Multi-Model Ensemble Approach for Statistically Downscaling General Circulation Model Outputs to Precipitation

D.A. Sachindra[a], F. Huang[a], A.F. Barton[a,b], and B.J.C. Perera[a]

[a]*Victoria University, P.O. Box 14428, Melbourne, Victoria 8001, Australia*

[b]*University of Ballarat, PO Box 663, Ballarat, Victoria 3353, Australia*

Email: sachindra.dhanapalaarachchige@live.vu.edu.au

*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. Telephone: +61 03 9919 4907.

Abstract

Two statistical downscaling models were developed for downscaling monthly GCM outputs to precipitation at a site in north-western Victoria, Australia. The first downscaling model was calibrated and validated with the NCEP/NCAR reanalysis outputs over the periods of 1950-1989 and 1990-2010 respectively. The projections of precipitation into future were produced by introducing the outputs of HadCM3, ECHAM5, GFDL2.0 and GFDL2.1, pertaining to A2 and B1 greenhouse gas emission scenarios to this downscaling model. In this model, the input data used in the development and future projections are not homogeneous, as they originate from two different sources. As a solution to this issue, the second downscaling model was developed and precipitation projections into future were produced with a homogeneous set of inputs. To produce a homogeneous set of inputs to this model, regression relationships were formulated between the NCEP/NCAR reanalysis outputs and the 20th century climate experiment outputs corresponding to the variables used in the first downscaling model obtained from the ensemble consisted of HadCM3, ECHAM5 and GFDL2.0. The outputs of these relationships pertaining to the periods of 1950-1989 and 1990-1999, were used for the calibration and validation of this downscaling model respectively. Using the outputs of HadCM3, ECHAM5 and GFDL2.0 pertaining to A2 and B1 emission scenarios on these relationships, inputs for the second downscaling model pertaining to the period of 2000-2099 were generated. The first downscaling model with NCEP/NCAR reanalysis outputs, showed a high Nash-Sutcliffe Efficiency (NSE) of 0.75 over the period 1950-1999. When this downscaling model was run with the 20th century climate experiment outputs of HadCM3, ECHAM5, GFDL2.0 and GFDL2.1, it exhibited limited performances over the period 1950-1999, which was

indicated by relatively low NSEs of -0.62, -2.54, -0.40 and -0.48 respectively. The second downscaling model displayed a NSE of 0.35 over the period 1950-1999.

Keywords: Statistical downscaling, Precipitation, General Circulation Models, Multi-model ensemble

1. INTRODUCTION

General Circulation Models (GCMs) are widely used for the simulation of future climate (King *et al.*, 2012). GCMs are based on atmospheric physics and they project the climate at global or continental scales into future, with acceptable reliability, considering the concentrations of greenhouse gases (GHG) in the atmosphere (Huth, 2002). However, due to the coarse spatial resolution of GCM projections, their direct use at regional/local scale is limited (Jeong *et al.*, 2012). Dynamic and statistical downscaling techniques have been developed to relate the coarse resolution GCM outputs with regional scale hydroclimatic variables such as precipitation, temperature, streamflow etc. IPCC (2007) defined downscaling as the process of producing sub-spatial scale climatic information from the coarse scale GCM outputs. In dynamic downscaling, a regional climate model (RCM) is nested in a GCM to provide the boundary conditions for the simulation of climate at regional scale. Owing to the physics based structure, dynamic downscaling can yield more realistic predictions of regional climate. Furthermore, dynamic downscaling can produce spatially continuous projections of climate, while preserving some spatial correlations (Maurer and Hidalgo, 2008). However, dynamic downscaling suffers from high computational costs and the proper function of the parameterization schemes (sub-grid scale processes are parameterised in RCMs) in future climate is uncertain (IPCC, 2007). In statistical downscaling, empirical relationships are built between the GCM outputs and observations of hydroclimatic variables. These relationships are used with GCM outputs pertaining to future for the projection of point scale hydroclimatology. Statistical downscaling techniques are computationally cheaper, therefore they are widely used in the projection of local scale climate. Statistical downscaling methods are applicable to

finer spatial scales (point scale) than spatial scales at which most dynamic downscaling methods produce their outputs (Willems *et al.,* 2012). Also with a relatively limited knowledge in atmospheric physics a statistical downscaling model can be developed. Furthermore statistical downscaling methods are capable in producing data of predictands that are not included in the outputs of GCMs such as streamflows, leaf wetness etc. All statistical downscaling techniques are dependent on the assumption that the relationships derived between GCM outputs and catchment scale hydroclimatic variables for the past, are also valid for the changing climate in future (Wilby and Wigly, 2000). However, in the presence of human induced climate change, this assumption seems to be doubtful.

Owing to the computational limitations, various assumptions and approximations are employed in GCMs. The assumptions and approximations employed in GCMs cause noise in their outputs. Furthermore, these assumptions and approximations vary from one GCM to another (Sachindra *et al.,* 2013b). Due to this inter-GCM structural difference, they tend to produce different climatic projections (Yu *et al.,* 2002). Therefore, when outputs of GCMs are either statistically or dynamically downscaled, different GCMs tend to produce different predictions at the catchment scale. Decision making based on a wide range of predictions is a difficult task. Ensemble predictions have become popular in the study of climate change as they aid in combining multiple predictions into single prediction (Krishnamurti *et al.,* 1999; Kharin and Zwiers, 2002; Yun *et al.,* 2003, 2005). Two distinct types of ensembles are found in the literature; (1) an ensemble consisting of a set of predictions obtained from single model (e.g. numerical weather forecasting model) and (2) an ensemble of predictions obtained from

number of different models (Kharin and Zwiers, 2002). The former type of ensemble is more popular in numerical weather forecasting. The latter was designated the multi-model "superensemble" by Krishnamurti *et al.* (1999). However, Kharin and Zwiers (2002) commented that there should be a proper designation for an ensemble which contains the individual predictions derived from different GCMs or numerical weather models, as "superensemble" does not aptly describe the nature of this ensemble. In this paper, here onwards the ensemble of multiple GCMs or numerical weather models is designated the multi-model ensemble.

There are several techniques available for deriving ensemble predictions from predictions of single model or individual predictions from a group of models. The most widely used technique is averaging (Zhang and Huang, 2013). Maqsood *et al.* (2004) called this technique the basic ensemble method, as it is the simplest technique for producing an ensemble prediction. In this technique, the outputs of the members of the ensemble are added together and divided by the number of members in the ensemble. Averaging an ensemble of predictions is one of the approaches to reduce the noise in the individual predictions (Kharin *et al.*, 2001). According to Warner *et al.* (2010), often, the average of an ensemble of predictions (ensemble average) is more accurate than any individual prediction in the ensemble. Also Knutti *et al.* (2010) demonstrated that multi-model ensemble averages consistently outperform the performances of any single model involved in the ensemble. Fealy and Sweeney (2008) used averaging to obtain ensemble predictions of temperature, radiation and potential evapotranspiration which were downscaled form several GCMs. In averaging, simply, equal weights are assigned to each member in the ensemble. The drawback of averaging is that, although equal

weights are assigned to each ensemble member (e.g. GCM), their accuracies may vary from one ensemble member to another (Maqsood *et al.*, 2004, Yun *et al.*, 2005).

As a solution to the above issue, method of assigning weights to each member of the ensemble based on their performances is used. According to Zhang and Huang (2013), assigning weights to each individual member of the ensemble based on their performances can reduce the uncertainty in the final projection. The assigning of weights to each member of the ensemble is achieved in number of different ways. However, the basis of the weighting method is to give more preference to the model that presents less error in reproducing the observed climate of the past. Ingol-Blanco (2011) used the method of weighted averaging on an ensemble of five GCMs. In that study, the GCM outputs were downscaled to precipitation and temperature for the simulation of streamflows using a hydrologic model. The weights for each GCM of the ensemble for each calendar month was determined using the ratio between the root mean square error of a certain GCM in simulating streamflow and the sum of the root mean square errors of all GCM in the ensemble in simulating streamflow. Zhang and Huang (2013) used a set of model performance parameters for determining the weights for the members in an ensemble of 4 RCMs. In that study, the performance parameters considered for each RCM included, their ability to reproduce; inter-annual and seasonal circulation patterns of precipitation and temperature, extremes of daily precipitation and temperature, precipitation occurrence, and probability density functions of precipitation and temperature. The skill scores derived for each of the above parameters were combined to a weight value which can explain the overall performance of each RCM. This method of considering the ability of ensemble members in reproducing multiple climatic

characteristics is a better method of assigning weights to ensemble members, as it considers multiple characteristics of the members of the ensemble, unlike the other averaging approaches (e.g. simple averaging). However, according to Christensen *et al.* (2010) still there is large subjectivity in the selection of parameters (e.g. extremes of daily precipitation - Zhang and Huang, 2013) for the assessment of GCM performances and also how the final weights are derived from the model assessment. Furthermore, the technique used by Zhang and Huang (2013) is computationally costly as multiple characteristics of each of the members in the ensemble have to be assessed.

As another ensemble technique, Krishnamurti *et al.* (1999) proposed the regression of outputs of different climate models against observations, for deriving an ensemble prediction. They used the outputs of several GCMs for developing the multi-linear regression relationships between them and the corresponding observations. It was found that the multi-model ensemble developed using the multi-linear regression technique outperforms predictions of all individual models and also the ensemble prediction obtained by simple averaging. Yun *et al.* (2003) used this method prescribed by Krishnamurti *et al.* (1999), in a slightly different format employing singular value decomposition for improving the coefficients of the multi-linear regression equations. However, when the relationship between the observations (or reanalysis outputs) and GCM outputs is non-linear in nature, non-linear regression may better model this relationship, although its computational cost is higher due to complexity.

It is the common practice to develop (calibrate and validate) a statistical downscaling model with some reanalysis outputs (e.g. National Centre for Environmental

Predictions/National Centre for Atmospheric Research - NCEP/NCAR) and observations pertaining to the past climate. Thereafter, the outputs of a GCM (or GCMs) are introduced to the downscaling model for the projection of climate into future. The major flaw associated with this approach is that, the downscaling model is developed with some reanalysis outputs which are of higher quality (due to quality controlling and corrective procedures), but the future projections are produced with the outputs of a GCM (or GCMs) which are associated with higher degree of uncertainty. In other words, the input data sets used in the development phase of the downscaling model and the future projection phase are not homogeneous, in terms of quality, as they originate from two different sources with different degrees of accuracy. Therefore in the above approach there is no smooth transition from the model development phase to the future projection phase. As a solution to this issue, Sachindra *et al.* (2013b) proposed a statistical downscaling model purely developed (calibrated and validated) with the 20th century climate experiment outputs of HadCM3, for downscaling monthly precipitation, which can be used for downscaling the future projections of climate of the same GCM to catchment scale. However, in that study it was found that the performances of this downscaling model in the calibration and validation periods were quite limited in comparison with another downscaling model developed with NCEP/NCAR reanalysis outputs for the same purpose. Furthermore in that study, a large mismatch was identified between the raw precipitation output of HadCM3 and that of NCEP/NCAR reanalysis data. This mismatch reflected the bias prevalent in HadCM3 outputs.

This paper discusses the development (calibration and validation) of two statistical downscaling models based on the multi-linear regression (MLR) technique for

downscaling GCM outputs to monthly precipitation at a station located in the Grampians water supply system in north-western Victoria, Australia. Also these two downscaling models were used for the projection of precipitation into future at this station with the outputs of GCMs. The first downscaling model was developed with the NCEP/NCAR reanalysis outputs. The precipitation projections into future were produced by introducing the outputs of HadCM3, ECHAM5, GFDL2.0 and GFDL2.1 to this downscaling model, individually. Smith and Chandler (2009) stated that HadCM3, ECHAM5, GFDL2.0, GFDL2.1 and MIROC3.2 provide good representation of precipitation over the Australian continent and credible simulation of El Niño Southern Oscillation (ENSO). Hence in this study, HadCM3, ECHAM5, GFDL2.0 and GFDL2.1 were used to provide inputs to the downscaling models. MIROC3.2 was not employed in this study due to the limited availability of its output data. The second downscaling model was developed (calibrated and validated) with a set of input data derived by building relationships (equations) between the 20th century climate experiment outputs of HadCM3, ECHAM5 and GFDL2.0 and the corresponding NCEP/NCAR reanalysis outputs, using the MLR technique. GFDL2.1 was not used in the second downscaling model since it showed limited performance in terms of coefficient of determination in comparison with its older version GFDL2.0, when used with the first downscaling model (as described later in Section 4.5). The MLR equations were used for deriving the inputs for the downscaling model for producing projections into future with the outputs of the above three GCMs pertaining to the future climate. Therefore this approach allows for the use of a homogeneous set of inputs for the development of the downscaling model and also for producing climatic projections into future, while maintaining a certain degree of consistency with the NCEP/NCAR reanalysis outputs by

reducing the mismatch (bias) of GCM outputs with the NCEP/NCAR reanalysis outputs, to a certain extent. The equidistant quantile mapping technique (Li *et al.*, 2010) was used for the correction of bias in the precipitation downscaled by both downscaling models for the past and future climate. A performance comparison of the above two statistical downscaling models for the calibration and validation periods is also provided in this paper.

Section 2 of this paper provides a short account on the study area and the data used in the present research. Section 3 details the generic methodology. Section 4 explains the application of this generic methodology to the precipitation station of interest along with the results. Finally, a summary on the work and conclusions yielded in this study are provided in Section 5.

2. STUDY AREA AND DATA

For the case study, the precipitation station at the Halls Gap post office (latitude -37.14˚S, longitude 142.52˚E) was selected. This precipitation station is located in the Grampians water supply system in north-western Victoria, Australia, within the operational area of the Grampians Wimmera Mallee Water Corporation (GWMWater) (www.gwmwater.org.au). In order to demonstrate of the methodology only the observation station at the Halls Gap post office was considered in this study. Furthermore, during the period 1950-2010, the observations of precipitation at the Halls Gap post office displayed high correlations (magnitudes above 0.80 at $p \leq 0.05$) with those at other surrounding stations (Eversley, Great Western, Polkemmet, Lake Lonsdale, Moyston post office, Wartook reservoir, Hamilton airport, Tottington,

Stawell, Balmoral post office and Ararat prison) in the study area. Therefore it was assumed that the projections produced at the Halls Gap post office are also valid for the other stations that were highly correlated with it. Figure 1 shows the location of this precipitation station in the operational area of the GWMWater.

Figure 1 Location of the precipitation station

A record of observed daily precipitation for the period 1950 – 2010 was obtained from the SILO database of Queensland Climate Change Centre of Excellence at http://www.longpaddock.qld.gov.au/silo/, for the calibration and validation of the two downscaling models. In the period 1950-2010, the daily observations of precipitation at the Halls Gap post office contained 31.2% missing data and these missing data have been filled by the Queensland Climate Change Centre of Excellence in the SILO database, using the spatial interpolation method detailed in Jeffrey *et al.* (2001). These daily data in the SILO database were used to compute precipitation totals in each month over the period 1950-2010. The observed precipitation record was also used in the bias-correction, as the reference data set. In order to provide inputs to the first downscaling model and to build the MLR relationships with the GCM outputs for developing of the second downscaling model, monthly NCEP/NCAR reanalysis outputs were downloaded from the physical sciences division of National Oceanic and Atmospheric Administration/Earth System Research Laboratory (NOAA/ESRL) at http://www.esrl.noaa.gov/psd/.

