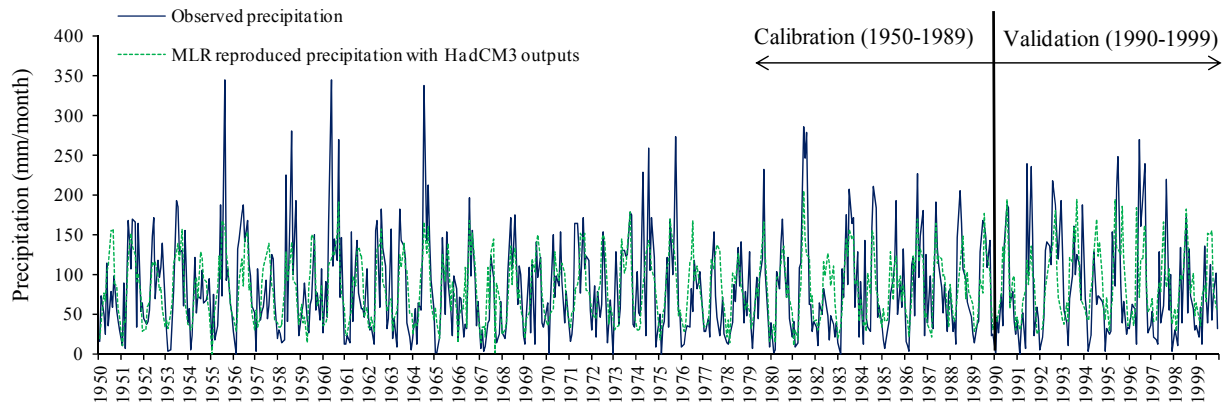


Figure 4 Observed and Model<sub>(HadCM3)</sub> reproduced monthly precipitation (1950 to 1999)

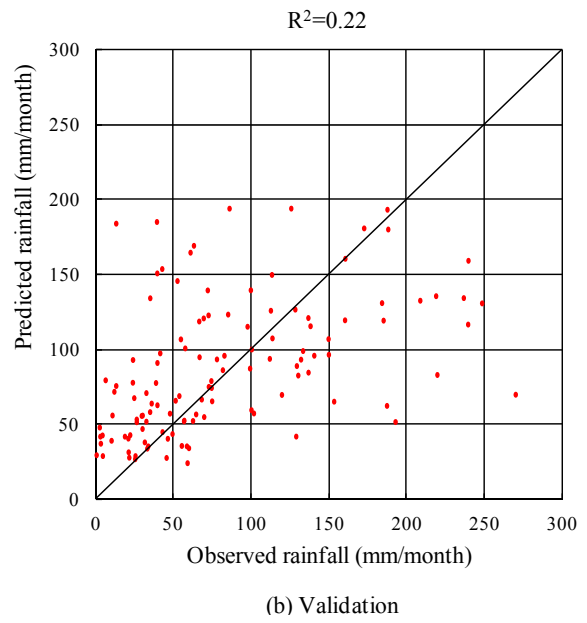
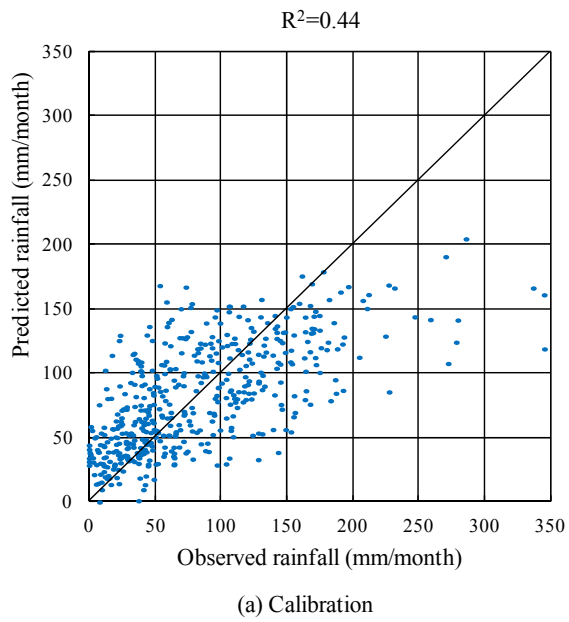


