Hydrology and Earth System Sciences Discussions

Statistical downscaling of precipitation: state-of-the-art and application of bayesian multi-model approach for uncertainty assessment

M. Z. Hashmi, A. Y. Shamseldin, and B. W. Melville

Department of Civil and Environmental Engineering, School of Engineering, The University of Auckland, New Zealand

Correspondence to: M. Z. Hashmi (mhas074@aucklanduni.ac.nz)

6535

Abstract

Global Circulation Models (GCMs) are a major tool used for future projections of climate change using different emission scenarios. However, for assessing the hydrological impacts of climate change at the watershed and the regional scale, the GCM outputs cannot be used directly due to the mismatch in the spatial resolution between the GCMs and hydrological models. In order to use the output of a GCM for conducting hydrological impact studies, downscaling is used. However, the downscaling results may contain considerable uncertainty which needs to be quantified before making the results available. Among the variables usually downscaled, precipitation downscaling is quite challenging and is more prone to uncertainty issues than other climatological variables. This paper addresses the uncertainty analysis associated with statistical downscaling of a watershed precipitation (Clutha River above Balclutha, New Zealand) using results from three well reputed downscaling methods and Bayesian weighted multi-model ensemble approach. The downscaling methods used for this study belong to the following downscaling categories; (1) Multiple linear regression; (2) Multiple non-linear regression; and (3) Stochastic weather generator. The results obtained in this study have shown that this ensemble strategy is very efficient in combining the results from multiple downscaling methods on the basis of their performance and quantifying the uncertainty contained in this ensemble output. This will encourage any future attempts on quantifying downscaling uncertainties using the multi-model ensemble framework.

1 Introduction

At present, there is a wide variety of Global Circulation Models (GCMs) which can be used for future projections of climate change using different emission scenarios (SRES, 2000). However, for assessing the hydrological impacts of climate change at the watershed and the regional scale, the GCM outputs cannot be used directly due

6536

to the mismatch in the spatial resolution between the GCMs and hydrological models. In principle, hydrological models run on a very fine spatial resolution (in the order of tens of kilometers, or even less) while the GCMs have spatial resolutions in the range of hundreds of kilometers. In order to use the output of a GCM for conducting hydrological impact studies, downscaling is used. It is simply a process of converting the coarse spatial resolution of the GCM output into a fine resolution which can involve generating point/station data of a specific area by using the GCM climatic output variables.

In broad terms, downscaling techniques can be classified into dynamical and statistical. Further details about the underlying principles and review of their applications can be found in Hewitson and Crane (1996), Xu (1999), Wilby et al. (2004) and Fowler et al. (2007). This paper focuses on studying the uncertainty issues associated with the use of statistical downscaling. There are a variety of techniques available for statistical downscaling. Each of these techniques has certain strengths and weaknesses depending on the working principles involved in the operation of the technique (Xu, 1999). Although the statistical downscaling is very popular and extensively used in many studies (Christensen et al., 2007), it usually performs well only for the conditions and regions where it was originally developed. It is also affected by four major sources of uncertainties which can significantly influence the reliability of its results and hence future projections of climate change. These four major sources of uncertainties are; (1) parent GCM; (2) climate change emission scenarios; (3) observed data; and (4) method used for downscaling.

In the context of climate change studies, these four major uncertainty sources form what is known as the *cascade of uncertainty* (Mitchell and Hulme, 1999). However, in climate change studies the uncertainty analysis has focused on studying the uncertainties associated with the first source (parent GCM) and the second source (emission scenarios). There is very limited research regarding uncertainty analysis associated with statistical downscaling, this paper addresses this research gap.

The Bayesian modeling framework introduced by Tebaldi et al. (2005), henceforth, referred to as TME, is used herein for developing a weighted multi-model ensemble

(WMME) of different statistical downscaling models to quantify the uncertainty in downscaling of monthly precipitation. Three statistical downscaling models are used to test the multi-model ensemble on the Clutha catchment in New Zealand. The first selected downscaling model is a weather generator model known as the Long Ashton Research Station Weather Generator (LARS-WG) developed by Semenov and Barrow (1997). The second model is the Statistical DownScaling Model (SDSM) developed by Wilby et al. (2002). The third model is an artificial intelligence data driven model developed by the authors using the Gene Expression Programming (GEP) to create symbolic downscaling functions. The choice of these downscaling models is consistent with the choices made in many of the recent studies (e.g. Hessami et al., 2008; Khan et al., 2006; Semenov 2008; Liu et al., 2008).