The monthly outputs of HadCM3(2.75˚ latitude by 3.75˚ longitude), ECHAM5(1.865˚ latitude by 1.875˚ longitude), GFDL2.0(2.022˚ latitude by 2.50˚ longitude) and GFDL2.1(2.022˚ latitude by 2.50˚ longitude) for the 20th 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, to provide the inputs to both downscaling models for the reproduction of the past observed precipitation. For the validation of the bias-correction, HadCM3 outputs pertaining to the COMMIT GHG emission scenarios were obtained from https://esgcet.llnl.gov:8443/index.jsp for the period 2000-2099. The COMMIT GHG emission scenarios assumed that the atmospheric GHG concentrations observed at year 2000 remain constant throughout the 21st century. The monthly outputs of HadCM3, ECHAM5, GFDL2.0 and GFDL2.1 for the A2 and the B1 GHG emission scenarios (IPCC, 2000) were obtained for the period 2000 – 2099, from https://esgcet.llnl.gov:8443/index.jsp, in order to provide inputs to the two downscaling models, for producing the projections of catchment scale precipitation into future. The A2 emission scenario refers to a future world with a greater economic focus and the B1 scenario describes a future world oriented more towards environmental protection than rapid economic development.

3. GENERIC METHODOLOGY

3.1. Atmospheric domain and predictor selection

As the first step of the study, an atmospheric domain over the study area was defined. In a downscaling study, the atmospheric domain enables the consideration of the influences of the surrounding atmosphere on the climate at the station of interest.

A set of probable predictors for this study was first selected based on the past literature on downscaling precipitation and also hydrology. These probable predictors are the most likely variables to influence precipitation at the catchment scale. The potential predictors which are the most influential variables on precipitation were extracted from the set of probable variables. The relationship between the predictors (e.g. reanalysis or GCM output) and predictands (e.g. precipitation) vary seasonally due to the seasonal variations of the atmospheric motions (Karl *et al.*, 1990). Therefore to represent the seasonality of the predictor-predictand relationships, the potential predictors were selected from the pool of probable predictors, for each calendar month separately. Using these sets of potential predictors, a separate downscaling model was developed for each calendar month. This approach aids in better modelling the seasonal variations of precipitation. Throughout this study, the MLR technique was used in developing the statistical downscaling models as it has been identified as a potential technique for developing downscaling models in the studies by Sachindra *et al.* (2013a, 2013b).

The data for probable predictors were extracted from the NCEP/NCAR reanalysis database. These reanalysis data for probable predictors and observed precipitation were split chronologically into 20 year time slices. Then the Pearson correlation coefficients

(Pearson, 1895) between them were computed for each time slice and the whole period, at all grid points in the atmospheric domain, for each calendar month. As recommended by Sachindra *et al.* (2013a), the probable predictors that showed the best statistically significant correlations ($p \leq 0.05$), consistently in all 20 year time slices and the whole period were selected as potential predictors. These consistently correlated variables referred to the predictor variables which displayed no sign variations and no large variations in the magnitude in their correlations with the observed precipitation, over all time slices and the whole period of the study.

3.2. Downscaling model with NCEP/NCAR outputs

The development (calibration and validation) of the first downscaling model was done by using the NCEP/NCAR reanalysis outputs corresponding to the potential predictors identified in the previous sub-section. The first two third of the reanalysis and observed precipitation was allocated for the calibration of this downscaling model and the remaining one third was used for the validation of the model. The potential predictors for both the calibration and validation phases were standardised by subtracting the mean and then dividing by the standard deviation of the NCEP/NCAR reanalysis outputs pertaining to the calibration period of the downscaling model. The means and the standard deviations of the NCEP/NCAR reanalysis outputs pertaining to the calibration period were used for standardising the input data to the downscaling model in calibration, validation and future projection periods (Sachindra *et al.,* 2013c). In the calibration of this model, first, the three potential predictors which showed the best correlations with precipitation over the whole period of the study were introduced to the model. The optimum values for the coefficients and constants in the MLR equations

were found by minimising the sum of squared errors between the observed precipitation and precipitation reproduced by the downscaling model. Thereafter the model validation was performed with this calibrated downscaling model as an independent simulation. The Nash-Sutcliffe efficiency (NSE) (Nash and Sutcliffe, 1970) was used in assessing the performances of the downscaling model in the calibration and validation periods. Thereafter the next potential predictors which displayed the best correlation with observed precipitation over the whole period were introduced to the model, one at a time. This procedure was continued until the model performance in validation reaches a maximum, in terms of NSE. This process of stepwise model development enabled finding the best set of potential predictors and the best downscaling model for a calendar month. In this manner, downscaling models were calibrated and validated for each calendar month, while identifying the best potential predictors for each calendar month. For the development of this downscaling model, the statistics toolbox in MATLAB (R2008b) was used.

The 20^{th} century climate experiment outputs and the outputs of future GHG emission scenarios of several GCMs (in this study HadCM3, ECHAM5, GFDL2.0 and GFDL2.1) corresponding to the best potential predictors identified in the model development stage were standardised with the means and the standard deviations of the NCEP/NCAR reanalysis outputs pertaining to the model calibration period, for each calendar month. The 20^{th} century climate experiment outputs of GCMs represent the climate in the last century. Therefore they can be used on the downscaling model to reproduce the past climatic observations at the catchment scale. The standardised 20^{th} century climate experiment outputs of each GCM were introduced to the downscaling model for the

reproduction of past observed precipitation at the station of interest. This was important in judging how this downscaling model which was built with reanalysis data will perform in downscaling GCM outputs to precipitation at the catchment scale. The standardised GCM outputs for future were applied to the downscaling model to project precipitation into future, at the station of interest. These two sets of precipitation (reproduced past and projected future) were bias-corrected using the equidistant quantile mapping technique. In equidistant quantile mapping (Li *et al.,* 2010), the empirical cumulative distribution functions (CDFs) are computed for the observed precipitation, and also for the past and future precipitation produced by the downscaling model with GCM outputs as inputs to it. As the first step, in equidistant quantile mapping, the CDF of precipitation reproduced by the downscaling model for the past was exactly mapped onto the CDF of the observed precipitation. This made all the statistical moments of model-reproduced past precipitation to match with those of observed precipitation. In the bias-correction of precipitation downscaled by the model pertaining to future, the difference between the CDF of the downscaled precipitation (to be corrected) corresponding to future and the CDF of the precipitation of the past climate reproduced by the downscaling model with 20^{th} century climate experiment outputs of GCM, was added to the bias-corrected version of the latter CDF (Salvi *et al.,* 2011). A detailed description on equidistant quantile mapping is found in the studies by Li *et al.* (2010), Salvi *et al.* (2011) and Sachindra *et al.* (2013c).

Principal component (PC) analysis is a technique used in order to extract the variance present in a large set of predictors to a limited number of PCs (e.g. Tripathi *et al.,* 2006; Anandhi *et al.,* 2008). In this manner, principal component analysis (PCA) can

minimise the redundancy errors caused by the inter-correlated predictors used as inputs to a downscaling model. However, caution must be exercised in applying PCA in downscaling as it can cause serious errors in the outputs of the downscaling model. The coefficients of the PCs extracted from the potential predictor data for the calibration phase of the downscaling model, becomes a fixed component of the downscaling model in its validation and future projection phases. In such case, some PCs pertaining to the validation period or the future projection period may become markedly correlated. In theory, PCs need to be near zero correlated with each other. Use of PCs that are considerably correlated with each other, in the validation and future projection phases can cause serious errors in model predictions. This phenomenon was well documented by Sachindra *et al.* (2013a). Hence PCA was not used in preparing inputs to the downscaling models developed in the present study.

3.3. Downscaling model with multi-model ensemble outputs

For the development of the second downscaling model, the same atmospheric domain and the same sets of best potential predictors identified during the development process of the first downscaling model were used. It was assumed that these best potential predictors are valid for both downscaling models, as the predictand (in this case monthly precipitation) and the point at the catchment (precipitation station) remained the same for both models. Unlike in the first downscaling model, in this downscaling model, the calibration and validation processes were not performed using the NCEP/NCAR reanalysis outputs. Instead, the 20th century climate experiment outputs of an ensemble of GCMs were regressed with the NCEP/NCAR reanalysis outputs using multi-linear regression to create a new set of inputs for the development of this

downscaling model. This new set of inputs used for the development of the downscaling model is called the multi-model ensemble output in this paper. This approach of producing a multi-model ensemble output enables the use of a homogeneous set of inputs to the downscaling model in calibration, validation and future projection periods. Furthermore, since the outputs of the ensemble members (GCMs used) are regressed against the reanalysis outputs, the presence of any mismatch between the GCM outputs and reanalysis outputs in the multi-model ensemble output is reduced. Thus, the above procedure of the multi-model ensemble outputs can also act as a simple bias-correction. The use of multiple GCMs in producing the inputs to the downscaling model can avoid the dependence on a single GCM in projecting catchment scale climate into future.

There were three distinct stages in the establishment of the second downscaling model. At the first stage, the inputs (multi-model ensemble outputs) to the downscaling model were produced. At the second stage, the downscaling model was calibrated and validated with these inputs developed in the first stage. At the third stage, the multi-model ensemble outputs pertaining to climate in the future were introduced to this downscaling model for the projection of catchment level precipitation at the station of interest into future.

At stage one, for developing the inputs for the second downscaling model, the NCEP/NCAR reanalysis outputs and the 20^{th} century climate experiment outputs of the ensemble members (in this study HadCM3, ECHAM5, GFDL2.0) were standardised with the means and the standard deviations of NCEP/NCAR reanalysis outputs pertaining to the calibration period of the downscaling model, for each calendar month.

This downscaling model also had the same calibration period as the first downscaling model which was detailed in Section 3.2. Then for each best potential predictor in each calendar month, a MLR equation was fitted between the NCEP/NCAR reanalysis outputs and the pertaining 20^{th} century climate experiment outputs of the ensemble members, corresponding to the calibration period of the downscaling model. The 20^{th} century climate experiment outputs of the ensemble members, pertaining to the validation period of the downscaling model were introduced to the above developed MLR equations for generating the inputs for the validation phase of the downscaling model.

At stage two, for each calendar month, the multi-model ensemble outputs pertaining to the best set of potential predictors were introduced to the downscaling model and it was calibrated by minimising the sum of squared errors between the observed precipitation and precipitation reproduced by the downscaling model. Since the best potential predictors were already selected in the development of the first downscaling model, in building the second downscaling model, stepwise addition of predictors was not needed. Following the calibration, the downscaling model was validated with the multi-model ensemble outputs relevant to the validation phase of this model. For the development of this downscaling model, the statistics toolbox in MATLAB (Version - R2008b) was used.

At stage three, the outputs of the GCMs corresponding to climate in future were standardised with the means and the standard deviations of the NCEP/NCAR reanalysis outputs pertaining to the calibration period of the downscaling model. These

standardised GCM outputs (corresponding to GHG emission scenarios) were applied on the MLR equations developed between the NCEP/NCAR reanalysis outputs and the 20th century climate experiment outputs of GCMs to generate the inputs to the downscaling model for producing precipitation projections into future. These multi-model ensemble outputs for the future climate were introduced to the downscaling model to downscale them to catchment scale precipitation. As described Section 3.2, the equidistant quantile mapping technique was applied to correct the bias in the precipitation output of this downscaling model.

3.4. Evaluation of the performances of downscaling models

The performances of both downscaling models in the calibration and validation periods were assessed numerically and graphically. The numerical assessment of the two downscaling models was performed by comparing the statistics of the precipitation reproduced by the downscaling model with those of the observed precipitation. These statistics included the average, standard deviation, coefficient of variation, NSE, Seasonally Adjusted NSE (SANS) (Wang, 2006; Sachindra *et al.*, 2013a) and the coefficient of determination (R^2). Scatter and time series plots were used to provide a graphical comparison between the model predictions and observations.

4. APPLICATION

The generic methodology described in Section 3 was applied to the precipitation station at Halls Gap post office in north western Victoria, Australia (see Figure 1).

4.1. Atmospheric domain and predictor selection

An atmospheric domain which covered the range of longitudes from 135.0˚E to 150.0˚E and the range of latitudes from -30.0˚S to -42.4˚S was selected for this study. This domain contained 7 and 6 grid points in the longitudinal and latitudinal directions respectively. Hence, it contained a total of 42 grid points. The atmospheric domain used in this study is shown in Figure 1. The grid points in Figure 1 were 2.5˚ apart in both longitudinal and latitudinal directions. This spatial resolution was maintained to comply with the resolution of the NCEP/NCAR reanalysis outputs. Note that, all GCM outputs used in this study were first interpolated to the NCEP/NCAR reanalysis grid shown (atmospheric domain) in Figure 2, using the inverse distance weighted method. The same atmospheric domain over the same study area was used by Sachindra *et al.* (2013a) for downscaling NCEP/NCAR reanalysis outputs to monthly streamflows at a station which is located close to the precipitation station considered in this study.

Figure 2 Atmospheric domain for downscaling

A set of probable predictors was selected following the studies of Anandhi *et al.* (2008) and Timbal *et al.* (2009), on downscaling GCM outputs to precipitation. Also the hydrology of the precipitation generation was considered in the selection of probable predictors. Timbal *et al.* (2009) identified the predictors influential on the precipitation in south eastern Australia, which covers the study area of this research. The 23 probable predictors used in the current downscaling exercise included; 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. The monthly NCEP/NCAR reanalysis data pertaining to the above 23 probable predictors at the 42 grid locations shown in Figure 2 were obtained from http://www.esrl.noaa.gov/psd/, for the period 1950-2010.

The NCEP/NCAR reanalysis data for the above 23 probable predictors and the observed precipitation were separated into three 20 year time slices; 1950-1969, 1970-1989, and 1990-2010. Following this, the Pearson correlation coefficients between the probable predictors and observed precipitation were computed for the three time slices and the whole period (1950-2010) at each grid point shown in Figure 1. The predictors which displayed the best statistically significant correlations ($p \leq 0.05$) in the three time slices and the whole period were selected as the potential predictor set. For extracting potential predictors, this process was performed for all calendar months separately.

4.2. Downscaling model with NCEP/NCAR outputs

The NCEP/NCAR reanalysis data for the potential predictors (selected as described in Section 4.1) and the observed precipitation data were split into two groups, chronologically. The first group contained the potential predictor and observed precipitation data from 1950 to 1989, and these data were used for the calibration of the first downscaling model. The rest of the data from 1989 to 2010 was used for the validation of this downscaling model. The first downscaling model which was developed with the NCEP/NCAR reanalysis outputs is referred to as $\text{Model}_{(\text{NCEP/NCAR})}$,

in this paper. The means and the standard deviations of potential predictors (NCEP/NCAR reanalysis data) corresponding to the calibration period (1950-1989) were used in standardising the potential predictors for both the calibration and validation periods. The standardised potential predictors were ranked according to the magnitude of the correlation with observed monthly precipitation over the period 1950-2010. Initially, the three best potential predictors (with ranks 1, 2 and 3) were introduced to the downscaling model and it was calibrated by minimising the sum of squared errors between the model predictions and observations. Thereafter the model was validated as an independent simulation, while keeping the optimum values of constants and coefficients (of the MLR equations) which were yielded in calibration fixed. Thereafter the 4^{th}, 5^{th} etc ranked potential predictors were introduced to the downscaling model and it was calibrated and validated. This stepwise addition of potential predictors was performed until the model performance in validation in terms of NSE reaches its maximum. Thus, the best sets of potential predictors were identified for each calendar month. Table i shows the best sets of potential predictors selected for each calendar month. The same sets of the best potential predictors were used in the development of the second downscaling model, which is detailed in Section 4.3.