Figure 5 Scatter plots of observed and Model<sub>(HadCM3)</sub> reproduced monthly precipitation for calibration (1950-1989) and validation (1990-1999)



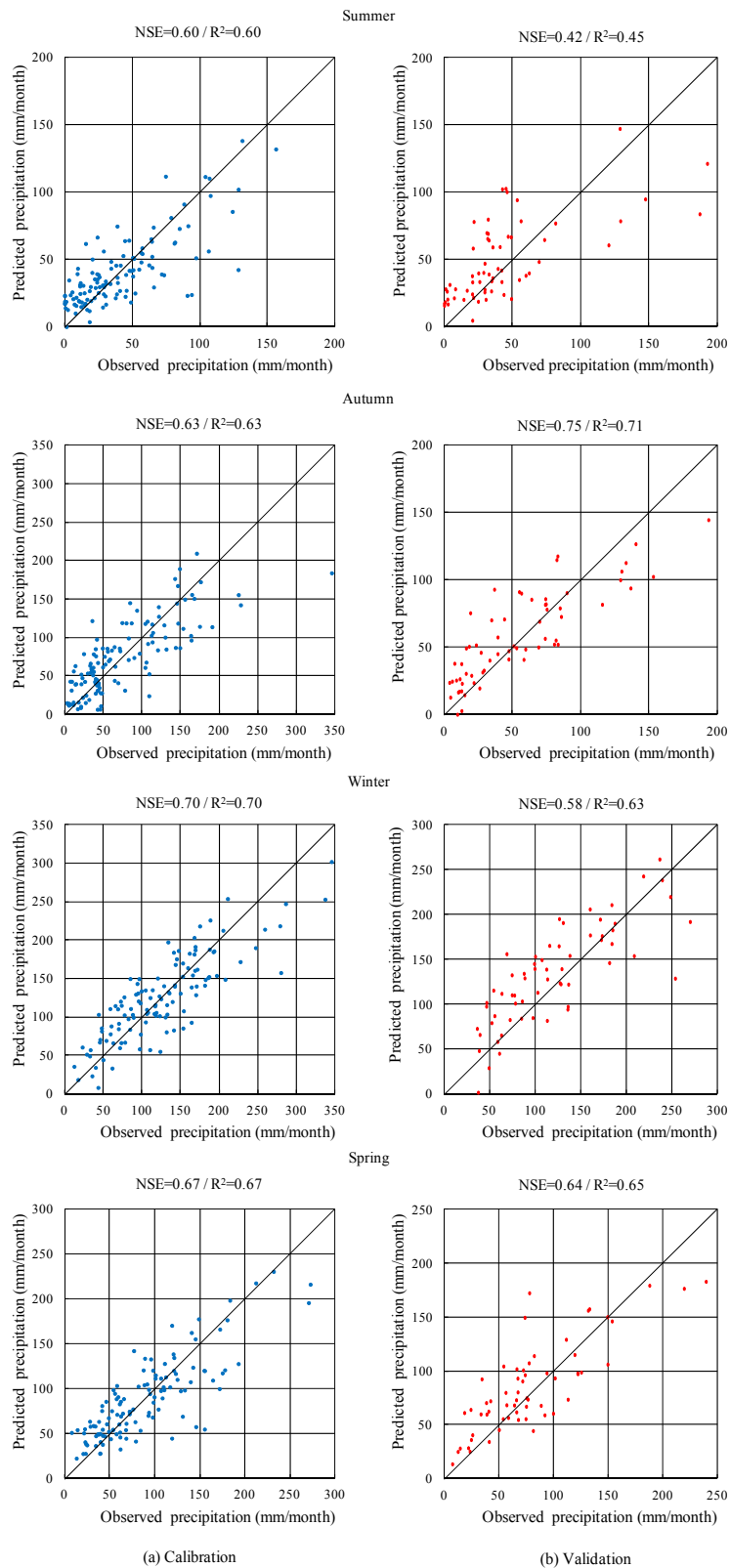


Figure 6 Seasonal scatter plots of observed and Model (NCEP/NCAR) reproduced monthly precipitation for calibration (1950-1989) and validation (1990-2010)