Bayesian techniques are powerful in producing improved forecasts by ensembling information from different sources and providing a measure of forecast uncertainty. They have many advantages over other probabilistic uncertainty analysis methods (Qian et al., 2003). The advantages include their ability to deal efficiently with uncertainty, incorporate prior knowledge as well as information from multiple sources and update probability distributions on the basis of new incoming information. In climate change studies and weather forecasting, examples of the use of Bayesian techniques in MME can be found in Coelho et al. (2006), Tebaldi et al. (2005, 2004), Giorgi and Mearns (2002) and Min and Hense (2006, 2007).

In this paper a brief overview of the different statistical downscaling techniques is given first and a review on uncertainty assessment in downscaling is then presented. Secondly, a discussion about how to deal with downscaling uncertainty using multi-model ensembles is provided. Thirdly, the multi-model ensemble approach developed in this paper and its application using three selected downscaling methods are presented. Finally, the conclusions of the presented study are given.

2 Study watershed and data sources

The Clutha River is the biggest river in the South Island of New Zealand. Its catchment above Balclutha (shown in Fig. 1) has been selected as a case study. It is the second longest river in length (340 km) and the largest river in terms of volume and catchment area in New Zealand (McKerchar and Henderson, 2003). Its long term annual mean flow is approximately 614 m^3/s. The catchment area is around 20 515 km^2 up to Balclutha and mean annual precipitation is around 1448 mm (National Institute of Water and Atmospheric Research (NIWA), New Zealand, web model: http://wrenz.niwa.co.nz/webmodel). The Clutha starts from the high Southern Alps glaciers, and the river being very hazardous due to its flooding potential. The climatic data used in this study were obtained from NIWA. Daily average precipitation data of 23 stations within Clutha at Balclutha watershed having record lengths of thirty years or more were acquired for the period 1961–2000. The corresponding length of large scale predictors of Hadley Center's GCM "HadCM3" for the baseline period as well as the A2 scenario run for 2070–2099 (also called 2080s) and re-analysis data predictors of the National Center of Environmental Prediction (NCEP) on HadCM3 computational grid were obtained from the Canadian Climate Impacts Scenarios (CCIS) website (http://www.cics.uvic.ca/scenarios/index.cgi/) at daily time steps. The A2 scenario is considered as the worst case scenario among the IPCC's four basic storylines (SRES, 2000). In climate change studies, the period of 1961–2000 is used to represent the current climate (e.g. Wilby et al., 2002).

In order to get a downscaled time series using a weather generator, the mean daily precipitation output of HadCM3 covering the whole globe, was obtained from Program for Climate Model Diagnosis and Inter-comparison (PCMDI) website (https://esg.llnl.gov:8443/) for the present period of 1961–1989 (called the 20th century run) and the A2 scenario run for 2080s.

6539

3 Review of statistical downscaling approaches

Statistical downscaling is based on the use of statistical tools and rules to develop local scale hydro-meteorological data using the GCM outputs. Statistical downscaling approaches can be classified into three broad categories, namely; (1) weather typing; (2) weather generators; and (3) regression-based downscaling. Although these broad downscaling categories may appear to differ radically in their operation, they generally embody three basic assumptions (cf. Hewitson and Crane, 1998, 2006);

1. selected predictor variables are relevant to the study and the host GCM is able to simulate them realistically.

2. the empirical relationships/rules developed under the present climate conditions are also valid for future climate change conditions.

3. selected predictor variables are able to capture the climate change signal.

The "weather typing" category involves grouping of the local meteorological variables on the basis of different atmospheric circulation patterns/types. The GCM projected change in atmospheric circulation types is used to project change in those local variables. Further discussion on the operation of this approach and its applications to climate change studies can be found in Ladd and Driscoll (1980) and Conway and Jones (1998). A modified version of weather typing includes pattern scaling developed by Ruosteenoja et al. (2007). Being subjective in nature, this approach is less popular than other downscaling approaches (Wilby et al., 2004) and hence was not used in this study in developing the multi-model downscaling ensemble.