Table i Best sets of potential predictors for each calendar month

4.3. Downscaling model with multi-model ensemble outputs

Unlike the first downscaling model (Section 4.2), the second downscaling model was not calibrated and validated with the NCEP/NCAR reanalysis data. For the calibration and validation of this downscaling model a set of input data was created using the

outputs of an ensemble of three GCMs which consisted of; HadCM3, ECHAM5 and GFDL2.0. Since these three GCMs originate from three independent organisations it was assumed that they possess different internal structures and hence it was assumed that they are quite independent of each other. Due to the limited performances shown by GFDL2.1 when its outputs were used with $Model_{(NCEP/NCAR)}$ (see Figure 7 and Table iv later in this paper), it was not used in producing inputs to the second downscaling model, described in this section. Considering the above reasons, these three GCM were selected for producing the input for this downscaling model. The 20^{th} century climate experiment data of HadCM3, ECHAM5 and GFDL2.0 relevant to the best sets of potential predictors shown in Table i were obtained for the period 1950-1999. These data were then regresses with the corresponding NCEP/NCAR reanalysis outputs for the period 1950-1989, using the MLR technique. This yielded a MLR relationship between NCEP/NCAR reanalysis outputs and the 20^{th} century climate experiment outputs of the ensemble members, for each best potential predictor for each calendar month. Using the outputs of the ensemble members in these relationships, a single output corresponding to all ensemble members (called multi-model ensemble outputs in this paper) for each best potential predictor in each calendar month was derived for the period 1950-1999. These multi-model ensemble outputs were used as input to the second downscaling model in its calibration and validation. Following the procedure detailed in Section 3.3, this downscaling model was calibrated and validated. Unlike in the development of the $Model_{(NCEP/NCAR)}$, the stepwise construction procedure was not adopted in the development of this downscaling model, as the best potential predictors for each calendar month had been already identified. This second downscaling model is referred to as $Model_{(Ensemble)}$ in this paper.

4.4. Calibration and validation results for $Model_{(NCEP/NCAR)}$ and $Model_{(Ensemble)}$

Figure 3 shows the time series plots for the observed precipitation and the precipitation reproduced by the $Model_{(NCEP/NCAR)}$ during the period 1950-2010. According to the time series plots, it was evident that this model was able to reproduce the pattern of precipitation well in both the calibration and validation periods. This was valid even for the period 1997-2010, that included the Millennium drought in Victoria, which caused a significant drop in precipitation.

Figure 3 Observed and $Model_{(NCEP/NCAR)}$ reproduced monthly precipitation (1950 to 2010)

Figure 4 displays the scatter plots for the calibration (1950-1989) and validation (1990-2010) periods of the $Model_{(NCEP/NCAR)}$. During both the calibration and validation phases, the $Model_{(NCEP/NCAR)}$ showed good accuracy in reproducing the observed precipitation. However, there was a trend of under-predicting high precipitation values and over-predicting near zero precipitation values in both phases. The coefficient of determination (R^2) was high and quite comparable for both calibration (R^2 = 0.74) and validation (R^2 = 0.72) periods of this model. Also the NSEs were high for both calibration (NSE = 0.74) and validation (NSE = 0.70) periods of the model.

Figure 4 Scatter plots of observed and $Model_{(NCEP/NCAR)}$ reproduced monthly precipitation for calibration (1950-1989) and validation (1990-2010)

Figure 5 shows the time series plots for the observed precipitation and precipitation reproduced by the $Model_{(Ensemble)}$ for the period 1950-1999. This downscaling model was able to capture the pattern of observed precipitation throughout the calibration and validation periods, but it showed limited capability in correctly reproducing the extreme precipitation values. The latter trend was less evident in the time series of precipitation reproduced by $Model_{(NCEP/NCAR)}$.

Figure 5 Observed and $Model_{(Ensemble)}$ reproduced monthly precipitation (1950 to 1999)

Figure 6 presents the scatter plots for the calibration (1950-1989) and validation (1990-1999) phases of the $Model_{(Ensemble)}$. In the calibration and validation phases this model, it showed a large under-predicting trend for the high precipitation values. The low precipitation values were largely over-predicted in the calibration period of this model. The under-predicting trend of high precipitation values was more severe in the outputs of this downscaling model, in comparison with that of $Model_{(NCEP/NCAR)}$. During the validation period the scatter was quite high for the precipitation reproduced by the $Model_{(Ensemble)}$. The R^2 for the calibration and validation were 0.47 and 0.13 respectively. The NSEs for the calibration and validation were 0.47 and -0.08 respectively.

Figure 6 Scatter plots of observed and $Model_{(Ensemble)}$ reproduced monthly precipitation for calibration (1950-1989) and validation (1990-1999)

According to the scatter plots in Figures 4 and 6, it can be seen that $Model_{(NCEP/NCAR)}$ is more capable in reproducing the observed precipitation than its counterpart $Model_{(Ensemble)}$. This was due to the fact that the quality of NCEP/NCAR reanalysis outputs used as inputs to the $Model_{(NCEP/NCAR)}$ were much higher than that of GCM outputs used in the preparation of inputs to the $Model_{(Ensemble)}$. The NCEP/NCAR reanalysis outputs are quality controlled and corrected against observations (Kalnay *et al.,* 1996), and therefore they are inherently more accurate than any GCM output.

Table ii provides the performance statistics of $Model_{(NCEP/NCAR)}$ and $Model_{(Ensemble)}$ for the calibration and validation periods. $Model_{(NCEP/NCAR)}$ and $Model_{(Ensemble)}$ had the same calibration period from 1950 to 1989. However, $Model_{(NCEP/NCAR)}$ was validated over the period 1999 to 2010 and $Model_{(Ensemble)}$ was validated for the period 1990 to 1999. This was because the 20^{th} century climate experiment outputs of HadCM3, ECHAM5 and GFDL2.0 terminated at 1999. Both downscaling models reproduced the average of the observed precipitation during the calibration period quite accurately. In validation, both downscaling models over-estimated the average of precipitation. The standard deviation of the observed precipitation was under-estimated by both downscaling models in their calibration and validation phases. However, the severity of the under-estimation of the standard deviation was relatively higher for the $Model_{(Ensemble)}$. The under-estimation of the standard deviation was due to the inherent characteristic of statistical downscaling models in failure to properly explain the entire variance of the predictand (Thripathi *et al.,* 2006). Similar to the standard deviation, the coefficient of variation was also under-predicted by both $Model_{(NCEP/NCAR)}$ and $Model_{(Ensemble)}$, in calibration and validation. The NSE, SANS and R^2 for the $Model_{(NCEP/NCAR)}$ were

comparable in calibration and validation. However, for the $Model_{(Ensemble)}$ the NSE, SANS and R^2 in validation, were quite low in comparison to those statistics in its calibration phase.

Table ii Performances of the two downscaling models in calibration and validation

Table iii provides the seasonal statistics of the precipitation reproduced by the two downscaling models. In all four seasons, both downscaling models perfectly reproduced the average of observed precipitation in their calibration phases. In validation, $Model_{(NCEP/NCAR)}$ showed an over-prediction of average of precipitation in all four seasons and $Model_{(Ensemble)}$ under-estimated the average of precipitation in summer and over-predicted it in other seasons. $Model_{(NCEP/NCAR)}$ and $Model_{(Ensemble)}$ under-estimated the standard deviation of observed precipitation during all seasons in the calibration phase. This was a more noticeable characteristic in the precipitation reproduced by $Model_{(Ensemble)}$. In validation, $Model_{(Ensemble)}$ indicated an over-estimation of the standard deviation of precipitation in autumn, during the rest of the seasons it was under-predicted. $Model_{(NCEP/NCAR)}$ continued to show an under-predicting trend for the standard deviation in validation, in all seasons. The coefficient of variation was also under-estimated by both downscaling models, in all seasons. $Model_{(NCEP/NCAR)}$ showed relatively better NSEs and R^2 values than its counterpart $Model_{(Ensemble)}$ in calibration and validation, during all seasons. $Model_{(Ensemble)}$ exhibited negative NSEs and near zero R^2 values in the validation period for all four seasons.

Table iii Seasonal performances of the two downscaling models

4.5. Reproduction of past observed precipitation with GCM outputs applied to $Model_{(NCEP/NCAR)}$

The $Model_{(NCEP/NCAR)}$ outperformed its counterpart $Model_{(Ensemble)}$ in calibration and validation. However, it does not guarantee that the $Model_{(NCEP/NCAR)}$ will be better than the $Model_{(Ensemble)}$ in precipitation projections produced for future. This is because the $Model_{(NCEP/NCAR)}$ was calibrated and validated with a better quality reanalysis data set, and the projections into future by this downscaling model are produced with the outputs of GCMs pertaining to future. Unlike reanalysis outputs, GCM outputs are associated with higher degree of uncertainty, as they do not undergo any correction against observations. Therefore it was interesting to see how the $Model_{(NCEP/NCAR)}$ would reproduce the past observed precipitation with the 20^{th} century climate experiment outputs of several GCMs. This analysis was important as $Model_{(NCEP/NCAR)}$ is meant to be used with GCM outputs pertaining to future climate for the projection of catchment scale precipitation into future. The 20^{th} century climate experiment outputs of HadCM3, ECHAM5, GFDL2.0 and GFDL2.1 for the period 1950-1999 were standardised with the means and the standard deviations of NCEP/NCAR reanalysis outputs corresponding to the calibration period (1950-1989) of the $Model_{(NCEP/NCAR)}$. Thereafter these standardised GCM outputs were introduced to the $Model_{(NCEP/NCAR)}$ for reproducing the past observed precipitation in the period 1950-1999.

Figures 7c, 7d, 7e and 7f show the precipitation reproduced by $Model_{(NCEP/NCAR)}$ when it was run with the 20^{th} century climate experiment outputs of HadCM3, ECHAM5,

GFDL2.0 and GFDL2.1 for the period 1950-1999, respectively. Also the precipitation reproduced by $Model_{(NCEP/NCAR)}$ with NCEP/NCAR reanalysis data (see Figure 7a) and that of $Model_{(Ensemble)}$ with multi-model ensemble outputs (see Figure 7b) for the period 1950-1999 also shown, as references.

Figure 7 Precipitation reproduced by $Model_{(NCEP/NCAR)}$ and $Model_{(Ensemble)}$ for period 1950-1999

$Model_{(NCEP/NCAR)}$ with NCEP/NCAR outputs showed the least scatter in all six scatter plots in Figure 7. This was numerically proven by the relatively high R^2 value of 0.75. The scatter of the precipitation reproduced by $Model_{(Ensemble)}$ when it was run with the multi-model ensemble outputs was much higher than that of $Model_{(NCEP/NCAR)}$ when it was run with the NCEP/NCAR outputs. However the scatter of precipitation reproduced by $Model_{(NCEP/NCAR)}$ with the 20^{th} century climate experiment outputs of HadCM3, ECHAM5, GFDL2.0 and GFDL2.1 were quite high with a relatively low R^2 values further reinforced this fact. Overall, when $Model_{(NCEP/NCAR)}$ was run with HadCM3, ECHAM5, GFDL2.0 and GFDL2.1 outputs displayed an over-estimating trend for precipitation. This trend was more evident when this downscaling model was run with the 20^{th} century climate experiment outputs of ECHAM5 than with the outputs of any other GCM. According to the scatter plots 7b, 7c, 7d, 7e and 7f, it was seen that the $Model_{(Ensemble)}$ can perform better than its counterpart $Model_{(NCEP/NCAR)}$, when it was run with HadCM3, ECHAM5, GFDL2.0 and GFDL2.1 outputs. The bias prevalent in the outputs of GCMs was one of the reasons for the large scatter exhibited by

$Model_{(NCEP/NCAR)}$ when it was run with the 20th century climate experiment outputs of HadCM3, ECHAM5, GFDL2.0 and GFDL2.1.

Table iv shows the statistics of the precipitation reproduced by the $Model_{(NCEP/NCAR)}$ for the period 1950-1999, when it was run with the 20th century climate experiment outputs of HadCM3, ECHAM5, GFDL2.0 and GFDL2.1. Also the statistics of precipitation reproduced by $Model_{(NCEP/NCAR)}$ with NCEP/NCAR reanalysis outputs and $Model_{(Ensemble)}$ with multi-model ensemble outputs are also shown in the same table. $Model_{(NCEP/NCAR)}$ with NCEP/NCAR reanalysis outputs and $Model_{(Ensemble)}$ with multi-model ensemble outputs reproduced the average of precipitation quite accurately. When the $Model_{(NCEP/NCAR)}$ was run with HadCM3, ECHAM5, GFDL2.0 and GFDL2.1 outputs, the average precipitation was over-estimated by a large margin compared to the observed. This trend was highest when the $Model_{(NCEP/NCAR)}$ was run with the 20th century climate experiment outputs of ECHAM5. $Model_{(NCEP/NCAR)}$ was able to reproduce the standard deviation of precipitation closely, when it was run with HadCM3 outputs and also with the outputs of GFDL2.0. The 20th century climate experiment outputs of ECHAM5, made the $Model_{(NCEP/NCAR)}$ to largely over-estimate the standard deviation of the observed precipitation. $Model_{(NCEP/NCAR)}$ with outputs of ECHAM5 over-estimated the coefficient of variation of precipitation, while all other models under-estimated the coefficient of variation. The minimum of precipitation was correctly produced by the $Model_{(NCEP/NCAR)}$ when it was run with NCEP/NCAR, HadCM3 and GFDL2.0 outputs. The maximum of precipitation was largely over-predicted by $Model_{(NCEP/NCAR)}$ when it was run with outputs of ECHAM5 and slightly

over-predicted when this downscaling model was driven with HadCM3 outputs. $Model_{(Ensemble)}$ severely under-estimated the maximum of the observed precipitation.

When $Model_{(NCEP/NCAR)}$ was run with the NCEP/NCAR outputs, it displayed the highest R^2 value of 0.75 and the highest NSE of 0.75 for the period 1950-1999. In reproducing the observed precipitation of the period 1950-1999, with the 20th century climate experiment outputs of HadCM3, ECHAM5, GFDL2.0 and GFDL2.1, $Model_{(NCEP/NCAR)}$ showed limited performances in comparison with its performances when it was run with the NCEP/NCAR reanalysis outputs. The $Model_{(Ensemble)}$ indicated a better R^2 value of 0.36 and a NSE of 0.35 when it was run with the multi-model ensemble outputs, in comparison to the R^2 values and NSEs produced by $Model_{(NCEP/NCAR)}$ when it was run with the outputs of HadCM3, ECHAM5, GFDL2.0 and GFDL2.1. $Model_{(NCEP/NCAR)}$ showed a NSE of -0.48 (R^2 = 0.03) when it was run with the outputs of GFDL2.1, while with the outputs of GFDL2.0 (i.e. this is the previous version of GDL2.1), it produced a NSE of -0.40 (R^2= 0.15). Due to this relatively limited performance exhibited by GFDL2.1 compared to that of GFDL2.0, only GFDL2.0 was used for producing the multi-model ensemble outputs for the $Model_{(Ensemble)}$.