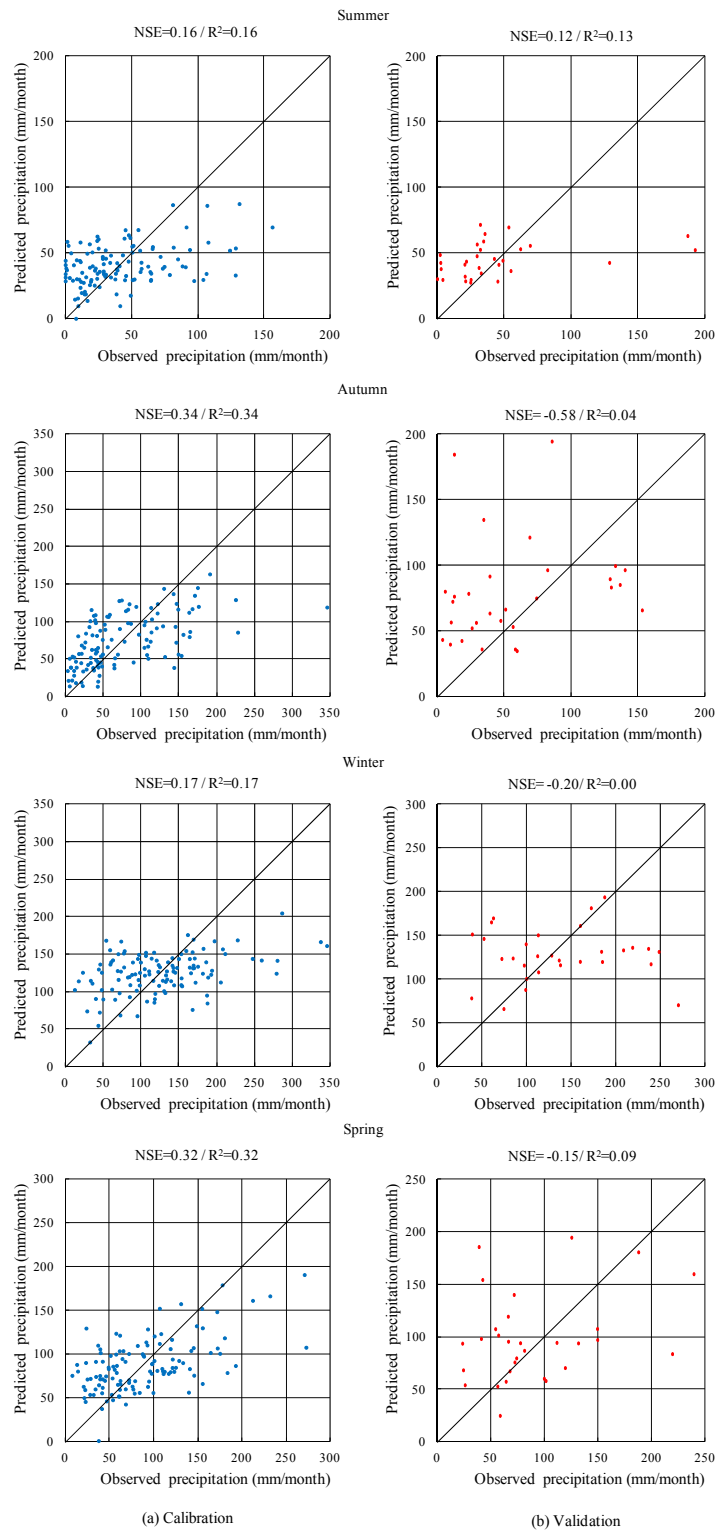


Figure 7 Seasonal scatter plots of observed and Model (HadCM3) reproduced monthly precipitation for calibration (1950-1989) and validation (1990-1999)

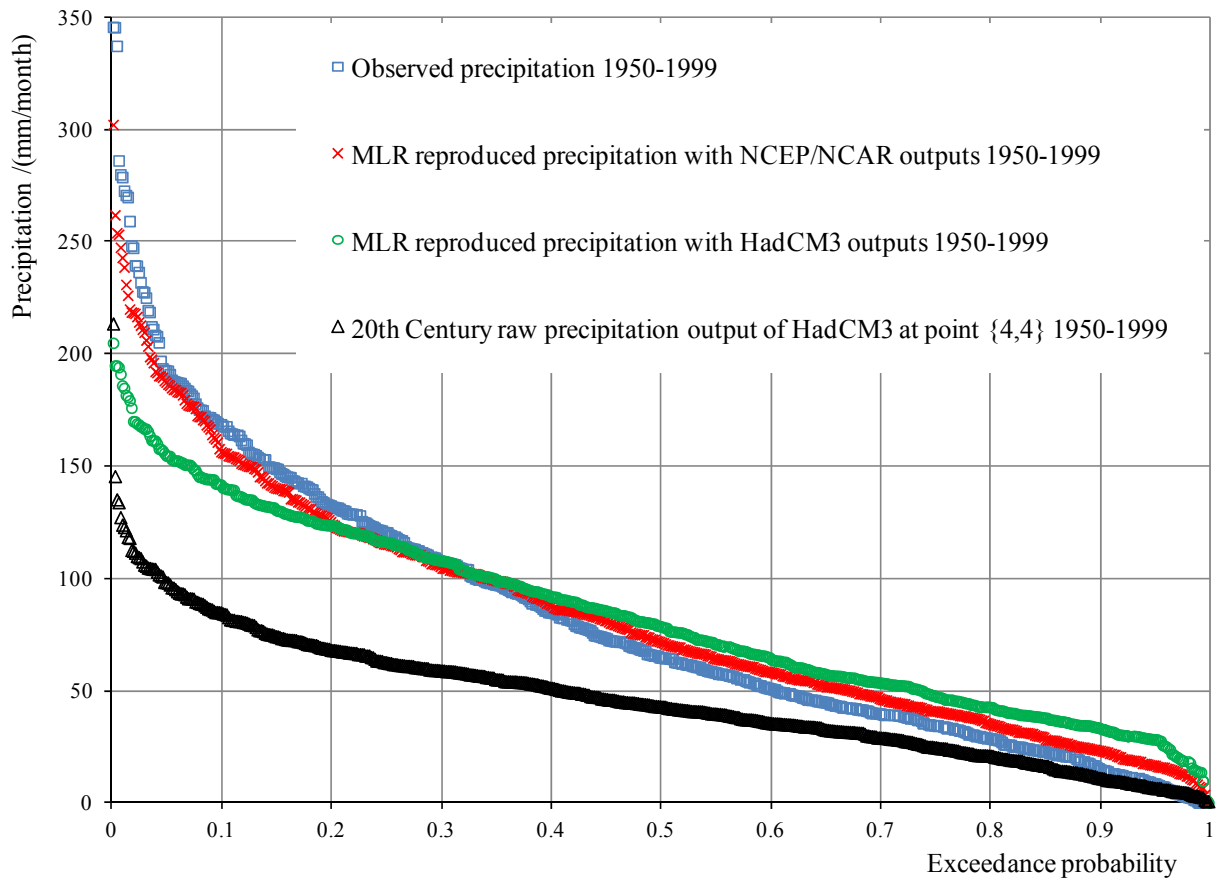


Figure 8 Precipitation probability exceedance curves (1950 to 1999)

Table 1 Final set of potential predictors used in the January downscaling model and their correlations with observed precipitation in each time slice and whole period.