Table iv Performances of $Model_{(NCEP/NCAR)}$ and $Model_{(Ensemble)}$ with inputs form different sources (1950-1999)

A correlation coefficient analysis performed between the 20th century climate experiment outputs of HadCM3 and NCEP/NCAR reanalysis outputs pertaining to the probable predictors used in this study revealed that all these correlations were quite

weak (e.g. 0.2 – 0.4) during the period 1950-1999. This indicated that there is large bias in the 20th century climate experiment outputs of HadCM3 relative to NCEP/NCAR reanalysis outputs. Hence it was assumed that such large bias is prevalent in the 20th century climate experiment outputs of ECHAM5, GFDL2.0 and GFDL2.1. Therefore it was argued that, even if the potential predictors were separately extracted for $Model_{(Ensemble)}$ considering the correlations between the multi-model ensemble outputs generated using the 20th century climate experiment outputs of HadCM3, ECHAM5 and GFDL2.0, the performances of $Model_{(Ensemble)}$ are unlikely to improve.

4.6. Bias-correction and future projections

Section 4.6.1 provides the details of the statistics of the bias-corrected precipitation outputs of $Model_{(NCEP/NCAR)}$ and $Model_{(Ensemble)}$ for the period 1950-1999. A validation of the bias-correction is provided in Section 4.6.2. The details of the statistics of the bias-corrected precipitation projections of $Model_{(NCEP/NCAR)}$ and $Model_{(Ensemble)}$, pertaining to the future period 2000-2099, are presented in Section 4.6.3.

4.6.1 Bias-correction of past precipitation of $Model_{(NCEP/NCAR)}$ and $Model_{(Ensemble)}$

According to Salvi *et al.* (2011), bias is the disagreement between the GCM outputs and observations. Bias in GCM outputs are due to various assumptions and approximations employed in the structure of the GCM which cause their outputs to deviate from the observations. Sachindra *et al.* (2013b) demonstrated the bias in the raw precipitation output of HadCM3 against the precipitation output of NCEP/NCAR, and emphasised the need of a correction to bias. According to Sharma *et al.* (2007) and Ojha *et al.*

(2012), the bias in GCM outputs should be corrected prior to their subsequent use. The correction of bias can be achieved in two different ways; (1) the correction of bias in raw GCM outputs (prior to downscaling) and (2) the correction of bias in the outputs (e.g. precipitation) of a downscaling model which was run with GCM outputs (following downscaling). The latter bias-correction approach involves less computational cost in comparison with the former method.

Sachindra *et al.* (2013c) found that when the scatter of the variable to be bias-corrected is quite large, any bias-correction technique could possibly lead to marginal improvement of the time series of the variable. In that study, monthly bias-correction (Johnson and Sharma, 2012), nested bias-correction (Johnson and Sharma, 2012) and equidistant quantile mapping (Li *et al.*, 2010) were used for the correction of bias in the monthly precipitation outputs of a downscaling model. There it was found that, even when the scatter of the variable is quite high, the statistical moments of it could be corrected successfully, by using the equidistant quantile mapping technique, but still with almost no improvement in the time series.

In the present study, since the scatter of precipitation reproduced by both downscaling models when they were run with the 20^{th} century climate experiment outputs of GCMs was quite high, it was decided to employ the equidistant quantile mapping technique for the correction of bias. The theory and application of the equidistant quantile mapping technique was well documented in the works of Li *et al.* (2010), Salvi *et al.* (2011) and Sachindra *et al.* (2013c). A brief account on the equidistant quantile mapping technique was included in Section 3.2 of this paper. The equidistant quantile mapping technique

was applied to the precipitation produced by $Model_{(NCEP/NCAR)}$ and $Model_{(Ensemble)}$ for the past period 1950-1999 and for the future period 2000-2099. This bias-correction procedure was performed for each calendar month separately.

Table v shows the statistics of precipitation reproduced by $Model_{(NCEP/NCAR)}$ when it was run with the outputs of HadCM3, ECHAM5, GFDL2.0 and GFDL2.1 and $Model_{(Ensemble)}$ when it was run with the multi-model ensemble outputs, for the period 1950-1999, before and after the bias-correction with the equidistant quantile mapping technique. The precipitation reproduced by $Model_{(NCEP/NCAR)}$ when it was run with the 20^{th} century climate experiment outputs of HadCM3, ECHAM5 and GFDL2.0 were added together and the time series of the average prediction was calculated. In the calculation of the time series of the average prediction of precipitation, predictions produced by $Model_{(NCEP/NCAR)}$ when it was run with the outputs of GFDL2.1 was not included, as GFDL2.1 was not included in the generation of the multi-model ensemble outputs for the $Model_{(Ensemble)}$. This enabled the comparison of the statistics of the average prediction with those produced by $Model_{(Ensemble)}$. The statistics of this average prediction calculated before and after the application of the bias-correction to the outputs of $Model_{(NCEP/NCAR)}$ are also shown in Table v. Note that, in the statistics of the average prediction, the bias-correction was done for the individual outputs of $Model_{(NCEP/NCAR)}$ rather than on the time series of the average prediction. All statistics prior to bias-correction are shown in brackets.

Table v Performances of $Model_{(NCEP/NCAR)}$ and $Model_{(Ensemble)}$ with inputs form different sources (1950-1999), before and after bias-correction

According to Table v, it was seen that for the past climate of the period 1950-1999, equidistant quantile mapping matches the statistical moments (e.g. average) of downscaled precipitation with those of observed precipitation perfectly. This was due to the fact, that in equidistant quantile mapping, the CDF of downscaled precipitation is mapped onto that of observed precipitation. However, in case of the average precipitation prediction calculated from the bias-corrected precipitation of $Model_{(NCEP/NCAR)}$ when it was run with the 20^{th} century climate experiment outputs of HadCM3, ECHAM5 and GFDL2.0, indicated a proper correction only to its average. The other statistics of this time series such as the standard deviation were not properly corrected by the bias-correction, as this time series was calculated from the bias-corrected precipitation outputs of $Model_{(NCEP/NCAR)}$ when it was run with the outputs of HadCM3, ECHAM5 and GFDL2.0, rather than from a bias-corrected version of the average prediction. It was seen that, even prior to the correction of bias, $Model_{(Ensemble)}$ can reproduce the average of precipitation with better accuracy than those obtained when $Model_{(NCEP/NCAR)}$ was run with the outputs of HadCM3, ECHAM5, GFDL2.0 and GFDL2.1 and also the average prediction calculated from the outputs of $Model_{(NCEP/NCAR)}$ when it was run with the outputs of HadCM3, ECHAM5 and GFDL2.0. Before and after the bias-correction, $Model_{(Ensemble)}$ displayed R^2 values higher than those for the $Model_{(NCEP/NCAR)}$ when it was run with the outputs of HadCM3, ECHAM5, GFDL2.0 and GFDL2.1. Hence, it was concluded that $Model_{(Ensemble)}$ when it was run with the multi-model ensemble outputs, is more reliable than $Model_{(NCEP/NCAR)}$ when it was run with GCM outputs. The MLR equations built between the NCEP/NCAR reanalysis outputs and GCM outputs for generating the multi-model

ensemble outputs for $Model_{(Ensemble)}$ are dependent on the performances of the individual members of the ensemble. Therefore, the multi-model ensemble predictions produced into future are depend on the past performances of the individual members of the ensemble (Yun *et al.,* 2005).

It was noticed that, after the bias-correction, the R^2 values and NSEs hardly improved for the precipitation reproduced by the $Model_{(NCEP/NCAR)}$ when it was run with the outputs of HadCM3, ECHAM5, GFDL2.0 and GFDL2.1, in spite the correction of the CDF. This was also true for the $Model_{(Ensemble)}$ when it was with the multi-model ensemble outputs. This indicated that, after the application of equidistant quantile mapping, the time series of precipitation has not improved significantly. Therefore it was realised that although equidistant quantile mapping can correct the statistical moments of a variable (e.g. precipitation), it cannot reduce the scatter of the variable successfully.

Theoretically, the bias correction of each individual GCM's outputs (HadCM3, ECAHM5, GFDL2.0 and GFDL2.1) against NCEP/NCAR reanalysis outputs and then applying them to $Model_{(NCEP/NCAR)}$ should improve the performances of this downscaling model. However, for any improvement to the precipitation reproduced by $Model_{(NCEP/NCAR)}$ run with the individually bias-corrected GCM outputs, the bias-correction should have significantly improved the time series of each output of each GCM. Sachindra *et al.* (2013c) compared the raw precipitation output of HadCM3 against the precipitation output of NCEP/NCAR reanalysis data pertaining to grid point {4,4} in the same atmospheric domain shown in Figure 2, and found that there is large

bias which is characterised by large scatter in the precipitation output of HadCM3. Then it was assumed that there can be such large bias in all predictors of all GCMs. Sachindra *et al.* (2013c) proved that when the bias in a certain variable is large, the bias-correction does not improve its time series. This fact was again proven in Table v where the R^2 values of precipitation have hardly improved in all cases, following the bias-correction. For a significant improvement to the outputs of a downscaling model, the time series of each GCM output should be considerably enhanced, however, the bias-correction techniques are not capable of this yet. Therefore it was realised that even if the bias-correction was applied to each individual GCM output, any improvement to the performance of the $Model_{(NCEP/NCAR)}$ is highly unlikely.

4.6.2 Validation of bias-correction

Although the equidistant quantile mapping technique, as the bias-correction method, matched all statistical moments of the precipitation reproduced by the downscaling models with those of past observed precipitation, how it will function in future climate is not certain. Therefore some validation of the equidistant quantile mapping technique is needed. For this validation, $Model_{(NCEP/NCAR)}$ was run with the outputs of HadCM3 pertaining to the COMMIT GHG emission scenario for the period 2000-2099. The outputs of ECHAM5 pertaining to the COMMIT emission scenario were not readily available for downloading. Therefore, the $Model_{(Ensemble)}$ was not run for the COMMIT emission scenario. The COMMIT GHG emission scenario assumes that the atmospheric GHG concentrations observed at year 2000 remain the same throughout the century (Ojha *et al.,* 2010). Since the atmospheric GHG concentrations observed at year 2000 remain the same throughout the 21st century, it can be assumed that this scenario can

also closely reflect the statistics of the climate in the latter half of the 20th century. Therefore it can be assumed that the statistics of climate downscaled with the GCM outputs (e.g. HadCM3) pertaining to the COMMIT GHG emission scenario for the period 2000-2099 can closely represent the statistics of the past observed climate of the period 1950-1999. Following this assumption, it was decided to run the $Model_{(NCEP/NCAR)}$ with the outputs of HadCM3 relevant to the COMMIT GHG emission scenario for the period 2000-2099, and compare the statistics of bias-corrected precipitation with those of past observed precipitation for the period 1950-1999.

The HadCM3 outputs pertaining to the COMMIT GHG emission scenario for the period 2000-2099 were standardised with the means and the standard deviations of NCEP/NCAR reanalysis outputs corresponding to the period 1950-1989. These were then introduced to the $Model_{(NCEP/NCAR)}$ to project precipitation at the point of interest into future. Then, using the equidistant quantile mapping technique, this precipitation projection for the future was bias-corrected. In this case, the difference between the CDF of precipitation reproduced by $Model_{(NCEP/NCAR)}$ when it was run with the outputs of HadCM3 for the COMMIT GHG emission scenario and the CDF of precipitation reproduced by $Model_{(NCEP/NCAR)}$ when it was run with the outputs of HadCM3 corresponding to the 20th century climate experiment, was added to the bias-corrected version of the latter CDF. This yielded the bias-corrected precipitation projections pertaining to the climate in future.

Table vi shows the statistics of observed precipitation for the period 1950-1999, and also the statistics of precipitation downscaled with the HadCM3 outputs corresponding

to the COMMIT GHG emission scenario for the period 2000-2099, before and after the bias-correction.

Table vi Performances of $Model_{(NCEP/NCAR)}$ with HadCM3 outputs pertaining to the COMMIT GHG emission scenario (2000-2099), before and after bias-correction

According to Table vi, it was seen that there is a large mismatch between the average of observed precipitation and that of precipitation reproduced by the $Model_{(NCEP/NCAR)}$ for COMMIT GHG emission scenario, prior to the application of the bias-correction. However, after the application of the equidistant quantile mapping technique, this mismatch in the average of precipitation was largely reduced. The standard deviation of precipitation produced by the downscaling model with HadCM3 COMMIT outputs was in close agreement with that of observations, despite slight over-estimation. Following the bias-correction, the coefficient of variation was successfully corrected to match with that of observed precipitation. The minimum of precipitation was slightly over-estimated by the downscaling model prior to bias-correction; however this was perfectly corrected by the bias-correction technique. The maximum of precipitation was further over-estimated by the bias-correction. Based on the above comparison, it was realised that, after the application of the equidistant quantile mapping technique, the statistics of the precipitation produced by $Model_{(NCEP/NCAR)}$ when it was run with the outputs of HadCM3 pertaining to the COMMIT GHG emission scenario could closely resemble those of past observed precipitation. This indicated that, the equidistant quantile mapping technique is also capable in bias-correcting the precipitation pertaining to the

future climate, hence, it was assumed that it will correct the bias in precipitation projections produced by the two downscaling models into future.

4.6.3 Bias-corrected future precipitation projections

For the projection of precipitation into future, the A2 and the B1 GHG emission scenarios defined by the IPCC were selected. The A2 GHG emission scenario refers to a future world with rapid economic growth. Hence it is associated with relatively high levels of GHG emissions. The B1 GHG emission scenario refers to a world with high level of focus on the environment and sustainable development; therefore it refers to relatively low levels of atmospheric GHGs. The A2 and B1 GHG emission scenarios referred to carbon dioxide concentrations of about 850 ppm and 550 ppm respectively, by the end of the 21st century (IPCC, 2000). Due to the distinct differences in the A2 and B1 emission scenarios, it was assumed that the precipitation projections derived from them could depict a diverse range in the precipitation regime under changing climate.

The monthly precipitation at the station of interest was produced for the period 2000-2099, by introducing the standardised outputs of HadCM3, ECHAM5, GFDL2.0 and GFDL2.1 pertaining to A2 and B1 GHG emission scenarios to the $Model_{(NCEP/NCAR)}$. Also, the multi-model ensemble outputs produced with the outputs of HadCM3, ECHAM5, and GFDL2.0 corresponding to A2 and B1 GHG emission scenarios were introduced to the $Model_{(Ensemble)}$ for the projection of precipitation into the future period 2000-2099. Following the procedure described in Section 3.2, the precipitation produced by $Model_{(NCEP/NCAR)}$ when it was run with outputs of HadCM3, ECHAM5,

GFDL2.0 and GFDL2.1 and the precipitation produced by $Model_{(Ensemble)}$ when it was run with multi-model ensemble outputs, for the future period 2000-2099, were bias-corrected. The procedure for the bias-correction of precipitation pertaining to the future climate was exactly the same as that used in the validation of the bias-correction described in Section 4.6.2.

Table vii shows the percentage changes of the statistics of the bias-corrected precipitation for the period 2000-2099, with respect to the statistics of the observed precipitation of the period 1950-1999. These precipitation projections into future were produced by the $Model_{(NCEP/NCAR)}$ with the outputs of different GCMs and by the $Model_{(Ensemble)}$ with the multi-model ensemble outputs, for the future period 2000-2099, under the A2 and the B1 GHG emission scenarios. The precipitation projections (into future) produced by $Model_{(NCEP/NCAR)}$ when it was run with the outputs of HadCM3, ECHAM5 and GFDL2.0 (after bias-correction) were added together to produce time series of average precipitation, for both A2 and the B1 GHG emission scenarios. The percentage changes in the statistics of these time series with respect to the statistics of the observed precipitation of the period 1950-1999 were also added to Table vii (see "*Average of* $Model_{(NCEP/NCAR)}$" in Table vii). These percentage changes in the statistics were compared with those of precipitation outputs produced by $Model_{(Ensemble)}$, at the end of this section.