Rank of variable	Potential variables for January	Grid location	Time slice	Correlation with precipitation
1	Surface precipitation rate	4,4	1950-1969	0.910
			1970-1989	0.581
			1990-2010	0.651
			<b>1950-2010</b>	<b>0.693</b>
2	1000hPa specific humidity	3,3	1950-1969	0.532
			1970-1989	0.532
			1990-2010	0.603
			<b>1950-2010</b>	<b>0.550</b>
3	850hPa meridional wind	3,6	1950-1969	-0.466
			1970-1989	-0.698
			1990-2010	-0.468
			<b>1950-2010</b>	<b>-0.548</b>
4	850hPa meridional wind	3,5	1950-1969	-0.400
			1970-1989	-0.724
			1990-2010	-0.522
			<b>1950-2010</b>	<b>-0.544</b>
5	1000hPa specific humidity	3,4	1950-1969	0.487
			1970-1989	0.553
			1990-2010	0.516
			<b>1950-2010</b>	<b>0.515</b>
6	850hPa meridional wind	2,6	1950-1969	-0.420
			1970-1989	-0.585
			1990-2010	-0.506
			<b>1950-2010</b>	<b>-0.513</b>
7	2m specific humidity	3,3	1950-1969	0.430
			1970-1989	0.550
			1990-2010	0.540
			<b>1950-2010</b>	<b>0.510</b>
8	Surface precipitation rate	3,3	1950-1969	0.608
			1970-1989	0.413
			1990-2010	0.523
			<b>1950-2010</b>	<b>0.498</b>
9	1000hPa specific humidity	4,4	1950-1969	0.595
			1970-1989	0.508
			1990-2010	0.412
			<b>1950-2010</b>	<b>0.494</b>
10	850hPa relative humidity	1,2	1950-1969	0.438
			1970-1989	0.475
			1990-2010	0.596
			<b>1950-2010</b>	<b>0.483</b>
11	2m specific humidity	3,4	1950-1969	0.448
			1970-1989	0.533
			1990-2010	0.482
			<b>1950-2010</b>	<b>0.482</b>

Table 2 Final sets of potential predictors for each calendar month

Month	Potential variables used in the model with grid locations
January	Surface precipitation rate {(3,3),(4,4)} 1000hPa specific humidity {(3,3),(3,4),(4,4)} 850hPa meridional wind {(2,6),(3,5),(3,6)} 850hPa relative humidity {(1,2)} 2m specific humidity {(3,3),(3,4)}
February	Surface precipitation rate {(3,4),(4,4),(4,5)}
March	Surface precipitation rate {(3,3),(3,4),(3,5),(4,3),(4,4),(4,5),(4,6)}
April	850hPa relative humidity {(4,3),(4,4)} Surface precipitation rate {(4,3)}
May	Surface precipitation rate {(4,4),(5,5)} 850hPa geopotential height {(4,3)}
June	Surface precipitation rate {(3,2),(3,3),(4,2),(4,3),(4,4),(4,5)} Mean sea level pressure {(4,3),(5,3)} 850hPa zonal wind {(2,4)} Surface pressure{(4,3),(5,3),(5,4)}
July	850hPa zonal wind {(1,3),(1,4)} 850hPa geopotential height {(4,3),(4,4),(4,5)}
August	Surface precipitation rate {(4,3),(5,4),(5,5)}
September	Surface precipitation rate {(2,1),(2,2),(3,2),(3,3),(3,5),(4,2),(4,3),(4,4),(4,5)} 850hPa relative humidity {(3,3)} 700hPa relative humidity {(3,4)}
October	Surface precipitation rate {(3,2),(4,2),(4,3),(4,4)} 850hPa relative humidity {(4,3)} 700hPa geopotential height {(1,1)}
November	850hPa relative humidity {(3,2),(3,3)} Surface precipitation rate {(4,3),(4,5)}
December	Surface precipitation rate {(2,1),(3,2),(4,3),(4,4),(5,5)} 850hPa relative humidity {(3,2)}

hPa = Atmospheric pressure in hectopascal; and the locations are given within brackets (see Figure 1)

Table 3 Performances of downscaling models in calibration and validation

Statistic	Calibration (1950-1989)			Validation (1990-2010) / (1990-1999)*			
	Observations	Model <sub>(NCEP/NCAR)</sub>	Model <sub>(HadCM3)</sub>	Observations		Model <sub>(NCEP/NCAR)</sub>	Model <sub>(HadCM3)</sub>
				1990-2010	1990-1999		
Avg	81.8	82.0	81.7	73.3	<b>81.8</b>	81.0	<b>87.6</b>
Std	61.7	53.2	41.1	56.9	<b>64.3</b>	51.9	<b>44.5</b>
C <sub>v</sub>	0.75	0.65	0.50	0.78	<b>0.79</b>	0.64	<b>0.51</b>
NSE		0.74	0.44			0.70	<b>0.17</b>
SANS		0.66	0.26			0.61	<b>-0.20</b>
R <sup>2</sup>		0.74	0.44			0.72	<b>0.22</b>

Avg = average of monthly precipitation in mm, Std = standard deviation of monthly precipitation in mm, C<sub>v</sub> = coefficient of variation, SANS = Seasonally Adjusted Nash Sutcliffe efficiency, NSE = Nash Sutcliffe efficiency, R<sup>2</sup> = coefficient of determination, \*Note: bold italicised values in the table refer to period 1990-1999.

Table 4 Seasonal performances of downscaling models

Model	Statistic	Calibration (1950-1989)				Validation (1990-2010)/(1990-1999)*			
		Season				Season			
		Summer	Autumn	Winter	Spring	Summer	Autumn	Winter	Spring
Observed		40.7	73.7	125.1	87.7	42.9/( <b>44.3</b> )	54.1/( <b>57.0</b> )	119.4/( <b>136.1</b> )	78.3/( <b>89.8</b> )
Model <sub>(NCEP/NCAR)</sub>	Avg	40.7	73.7	125.1	87.7	49.2	57.8	132.5	85.1
Model <sub>(HadCM3)</sub>		40.3	73.8	125.1	87.8	( <b>44.9</b> )	( <b>78.8</b> )	( <b>128.3</b> )	( <b>98.5</b> )
Observed		33.7	58.8	64.5	53.5	41.0/( <b>46.8</b> )	43.1/( <b>46.5</b> )	61.2/( <b>66.3</b> )	48.4/( <b>55.1</b> )
Model <sub>(NCEP/NCAR)</sub>	Std	26.0	46.6	54.1	43.9	29.8	33.1	54.1	41.7
Model <sub>(HadCM3)</sub>		15.6	34.4	26.7	30.5	( <b>12.7</b> )	( <b>39.0</b> )	( <b>30.0</b> )	( <b>42.0</b> )
Observed		0.83	0.80	0.52	0.61	0.96/( <b>1.06</b> )	0.80/( <b>0.82</b> )	0.51/( <b>0.49</b> )	0.62/( <b>0.61</b> )
Model <sub>(NCEP/NCAR)</sub>	C <sub>v</sub>	0.64	0.63	0.43	0.50	0.61	0.57	0.41	0.49
Model <sub>(HadCM3)</sub>		0.39	0.47	0.21	0.35	( <b>0.28</b> )	( <b>0.49</b> )	( <b>0.23</b> )	( <b>0.43</b> )
Model <sub>(NCEP/NCAR)</sub>	NSE	0.60	0.63	0.70	0.67	0.42	0.75	0.58	0.64
Model <sub>(HadCM3)</sub>		0.16	0.34	0.17	0.33	( <b>0.12</b> )	( <b>-0.58</b> )	( <b>-0.20</b> )	( <b>-0.15</b> )
Model <sub>(NCEP/NCAR)</sub>	R <sup>2</sup>	0.60	0.63	0.70	0.67	0.45	0.71	0.63	0.65
Model <sub>(HadCM3)</sub>		0.16	0.34	0.17	0.33	( <b>0.13</b> )	( <b>0.04</b> )	( <b>0.00</b> )	( <b>0.09</b> )

Avg = average of monthly precipitation in mm, Std = standard deviation of monthly precipitation in mm, C<sub>v</sub> = coefficient of variation, NSE = Nash Sutcliffe efficiency, R<sup>2</sup> = coefficient of determination, \*Note: bold italicised values in brackets in the table refer to period 1990-1999.