According to Table vii, $Model_{(NCEP/NCAR)}$ indicated a drop in the average of precipitation in summer when it was run with HadCM3 and GFDL2.1 outputs pertaining to the A2 and B1 GHG emission scenarios. Similarly, in autumn a decline in the average of

precipitation was indicated by $Model_{(NCEP/NCAR)}$ with the outputs of GFDL2.0 and GFDL2.1 under both A2 and B1 scenarios. HadCM3 and GFDL2.1 outputs on $Model_{(NCEP/NCAR)}$ produced an increase in the average of precipitation in winter for both emission scenarios. When $Model_{(NCEP/NCAR)}$ was run with the outputs of HadCM3 and ECHAM5, it showed a decline in the average of precipitation for A2 and B1 scenarios, in spring. It was seen that, when $Model_{(NCEP/NCAR)}$ was run with the outputs of different GCMs, it tended to produce mixed results for the average of the precipitation corresponding to future. When $Model_{(Ensemble)}$ was run with the multi-model ensemble outputs, it showed an increase in the average of precipitation in summer, autumn and winter. However, this rise in the average of precipitation in autumn and winter was quite limited. It was concluded that the average of precipitation produced by the downscaling models, pertaining to the future, is dependent on the inputs introduced to it.

$Model_{(NCEP/NCAR)}$ with HadCM3, ECHAM5 and GFDL2.0 outputs showed a rising trend in the standard deviation of precipitation in summer, for both A2 and B1. In autumn, $Model_{(NCEP/NCAR)}$ showed a decrease in the standard deviation of precipitation when it was run with the outputs of GFDL2.0 and GFDL2.1, for both emission scenarios. In winter, $Model_{(NCEP/NCAR)}$ indicated an increase in the standard deviation of precipitation with the outputs of all GCMs, for both emission scenarios. In spring, except with the outputs of ECHAM5, $Model_{(NCEP/NCAR)}$ indicated a rise in the standard deviation of precipitation for both scenarios with the outputs of all other GCMs. Overall, it was seen that for the majority of the cases, $Model_{(NCEP/NCAR)}$ shows a rise in the standard deviation of precipitation under both emission scenarios. Meanwhile $Model_{(Ensemble)}$

displayed a drop in the standard deviation of the precipitation during all four seasons under both A2 and B1 scenarios.

Apart from GFDL2.1 which showed a decrease in the maximum monthly precipitation produced by $Model_{(NCEP/NCAR)}$, in summer and autumn under B1 scenario, all other GCMs indicated an increase in the maximum monthly precipitation produced by $Model_{(NCEP/NCAR)}$ in all seasons, for both emission scenarios. This showed that the severity of maximum monthly precipitation values tends to increase in future with the rising GHG concentrations in the atmosphere. However, for both emission scenarios, $Model_{(Ensemble)}$ displayed a decrease in the maximum monthly precipitation in summer, autumn and winter and an increase in it only in spring. It was concluded that, in spring, there is greater likelihood for maximum monthly precipitation to increase.

In autumn and spring, $Model_{(NCEP/NCAR)}$ showed an increase in the percentage of months with zero precipitation with the outputs of all GCMs. In winter, $Model_{(NCEP/NCAR)}$ displayed limited changes in the percentage of months with zero precipitation for both emission scenarios, when it was run with the outputs of all GCMs. $Model_{(Ensemble)}$ showed a decrease in the percentage of months with zero precipitation only in summer and it indicated no change in the percentage of months with zero precipitation in the other seasons. Considering the outputs of $Model_{(NCEP/NCAR)}$ and $Model_{(Ensemble)}$ it can be concluded that, in future, the percentage of months with zero precipitation is unlikely to decrease in autumn, winter and spring.

HadCM3, ECHAM5 and GFDL2.0 outputs displayed a decrease in the percentage of months with above average precipitation in summer under A2 and B1 emission scenarios when those were applied on the $Model_{(NCEP/NCAR)}$. In spring, a rising trend was exhibited by the both $Model_{(NCEP/NCAR)}$ and $Model_{(Ensemble)}$ with the outputs of all GCMs, for both GHG emission scenarios. Hence, it was realised that, in spring, the percentage of months with above average precipitation is likely to decrease in future.

The average projections of precipitation calculated from the outputs of $Model_{(NCEP/NCAR)}$ when it was run with the outputs of HadCM3, ECHAM5 and GFDL2.0, indicated a drop in the average of precipitation in autumn and spring, for both emission scenarios. However, $Model_{(Ensemble)}$ showed a drop in the average of precipitation only in spring, under B1 emission scenario. It was seen that there is little agreement between the averages of precipitation of the average projections computed from the outputs of $Model_{(NCEP/NCAR)}$ and the projections produced by $Model_{(Ensemble)}$. However, the average projections computed from the outputs of $Model_{(NCEP/NCAR)}$ and the projections produced by $Model_{(Ensemble)}$ indicated a decline in the standard deviation in all seasons, under both A2 and B1 emission scenarios. In summer, autumn and winter, the average projections obtained from the outputs of $Model_{(NCEP/NCAR)}$ and the projections of $Model_{(Ensemble)}$ showed a decrease in the magnitude of the maximum monthly precipitation, for both emission scenarios. In summer, the average projections obtained from the outputs of $Model_{(NCEP/NCAR)}$ and the projections produced by $Model_{(Ensemble)}$ indicated a decrease in the percentage of months with zero precipitation and in other seasons no change in the percentage of months with zero precipitation was seen, under A2 and B1 emission scenarios. The average projections obtained from the outputs of $Model_{(NCEP/NCAR)}$ and

the projections produced by $Model_{(Ensemble)}$ indicated relatively small changes in the percentage of months with above average precipitation, in all seasons, for both A2 and B1 emission scenarios. It was seen that, the average projections obtained from the outputs of $Model_{(NCEP/NCAR)}$ when it was run with the outputs of HadCM3, ECHAM5 and GFDL2.0 and the projections produced by $Model_{(Ensemble)}$ when it was run with the multi-model ensemble outputs show similar trends in the standard deviation, the monthly maximum precipitation and the percentage of months with zero precipitation, in the majority of seasons, over the period 2000-2099.

For the management of water resources, the knowledge of long-term statistics of monthly precipitation such as average, variance, extremes etc, in advance, is of great importance. The average of the precipitation is an indication of the availability of water in the catchment and the variance provides the information on the possible fluctuations in the precipitation regime. The knowledge of the extreme precipitation is useful in the effective management of floods and droughts. Since the observation station at Halls Gap post office is located in an important agricultural and water supply area, the point specific precipitation projections produced in this study under different GCMs and different GHG emission scenarios will aid the water resources managers in planning and management of water resources.

Table vii Percentage changes of statistics of bias-corrected precipitation of $Model_{(NCEP/NCAR)}$ $Model_{(Ensemble)}$ and for period 2000-2099, with respect to statistics of observed precipitation of period 1950-1999

According to the median estimates obtained from the raw precipitation outputs of number of GCMs under B1 (low emissions), and A1F1 (high emissions) emission scenarios, the average of precipitation over the Wimmera region which include the present study area has indicated a decline in all seasons by year 2070. (Victorian Government Department of Sustainability and Environment, 2008). Furthermore, Smith and Chandler (2009) commented that raw precipitation outputs of HadCM3 under the A1B emission scenario (medium emissions) indicates a decline in the average of precipitation of about 15% in the period 2071-2099 in comparison to the average of observed precipitation pertaining to the period 1971-2000, over the Murray Darling basin in south east Australia. The present precipitation station at Halls Gap post office is also located within the Murray Darling basin and according to the present study the precipitation downscaled from the outputs of HadCM3 under A2 and B1 emission scenarios at this station for the period 2071-2099, showed a decrease in the average of precipitation of about 12% and 3.4% respectively in comparison with observations of the period 2071-2099. It is noteworthy to state that there was no evidence of any statistical downscaling exercise conducted in the study area in published literature, prior to the present study.

5. SUMMARY AND CONCLUSIONS

The common practice in building a statistical downscaling model is to perform the calibration and validation (development) a with some reanalysis outputs pertaining to the past climate and then produce the catchment scale projection into future by introducing outputs of a GCM corresponding to possible future climate. The major issue associated with this approach is that, in the calibration and validation phases, reanalysis

data are used as inputs to the downscaling model, while for producing projections into future, GCM outputs are used as inputs to the model. The two sets of input data for the downscaling model are obtained from two different sources and hence there is no homogeneity in them. The reanalysis outputs are quality controlled and corrected against observations and are of better quality than GCM outputs. Therefore when a downscaling model is developed in the above manner, it tends to perform better in calibration and validation, and its performances in future are doubtful.

Two statistical downscaling models for downscaling monthly GCM outputs to catchment scale precipitation were developed in this research using the multi-linear regression technique. The precipitation station located at Halls Gap post office in north western Victoria, Australia was selected as the case study. The first downscaling model ($Model_{(NCEP/NCAR)}$) was developed following the common practice, with the NCEP/NCAR reanalysis outputs and the outputs of HadCM3, ECHAM5, GFDL2.0 and GFDL2.1 pertaining to future climate were applied on this model for producing the catchment scale projections of precipitation into future. For the development of this downscaling model, a pool of probable predictors was selected considering the past literature and hydrology. The data for probable predictors obtained from the NCEP/NCAR reanalysis archive and observed precipitation data were split into three time slices; 1950-1969, 1970-1989, and 1990-2010. Then the probable predictors which displayed the best statistically significant correlations ($p \leq 0.05$) consistently with observed precipitation in all time slices and the whole period were selected as potential predictors, for each calendar month. Initially, the three best potential predictors which showed the best correlations with observed precipitation over the whole period of the

study (1950-2010) were introduced to the downscaling model and it was calibrated for the period 1950-1989 and validated for the period 1990-2010. In the same manner, the next best potential predictors were added to the downscaling model one at a time. This procedure of stepwise addition of predictors was continued until the model performance in terms of Nash-Sutcliffe efficiency in validation reaches a maximum. This way, the best potential predictors for each calendar month were identified.

The best sets of potential predictors identified during the development of the first downscaling model were also used in the second downscaling model. Therefore, in the calibration (1950-1989) and validation (1990-1999) of the second downscaling model, the stepwise addition of predictors was not practised. The second downscaling model ($Model_{(Ensemble)}$) was calibrated and validated with a set of data derived by regressing the 20^{th} century climate experiment outputs of HadCM3, ECHAM5 and GFDL2.0 corresponding to the best potential predictors with the reanalysis outputs (NCEP/NCAR) using the multi-linear regression technique. The outputs of these multi-linear regression equations were called the multi-model ensemble outputs in this paper. The same regression equations developed between the outputs of HadCM3, ECHAM5 and GFDL2.0 and outputs of NCEP/NCAR for the past climate were used to generate the multi-model ensemble outputs which are employed as inputs to this downscaling model for producing the projections of precipitation into future. This way, the second downscaling model was developed and projections into future were made with a homogeneous set of inputs to it.

Conclusions drawn from this study are:

1. The $Model_{(NCEP/NCAR)}$ with the NCEP/NCAR reanalysis outputs performed well in both calibration and validation. The $Model_{(Ensemble)}$ displayed relatively limited performances in calibration and validation. The limited performances of this model were more pronounced in its validation phase.

2. Both $Model_{(NCEP/NCAR)}$ and $Model_{(Ensemble)}$ showed a trend of under-predicting high precipitation values and over-predicting near zero precipitation values in their calibration and validation phases.

3. In reproducing the past observed precipitation of the period 1950-1999, with the 20^{th} century climate experiment outputs of HadCM3, ECHAM5, GFDL2.0 and GFDL2.1, $Model_{(NCEP/NCAR)}$ showed much limited performances, in comparison with its performances when it was run with the NCEP/NCAR reanalysis outputs.

4. When $Model_{(Ensemble)}$ was run with the multi-model ensemble outputs, it indicated relatively better performances in reproducing the past observed precipitation of the period 1950-1999, in comparison with the performances of $Model_{(NCEP/NCAR)}$ when it was run with the 20^{th} century climate experiment outputs of HadCM3, ECHAM5, GFDL2.0 and GFDL2.1, for the same period.

5. According to the comparison of performances of $Model_{(NCEP/NCAR)}$, when it was run with NCEP/NCAR reanalysis outputs and the 20^{th} century climate experiment outputs of HadCM3, ECHAM5, GFDL2.0 and GFDL2.1, it was

realised that there is a large mismatch between the reanalysis outputs and GCM outputs.

6. When $Model_{(NCEP/NCAR)}$ was run with the 20^{th} century climate experiment outputs of HadCM3, ECHAM5, GFDL2.0 and GFDL2.1, the averages of precipitation were over-estimated in the period 1950-1999. However, $Model_{(Ensemble)}$ was able to capture the average of observed precipitation with a better degree of accuracy in this period.

7. It was seen that following the bias-correction the average, the standard deviation, the coefficient of variation, the minimum and the maximum of the monthly precipitation reproduced by $Model_{(NCEP/NCAR)}$ with the 20^{th} century climate experiment outputs of HadCM3, ECHAM5, GFDL2.0 and GFDL2.1 and those reproduced by $Model_{(Ensemble)}$ with the multi-model ensemble outputs over the period 1950-1999, were near-perfectly corrected by the equidistant quantile mapping technique. However still $Model_{(Ensemble)}$ possesses the advantage of using homogeneous inputs to it during its calibration, validation and future projection phases, unlike its counterpart $Model_{(NCEP/NCAR)}$. Also the methodology used in the development of $Model_{(Ensemble)}$ enables the use of outputs from different GCMs to produce a single point scale precipitation projection into future.

8. When $Model_{(NCEP/NCAR)}$ was run with the outputs of HadCM3, ECHAM5, GFDL2.0 and GFDL2.1 pertaining to the A2 and B1 greenhouse gas emission

scenarios, it tended to produce a wide mix of results (increases and decreases) for the average of the precipitation corresponding to the future period 2000-2099, in comparison with that of observed precipitation of the period 1950-1999. When $Model_{(Ensemble)}$ was run with the multi-model ensemble outputs for the same period, it showed an increase in the average of precipitation in summer, autumn and winter in future, in comparison with that of observed precipitation of period 1950-1999. It was realised that the average of precipitation projected into future is dependent on the source of inputs to the downscaling model.

9. When $Model_{(NCEP/NCAR)}$ was run with the outputs of HadCM3, ECHAM5, GFDL2.0 and GFDL2.1, and when $Model_{(Ensemble)}$ was run with multi-model ensemble outputs, for both A2 and B1 emission scenarios, a rise in the magnitude of the maximum monthly precipitation was seen, in spring. It was concluded that, in future, during spring, there is greater likelihood for the magnitude of the maximum monthly precipitation to increase.

10. According to the precipitation projection produced by $Model_{(NCEP/NCAR)}$ and $Model_{(Ensemble)}$ into the future period 2000-2099, it was seen that the percentage of months with zero precipitation is unlikely to decrease in autumn, winter and spring.

11. In the majority of seasons, the average projections obtained from the outputs of $Model_{(NCEP/NCAR)}$ when it was run with the outputs of HadCM3, ECHAM5 and GFDL2.0 and the projections produced by $Model_{(Ensemble)}$ when it was run with

the multi-model ensemble outputs showed similar trends (increase/decrease) in the standard deviation, the monthly maximum precipitation and the percentage of months with zero precipitation, over the period 2000-2099.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the financial assistance provided by the Australian Research Council Linkage Grant scheme and 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

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

Christensen JH, Kjellström E, Giorgi F, Lenderink G, Rummukainen M. 2010. Weight assignment in regional climate models. *Climate Res.* 44: 179-194. DOI: 10.3354/cr00916

Commonwealth Scientific and Industrial Research Organisation. 2007. Wimmera region fact sheet: Murray-Darling basin sustainable yields project. Available at http://www.csiro.au/Outcomes/Water/Water-for-the-environment/Wimmera-region-

fact-sheet-Murray-Darling-Basin-Sustainable-Yields-Project.aspx. (Accessed on 20th April 2013).

Fealy R, Sweeney J. 2008. Statistical downscaling of temperature, radiation and potential evapotranspiration to produce a multiple GCM ensemble mean for a selection of sites in Ireland. *Irish Geography* 41: 1-27. DOI: 10.1080/00750770801909235

Huth R. 2002. Statistical downscaling of daily temperature in central Europe. *J. Climate* 15: 1731-1742. DOI: 10.1175/1520-0442(2002)015

Ingol-Blanco EM. 2011. '*Modeling Climate Change Impacts on Hydrology and Water Resources: Case Study Rio Conchos Basin',* PhD Thesis, pp61-63.The University of Texas, Austin, USA.

IPCC 2007. Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds.)]. Cambridge University Press: Cambridge; pp 996.

IPCC. 2000. '*IPCC special report on emissions scenarios - Summary for policymakers*'. Online report available at http://www.ipcc.ch/pdf/special-reports/spm/sres-en.pdf. 4-8.

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

Jeong DI, St-Hilaire A, Ouarda TBMJ, Gachon P. 2012. A multi-site statistical downscaling model for daily precipitation using global scale GCM precipitation outputs. *Int. J. Climatol.* Article in Press. DOI: 10.1002/joc.3598

Johnson F, Sharma A. 2012. A nesting model for bias correction of variability at multiple time scales in general circulation model precipitation simulations. *Water Resour. Res.* 48: 1-16. DOI: 10.1029/2011WR010464

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. *B. Am. Meteorol. Soc.* 77: 437-471. DOI: 10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2

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. *J. Climate.* 3: 1053-1079. DOI: 10.1175/1520-0442(1990)003<1053:AMORGC>2.0.CO;2

Kharin VV, Zwiers FW, Gagnon N. 2001. Skill of seasonal hindcasts as a function of the ensemble size. *Clim. Dynam.* 17: 835-843. DOI: 10.1007/s003820100149

Kharin VV, Zwiers FW. 2002. Climate predictions with multimodel ensembles. *J. Climate.* 15: 793-799. DOI: 10.1175/1520-0442(2002)015<0793:CPWME>2.0.CO;2

King LM, Irwin S, Sarwar R, McLeod AI, Simonovic SP. 2012. *Can. Water Res. J.* 37: 253-274. DOI: 10.4296/cwrj2011-938

Knutti R, Furrer R, Tebaldi C, Cermak J, Meehl GA. 2010. Challenges in Combining Projections from Multiple Climate Models. *J. Climate.* 23: 2739-2758, DOI: 10.1175/2009jcli3361.1

Krishnamurti TN, Kishtawal CM, LaRow TE, Bachiochi DR, Zhang Z, Williford CE, Gadgil S, Surendran S. 1999. Improved weather and seasonal climate forecasts from multimodel superensemble. *Science.* 285: 1548-1550. DOI: 10.1126/science.285.5433.1548

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. *J. Geophys. Res-Atmos.* 115: 1-20. DOI: 10.1029/2009JD012882

Maqsood I, Khan MR, Abraham A. 2004. An ensemble of neural networks for weather forecasting. *Neural Comput. Appl.* 13: 112-122. DOI: 10.1007/s00521-004-0413-4

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

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

Ojha CSP, Goyal MK, Adeloye AJ. 2010. Downscaling of precipitation for lake catchment in arid region in India using linear multiple regression and neural networks. *The Open Hydrology Journal* 4: 122-136. DOI: 10.2174/1874378101004010122.

Ojha R, Kumar DN, Sharma A, Mehrotra R. 2012. Assessing severe drought and wet events over India in a future climate using a nested bias correction approach. *J. Hydrologic Eng.* Article in press. DOI: 10.1061/(ASCE)HE.1943-5584.0000585

Pearson K. 1895. Mathematical contributions to the theory of evolution. iii. regression heredity and panmixia. *Philos. T. Roy. Soc. A* 187: 253-318. DOI: 10.1098/rsta.1896.0007

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

Sachindra DA, Huang F, Barton AF, Perera BJC. 2013b. Statistical Downscaling of General Circulation Model Outputs to Precipitation Part 1: Calibration and Validation. *Int. J. Climatol.* accepted with revison.

Sachindra DA, Huang F, Barton AF, Perera BJC. 2013c. Statistical Downscaling of General Circulation Model Outputs to Precipitation Part 2: Bias-correction and future projections. *Int. J. Climatol.* accepted with revison.

Salvi K, Kannan S, Ghosh S. 2011. 'Statistical downscaling and bias-correction for projections of Indian rainfall and temperature in climate change studies'. In proceeding of the *4th International Conference on Environmental and Computer Science*, Singapore, 16 Sep - 18 Sep 2011.

Sharma D, Gupta AD, Babel MS. 2007. Spatial disaggregation of bias-corrected GCM precipitation for improved hydrologic simulation: Ping River Basin, Thailand. *Hydrol. Earth Syst. Sci.* 11: 1373-1390. DOI: 10.5194/hess-11-1373-2007

Smith I, Chandler E. 2009. Refining rainfall projections for the Murray Darling basin of south-east Australia-the effect of sampling model results based on performance. *J. Climate Change* 102: 377-393. DOI: 10.1007/s10584-009-9757-1

Timbal B, Fernandez E, Li Z. 2009. Generalization of a statistical downscaling model to provide local climate change projections for Australia. *Environ. Modell. Softw.* 24: 341-358. DOI: 10.1016/j.envsoft.2008.07.007

Tripathi S, Srinivas VV, Nanjundiah RS. 2006. Downscaling of precipitation for climate change scenarios: a support vector machine approach. *J. Hydrol.* 330: 621-640. DOI: 10.1016/j.jhydrol.2006.04.030

Victorian Government Department of Sustainability and Environment. 2008. Climate change in the Wimmera. Available at http://www.climatechange.vic.gov.au/regional-projections/wimmera. (Accessed on 20th April 2013). 3-8.

Wang W. 2006. *Stochasticity, Nonlinearity and Forecasting of Streamflow Processes*. Deft University Press: Amsterdam; pp 72-73.

Warner TT. 2011. Numerical weather and climate prediction. Cambridge University Press: Cambridge; pp 252-283.

Wilby RL, Wigley TML. 2000. Precipitation predictors for downscaling: observed and general circulation model relationships. *Int. J. Climatol.* 20: 641-661. DOI: 10.1002/(SICI)1097-0088(200005)

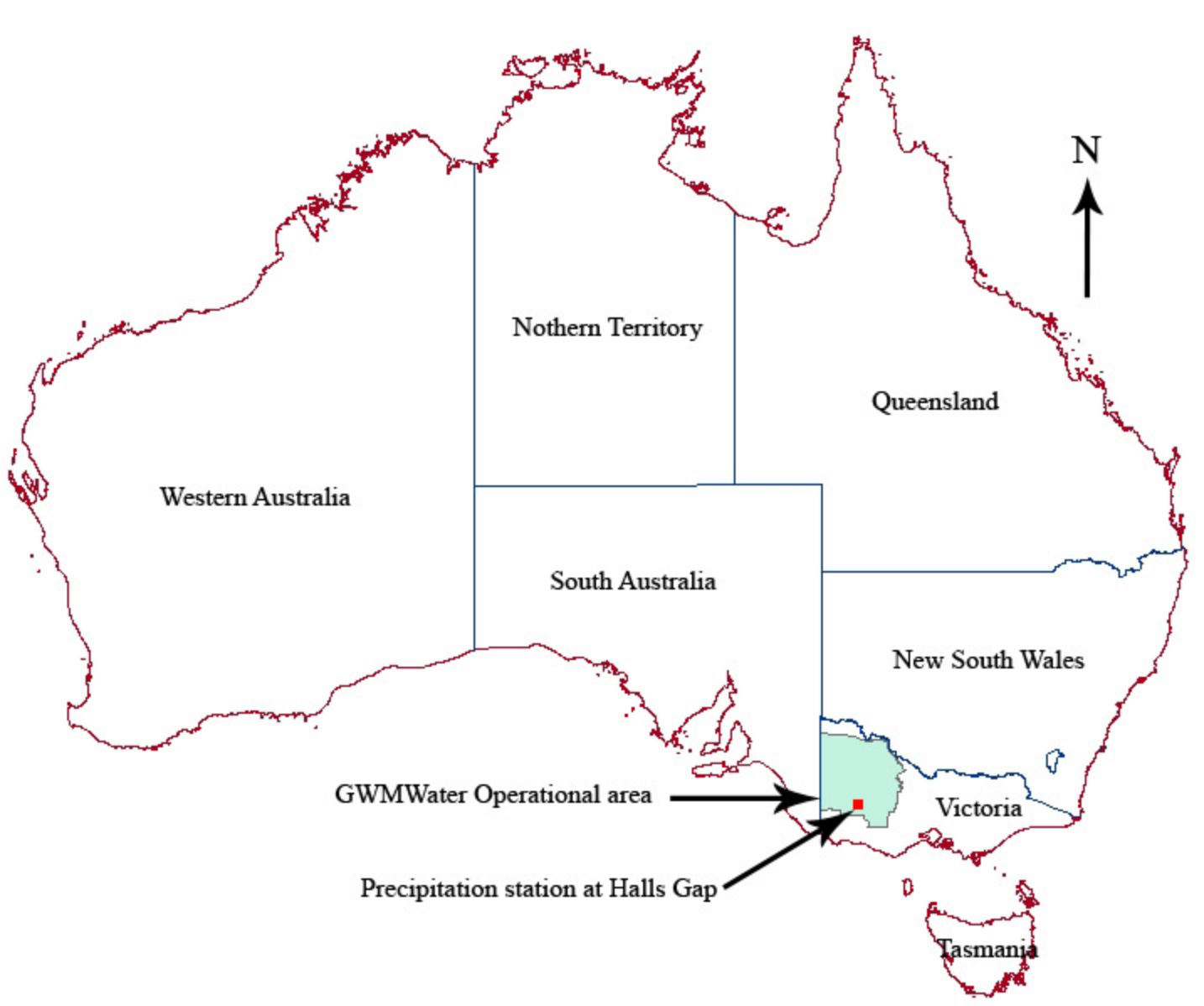
Willems P, Olsson J, Arnbjerg-Nielsen K, Beecham S, Pathirana A, Bülow Gregersen I, Madsen H, Nguyen VTV. 2012. Impacts of climate change on rainfall extremes and urban drainage systems. IWA publishing: London; pp 58-109.

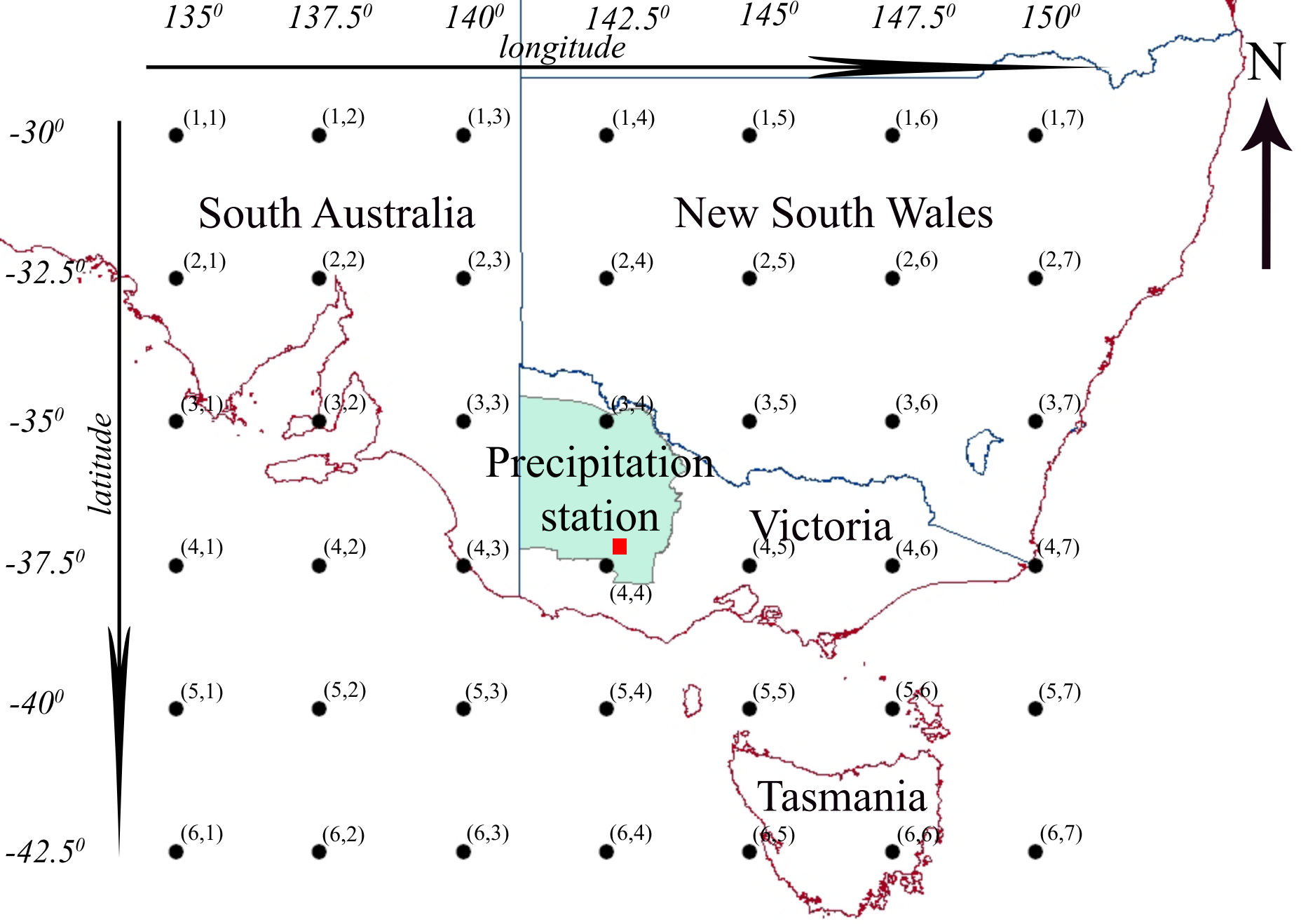
Yu PS, Yang TC, Wu CK. 2002. Impact of climate change on water resources in southern Taiwan. *J. Hydrol.* 260: 161-175. DOI: 10.1016/S0022-1694(01)00614-X

Yun WT, Stefanova L, Krishnamurti TN. 2003. Improvement of the multimodel superensemble technique for seasonal forecasts. *J. Climate.* 16: 3834-3840. DOI: 10.1175/1520-0442(2003)016<3834:IOTMST>2.0.CO;2

Yun WT, Stefanova L, Mitra AK, Kumar TSV, Dewar W, Krishnamurti TN. 2005. A multi-model superensemble algorithm for seasonal climate prediction using DEMETER forecasts. *Tellus A* 57: 280-289. DOI: 10.1111/j.1600-0870.2005.00131.x

Zhang H, Huang GH. 2013. Development of climate change projections for small watersheds using multi-model ensemble simulation and stochastic weather generation. *Clim. Dynam.* 40: 805-821. DOI: 10.1007/s00382-012-1490-1


N
Nothern Territory
Queensland
Western Australia
South Australia
New South Wales
GWMWater Operational area
Victoria
Precipitation station at Halls Gap
Tasmania


135⁰
137.5⁰
140⁰
142.5⁰
145⁰
147.5⁰
150⁰
longitude
N
-30⁰
-32.5⁰
-35⁰
-37.5⁰
-40⁰
-42.5⁰
latitude
South Australia
New South Wales
Victoria
Tasmania
Precipitation station
(1,1)
(1,2)
(1,3)
(1,4)
(1,5)
(1,6)
(1,7)
(2,1)
(2,2)
(2,3)
(2,4)
(2,5)
(2,6)
(2,7)
(3,1)
(3,2)
(3,3)
(3,4)
(3,5)
(3,6)
(3,7)
(4,1)
(4,2)
(4,3)
(4,4)
(4,5)
(4,6)
(4,7)
(5,1)
(5,2)
(5,3)
(5,4)
(5,5)
(5,6)
(5,7)
(6,1)
(6,2)
(6,3)
(6,4)
(6,5)
(6,6)
(6,7)

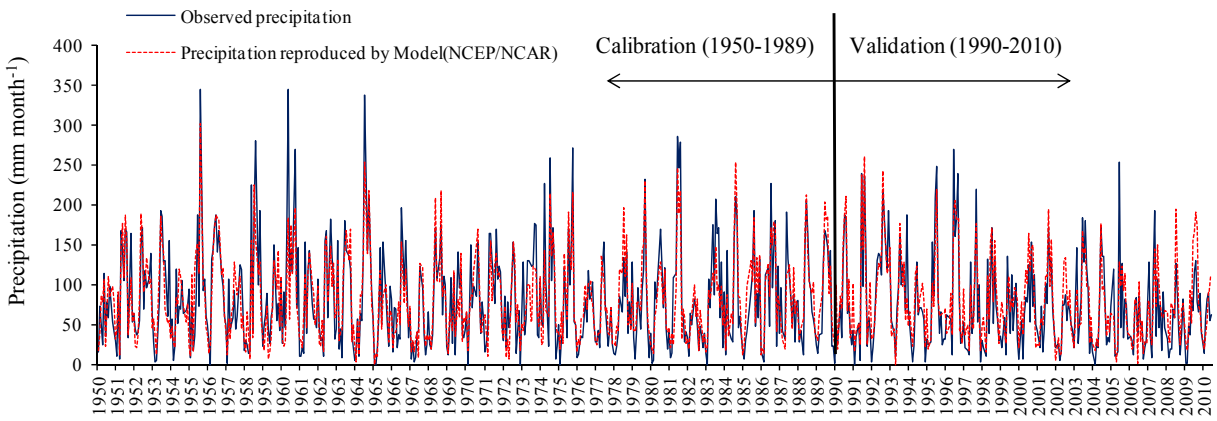


Figure 3 Observed and Model$_{(NCEP/NCAR)}$ reproduced monthly precipitation (1950 to 2010)

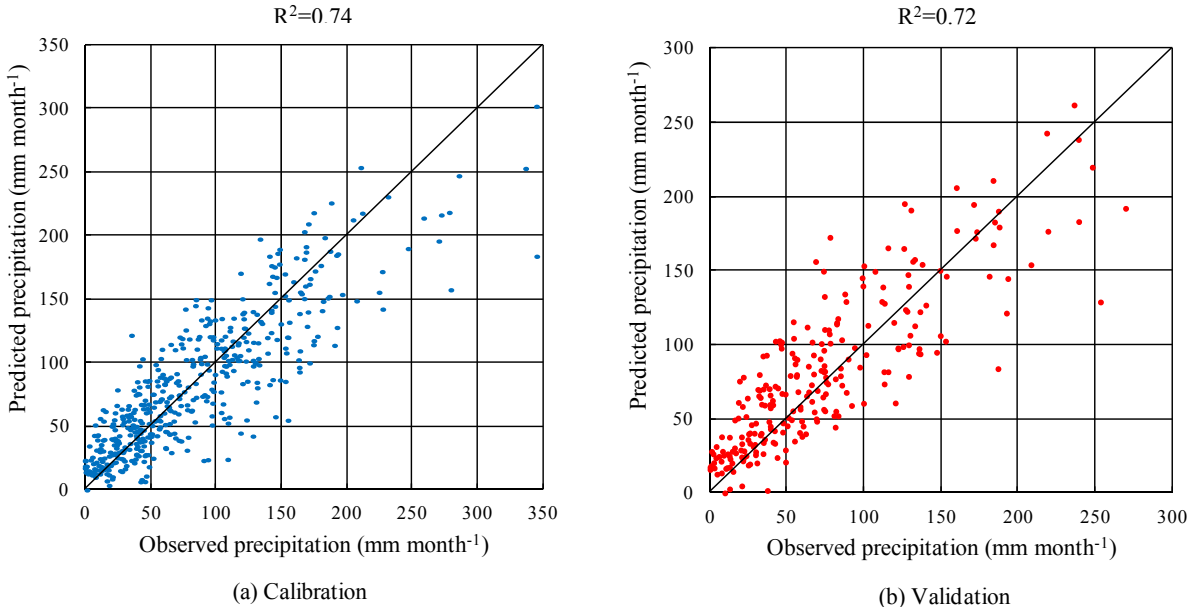


Figure 4 Scatter plots of observed and Model$_{(NCEP/NCAR)}$ reproduced monthly precipitation for calibration (1950-1989) and validation (1990-2010)

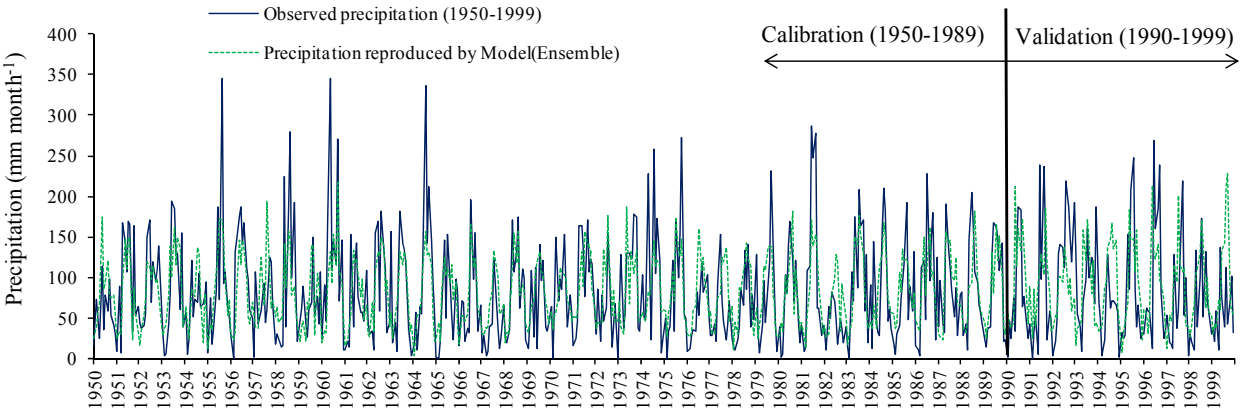


Figure 5 Observed and Model$_{(Ensemble)}$ reproduced monthly precipitation (1950 to 1999)

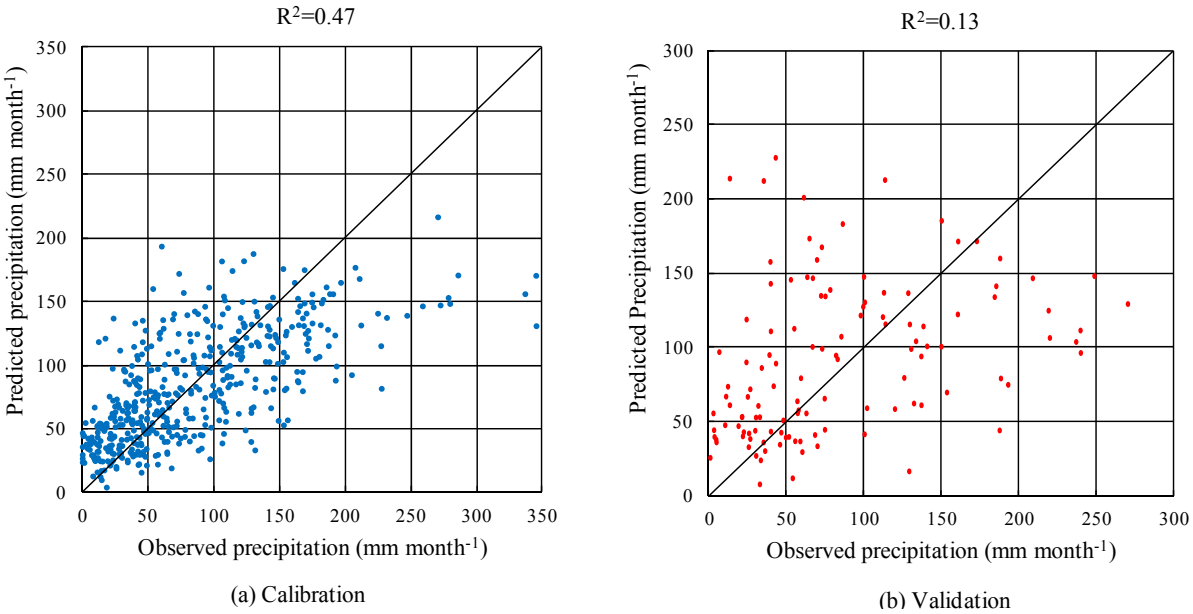
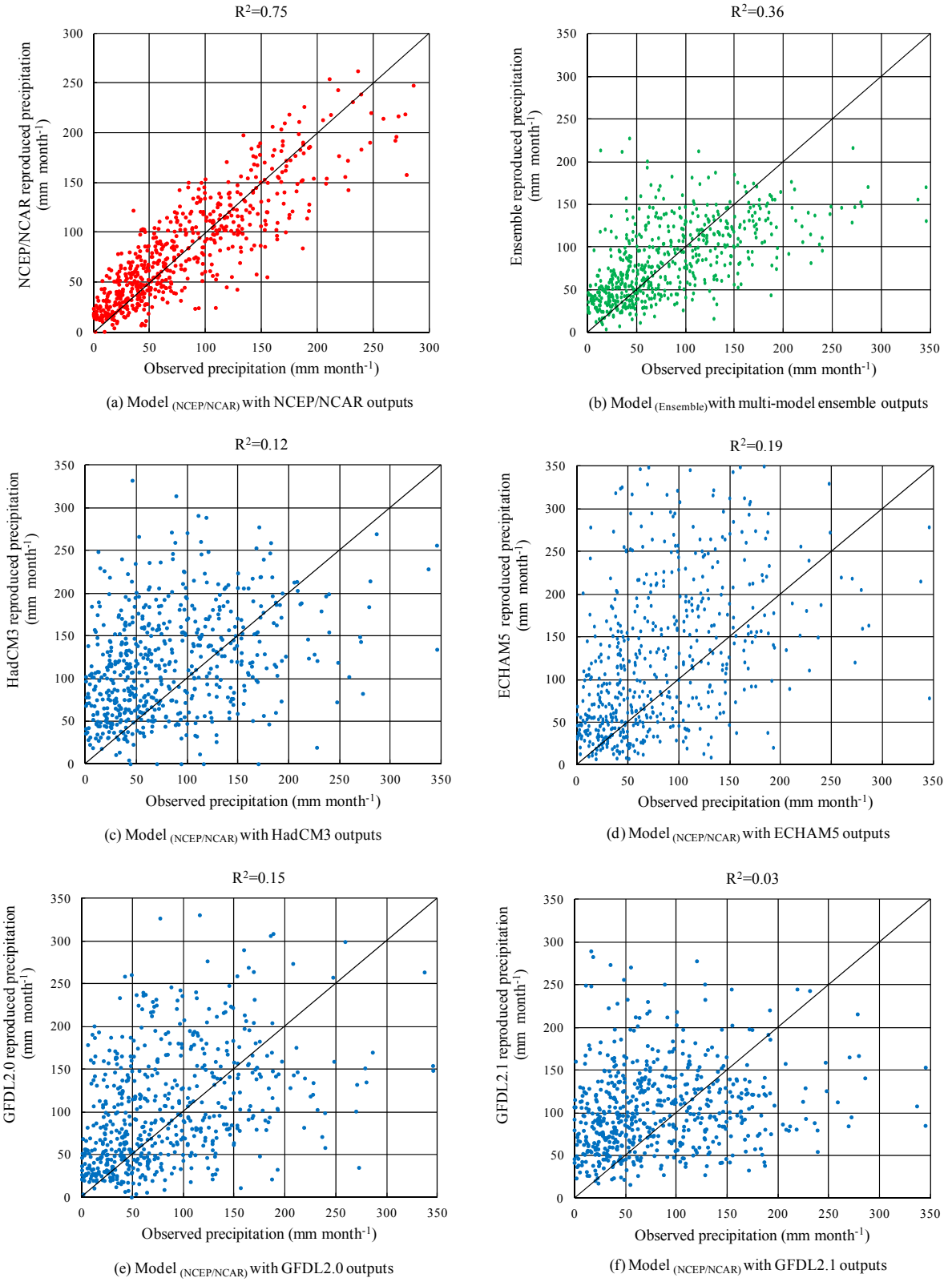


Figure 6 Scatter plots of observed and Model$_{(Ensemble)}$ reproduced monthly precipitation for calibration (1950-1989) and validation (1990-1999)



(a) Model $_{(NCEP/NCAR)}$ with NCEP/NCAR outputs

(b) Model $_{(Ensemble)}$ with multi-model ensemble outputs

(c) Model $_{(NCEP/NCAR)}$ with HadCM3 outputs

(d) Model $_{(NCEP/NCAR)}$ with ECHAM5 outputs

(e) Model $_{(NCEP/NCAR)}$ with GFDL2.0 outputs

(f) Model $_{(NCEP/NCAR)}$ with GFDL2.1 outputs

Figure 7 Precipitation reproduced by Model$_{(NCEP/NCAR)}$ and Model$_{(Ensemble)}$ for period 1950-1999

Table i Best sets of potential predictors for each calendar month

Month	Potential variables used in the models 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 ii Performances of the two downscaling models in calibration and validation

Statistic	Calibration (1950-1989)			Validation (1990-2010) /*(1990-1999)**			
	Observations	$Model_{(NCEP/NCAR)}$	$Model_{(Ensemble)}$	Observations		$Model_{(NCEP/NCAR)}$	$Model_{(Ensemble)}$
				1990-2010	1990-1999		
Avg	81.8	82.0	81.1	73.3	***81.8***	81.0	***90.7***
Std	61.7	53.2	42.2	56.9	***64.3***	51.9	***50.9***
C_v	0.75	0.65	0.52	0.78	***0.79***	0.64	***0.56***
NSE		0.74	0.47			0.70	***-0.08***
SANS		0.66	0.30			0.61	***-0.55***
R^2		0.74	0.47			0.72	***0.13***

Avg = average of monthly precipitation in mm $month^{-1}$, Std = standard deviation of monthly precipitation in mm $month^{-1}$, C_v = coefficient of variation, SANS = Seasonally Adjusted Nash Sutcliffe efficiency, NSE = Nash Sutcliffe efficiency, R^2 = coefficient of determination, *Note: bold italicised values in the table refer to period 1990-1999.

Table iii Seasonal performances of the two downscaling models

Model	Statistic	Calibration (1950-1989) Season				Validation (1990-2010)/*(1990-1999)** Season			
		Summer	**Autumn**	**Winter**	**Spring**	**Summer**	**Autumn**	**Winter**	**Spring**
Observed	Avg	40.7	73.7	125.1	87.7	42.9/***(44.3)***	54.1/***(57.0)***	119.4/***(136.1)***	78.3/***(89.8)***
$Model_{(NCEP/NCAR)}$		40.7	73.7	125.1	87.7	49.2	57.8	132.5	85.1
$Model_{(Ensemble)}$		40.7	73.7	125.1	87.7	***(41.1)***	***(83.6)***	***(137.9)***	***(100.3)***
Observed	Std	33.7	58.8	64.5	53.5	41.0/***(46.8)***	43.1/***(46.5)***	61.2/***(66.3)***	48.4/***(55.1)***
$Model_{(NCEP/NCAR)}$		26.0	46.6	54.1	43.9	29.8	33.1	54.1	41.7
$Model_{(Ensemble)}$		15.2	36.8	27.9	33.3	***(17.1)***	***(50.5)***	***(27.7)***	***(44.8)***
Observed	C_v	0.83	0.80	0.52	0.61	0.96/***(1.06)***	0.80/***(0.82)***	0.51/***(0.49)***	0.62/***(0.61)***
$Model_{(NCEP/NCAR)}$		0.64	0.63	0.43	0.50	0.61	0.57	0.41	0.49
$Model_{(Ensemble)}$		0.37	0.50	0.22	0.38	***(0.42)***	***(0.60)***	***(0.20)***	***(0.45)***
$Model_{(NCEP/NCAR)}$	NSE	0.60	0.63	0.70	0.67	0.42	0.75	0.58	0.64
$Model_{(Ensemble)}$		0.20	0.39	0.19	0.39	***(-0.04)***	***(-1.27)***	***(-0.28)***	***(-0.81)***
$Model_{(NCEP/NCAR)}$	R^2	0.60	0.63	0.70	0.67	0.45	0.71	0.63	0.65
$Model_{(Ensemble)}$		0.20	0.39	0.19	0.39	***(0.02)***	***(0.01)***	***(0.01)***	***(0.00)***

Avg = average of monthly precipitation in mm $month^{-1}$, Std = standard deviation of monthly precipitation in mm $month^{-1}$, C_v = coefficient of variation, NSE = Nash Sutcliffe efficiency, R^2 = coefficient of determination, *Note: bold italicised values in brackets in the table refer to period 1990-1999.

Table iv Performances of $Model_{(NCEP/NCAR)}$ and $Model_{(Ensemble)}$ with inputs form different sources (1950-1999)

Statistic	Observed (1950-1999)	***$Model_{(NCEP/NCAR)}$***					**$Model_{(Ensemble)}$**
		NCEP/NCAR	***HadCM3***	***ECHAM5***	***GFDL2.0***	***GFDL2.1***	**Ensemble**
Avg	**81.8**	*83.1*	*117.4*	*139.1*	*102.3*	*104.2*	83.6
Std	**62.1**	*53.8*	*61.8*	*112.1*	*65.7*	*49.8*	44.2
C_v	**0.76**	*0.65*	*0.52*	*0.81*	*0.64*	*0.48*	0.53
Min	**0.0**	*0.0*	*0.0*	*6.9*	*0.0*	*15.9*	4.1
Max	**345.4**	*301.9*	*358.1*	*597.7*	*330.6*	*289.4*	227.7
R^2	N/A	*0.75*	*0.12*	*0.19*	*0.15*	*0.03*	0.36
NSE	N/A	*0.75*	*-0.62*	*-2.54*	*-0.40*	*-0.48*	0.35

Avg = average of monthly precipitation in mm $month^{-1}$, Std = standard deviation of monthly precipitation in mm $month^{-1}$, C_v = coefficient of variation, Min = minimum precipitation in mm $month^{-1}$, Max = maximum precipitation in mm $month^{-1}$, R^2 = coefficient of determination. Bold text refers to the statistics of observed precipitation for the period 1950-1999, Italicised text refers to the performances of $Model_{(NCEP/NCAR)}$ with inputs from different sources, N/A = not applicable

Table v Performances of $Model_{(NCEP/NCAR)}$ and $Model_{(Ensemble)}$ with inputs form different sources (1950-1999), before and after bias-correction

Statistic	**Observed**	***Model*$_{(NCEP/NCAR)}$**					**Model$_{(Ensemble)}$**
		HadCM3	***ECHAM5***	***GFDL2.0***	***GFDL2.1***	***Average***	**Ensemble**
Avg	**81.8**	***81.8****/(117.4)*	***81.8****/(139.1)*	***81.8****/(102.3)*	***81.8****/(104.2)*	***81.8****/(119.6)*	***81.8****/(83.6)*
Std	**62.1**	***62.2****/(61.8)*	***62.2****/(112.1)*	***62.2****/(65.7)*	***62.2****/(49.8)*	***45.9****/(64.8)*	***62.2****/(44.2)*
C_v	**0.76**	***0.76****/(0.52)*	***0.76****/(0.81)*	***0.76****/(0.64)*	***0.76****/(0.48)*	***0.56****/(0.54)*	***0.76****/(0.53)*
Min	**0.0**	***0.0****/(0.0)*	***0.0****/(6.9)*	***0.0****/(0.0)*	***0.0****/(15.9)*	***2.0****/(16.4)*	***0.0****/(4.1)*
Max	**345.4**	***345.4****/(358.1)*	***345.4****/(597.7)*	***345.4****/(330.6)*	***345.4****/(289.4)*	***260.0****/(304.6)*	***345.4****/(227.7)*
R^2	N/A	***0.11****/(0.12)*	***0.12****/(0.19)*	***0.08****/(0.15)*	***0.11****/(0.03)*	***0.19****/(0.24)*	***0.27****/(0.36)*
NSE	N/A	***-0.34****/(-0.62)*	***-0.29****/(-2.54)*	***-0.43****/(-0.40)*	***-0.33****/(-0.48)*	***0.10****/(-0.43)*	***0.04****/(0.35)*

Avg = average of monthly precipitation in mm month^{-1}, Std = standard deviation of monthly precipitation in mm month^{-1}, C_v = coefficient of variation, Min = minimum precipitation in mm month^{-1}, Max = maximum precipitation in mm month^{-1}, R^2 = coefficient of determination. Italicised text in brackets refers to the statistics of precipitation reproduced by downscaling models prior to bias-correction, Bold italicised text refers to the statistics of precipitation reproduced by downscaling models following bias-correction, N/A = not applicable.

Table vi Performances of $Model_{(NCEP/NCAR)}$ with HadCM3 outputs pertaining to the COMMIT GHG emission scenario (2000-2099), before and after bias-correction

Statistic	**Observed (1950-1999)**	**Model$_{(NCEP/NCAR)}$ with HadCM3 COMMIT outputs (2000-2009)**	
		Before bias-correction	***After bias-correction***
Avg	**81.8**	*122.6*	***87.5***
Std	**62.1**	*67.9*	***69.8***
C_v	**0.76**	*0.55*	***0.79***
Min	**0.0**	*1.9*	***0.0***
Max	**345.4**	*387.5*	***424.1***

Avg = average of monthly precipitation in mm month^{-1}, Std = standard deviation of monthly precipitation in mm month^{-1}, C_v = coefficient of variation, Min = minimum precipitation in mm month^{-1}, Max = maximum precipitation in mm month^{-1}, Bold text refers to the statistics of observed precipitation of the period 1950-1999, Italicised text refers to the statistics of precipitation reproduced by downscaling model with the HadCM3 COMMIT outputs for the period 2000-2099 prior to the bias-correction, Italicised bold text refers to the statistics of precipitation reproduced by downscaling model with the HadCM3 COMMIT outputs for the period 2000-2099 after the bias-correction.

Table vii Percentage changes of statistics of bias-corrected precipitation of $Model_{(NCEP/NCAR)}$ $Model_{(Ensemble)}$ and for period 2000-2099, with respect to statistics of observed precipitation of period 1950-1999

Statistic	Downscaling Model	Input source	Summer		Autumn		Winter		Spring	
			A2	B1	A2	B1	A2	B1	A2	B1
Average of precipitation	$Model_{(NCEP/NCAR)}$	HadCM3	**-19% ↓**	**-15% ↓**	+9% ↑	**-4% ↓**	+4% ↑	+7% ↑	**-15% ↓**	**-19% ↓**
		ECHAM5	+14% ↑	+9% ↑	*0% =*	+1% ↑	**-14% ↓**	**-8% ↓**	**-39% ↓**	**-30% ↓**
		GFDL2.0	+5% ↑	**-2% ↓**	**-48% ↓**	**-42% ↓**	*0% =*	+8% ↑	+3% ↑	+15% ↑
		GFDL2.1	**-11% ↓**	**-15% ↓**	**-28% ↓**	**-23% ↓**	+3% ↑	+10% ↑	**-12% ↓**	+2% ↑
	$Model_{(Ensemble)}$	MME	+10% ↑	+6% ↑	+2% ↑	+1% ↑	+2% ↑	+3% ↑	+1% ↑	**-1% ↓**
	Average of $Model_{(NCEP/NCAR)}$		*0% =*	**-3% ↓**	**-13% ↓**	**-15% ↓**	**-3% ↓**	+2% ↑	**-17% ↓**	**-11% ↓**
Standard deviation of precipitation	$Model_{(NCEP/NCAR)}$	HadCM3	+20% ↑	+18% ↑	+15% ↑	+18% ↑	+11% ↑	+13% ↑	+29% ↑	+18% ↑
		ECHAM5	+15% ↑	+18% ↑	*0% =*	+8% ↑	+11% ↑	+9% ↑	+1% ↑	**-4% ↓**
		GFDL2.0	+18% ↑	+6% ↑	**-14% ↓**	**-3% ↓**	+17% ↑	+18% ↑	+27% ↑	+32% ↑
		GFDL2.1	*0% =*	**-17% ↓**	**-2% ↓**	**-6% ↓**	+11% ↑	+7% ↑	+1% ↑	+21% ↑
	$Model_{(Ensemble)}$	MME	**-15% ↓**	**-16% ↓**	**-20% ↓**	**-19% ↓**	**-16% ↓**	**-15% ↓**	**-23% ↓**	**-26% ↓**
	Average of $Model_{(NCEP/NCAR)}$		**-32% ↓**	**-28% ↓**	**-35% ↓**	**-25% ↓**	**-32% ↓**	**-35% ↓**	**-27% ↓**	**-30% ↓**
Monthly maximum precipitation	$Model_{(NCEP/NCAR)}$	HadCM3	+96% ↑	+31% ↑	+23% ↑	+21% ↑	+53% ↑	+43% ↑	+41% ↑	+17% ↑
		ECHAM5	+16% ↑	+43% ↑	+12% ↑	+22% ↑	+40% ↑	+24% ↑	+9% ↑	+11% ↑
		GFDL2.0	+35% ↑	+11% ↑	+10% ↑	+10% ↑	+39% ↑	+41% ↑	+30% ↑	+62% ↑
		GFDL2.1	+30% ↑	**-1% ↓**	+6%↑	**-10% ↓**	+14% ↑	+12% ↑	+17% ↑	+35% ↑
	$Model_{(Ensemble)}$	MME	**-10% ↓**	**-10% ↓**	**-5% ↓**	**-6% ↓**	**-3% ↓**	**-2% ↓**	+11% ↑	+6% ↑
	Average of $Model_{(NCEP/NCAR)}$		**-26% ↓**	**-4% ↓**	**-44% ↓**	**-12% ↓**	**-14% ↓**	**-22% ↓**	**-29% ↓**	**-22% ↓**
Months with zero precipitation	$Model_{(NCEP/NCAR)}$	HadCM3	+16% ↑	+20% ↑	+6% ↑	+10% ↑	+1% ↑	*0% =*	+8% ↑	+5% ↑
		ECHAM5	*0% =*	**-3% ↓**	+1% ↑	+1% ↑	+1% ↑	+2% ↑	+17% ↑	+3% ↑
		GFDL2.0	+2% ↑	+1% ↑	+27% ↑	+28% ↑	*0% =*	*0% =*	+2% ↑	+1% ↑
		GFDL2.1	**-2% ↓**	**-1% ↓**	+15% ↑	+14% ↑	+1% ↑	+1% ↑	+2% ↑	+1% ↑
	$Model_{(Ensemble)}$	MME	**-4% ↓**	**-4% ↓**	*0% =*	*0% =*	*0% =*	*0% =*	*0% =*	*0% =*
	Average of $Model_{(NCEP/NCAR)}$		**-4% ↓**	**-4% ↓**	*0% =*	*0% =*	*0% =*	*0% =*	*0% =*	*0% =*
Months with above average precipitation	$Model_{(NCEP/NCAR)}$	HadCM3	**-8% ↓**	**-2% ↓**	+1% ↑	*0% =*	**-1% ↓**	+1% ↑	**-4% ↓**	**-5% ↓**
		ECHAM5	**-5% ↓**	**-5% ↓**	**-1% ↓**	*0% =*	**-2% ↓**	+2% ↑	**-3% ↓**	**-6% ↓**
		GFDL2.0	**-6% ↓**	**-4% ↓**	**-5% ↓**	**-6% ↓**	+2% ↑	+2% ↑	**-4% ↓**	**-2% ↓**
		GFDL2.1	**-3% ↓**	*0% =*	**-7% ↓**	+1% ↑	*0% =*	+5% ↑	**-3% ↓**	**-2% ↓**
	$Model_{(Ensemble)}$	MME	+3% ↑	**-5% ↓**	+2% ↑	+2% ↑	**-4% ↓**	**-1% ↓**	**-3% ↓**	**-2% ↓**
	Average of $Model_{(NCEP/NCAR)}$		+4% ↑	**-1% ↓**	+3% ↑	+4% ↑	+1% ↑	+2% ↑	*0% =*	+1% ↑

Average of monthly precipitation, standard deviation of monthly precipitation and monthly maximum precipitation are in mm, MME = multi-model ensemble outputs, A2 = high emission scenario, B1 = low emission scenario, ↑ = percentage increase in 2000-2099 with respect to observations of period 1950-1999, ↓ = percentage decrease in 2000-2099 with respect to observations of period 1950-1999 (in bold), = No change in percentage in 2000-2099 with respect to observations of period 1950-1999 (in italics), *Average of* $Model_{(NCEP/NCAR)}$ = average time series computed from the outputs of $Model_{(NCEP/NCAR)}$ when it was run with the outputs of HadCM3, ECHAM5 and GFDL2.0.