Weather generators which are traditionally used for filling missing data and interpolating or extrapolating the data to indefinite length by using the statistical properties of the available observed data have become a very popular tool for downscaling (Dibike and Coulibaly, 2005). A comprehensive review of the theory behind weather generators and their evolution over time as well as their use for different purposes can be

6540

found in Wilks and Wilby (1999). In this study, a weather generator model known as LARS-WG, developed by Semenov et al. (1998), is used as a member of the ensemble downscaling model.

The regression based downscaling methods involve developing empirical relationships between large scale GCM data or observed data as "Predictor" variables and local or small scale climate variables as "Predictand" variables using traditional linear and non-linear regression methods (cf. Heyen et al., 1996; Wilby et al., 2002). Examples of traditional downscaling regression based methods include linear regression, canonical correlation analysis (CCA) and principle component analysis (PCA) (Dibike and Coulibaly, 2005). However, in recent years non-linear regression models based on soft computing data-driven modeling techniques have also been used. In the context of downscaling, recent studies have shown that artificial neural network (ANN) models are good multiple non-linear regression models (e.g. Mpelasoka et al., 2001) and can be considered as a global correlation detector, but suffer from conventional problems of a complex model, such as being trapped in local optimum during their calibration (Tripathi et al., 2006). Coulibaly (2004) and Liu et al. (2008) suggested the use of another sophisticated soft computing technique known as genetic programming to perform symbolic regression. The advantage of the genetic programming based tool for developing downscaling functions over ANN is that it provides parsimonious and transparent modeling solutions which can be expected to have less uncertainty in them. In this study, a variant of the genetic algorithms known as Gene Expression Programming (GEP) is used in conjunction with symbolic regression to develop a downscaling model used in the multi-model downscaling ensemble.

Different inter-comparison studies show advantages of certain statistical downscaling techniques over the others (e.g. Wilby et al., 1998) under certain conditions. However, it is important to note that none of the currently available statistical downscaling techniques was able to reproduce the precipitation with all its characteristics very well. Thus, there is no universal single statistical downscaling technique that works very well under all circumstances. According to the "Guidelines for Use of Climate Sce-

narios Developed from Statistical Downscaling Methods" (Wilby et al., 2004), the user should carefully select the downscaling method according to the nature of problem and predictands involved. It also recommends the use of more than one downscaling method to minimize the uncertainties. This paper endorses this recommendation by using three downscaling techniques and integrating their results using the Bayesian modeling framework to quantify the uncertainty in downscaling of precipitation.

3.1 Uncertainty assessment in statistical downscaling

Due to many noted reasons discussed in the introduction of this paper, the results obtained from downscaling models may have a considerable amount of uncertainty and a single deterministic point estimate will be wrong. In this case, a probabilistic approach for proper uncertainty assessment of the downscaling results should be adopted to obtain a range of projection of climatic variable values rather than a single value. To our present knowledge, there are only limited studies which deal with the uncertainty analysis of downscaling results and the first attempt to quantify the uncertainty of downscaling is made by Khan et al. (2006). However, more recently Wilby and Harris (2006) developed a framework for analysis of uncertainty arising from all sources and used a weighted multi-model ensemble strategy for minimizing uncertainty. Equal weights were assigned to the different statistical downscaling methods by employing simple averaging. Although a simple average approach has shown definite advantages over a single model approach in terms of robust uncertainty assessment (Hagedorn et al., 2005), it may not be acceptable in the case where the downscaling methods show diversity in the level of performance. In the multi-model ensemble of different statistical downscaling models developed in this study, different weights are assigned to the different models. Hence, the variation in model performance and uncertainties associated with the individual downscaling model used to produce the ensemble are taken into account.

4 Bayesian way of uncertainty assessment

Due to advancement in computational resources in more recent years, the Bayesian statistical techniques have become very popular for uncertainty analysis in different scientific research fields. They have many advantages over other probabilistic uncertainty analysis methods, which are discussed earlier.

According to the Bayes' theorem (Bayes, 1763), as soon as the observed data "y" related to a parameter of interest "x" is available, the prior distribution of the variable "x", $P(x)$ is updated to obtain its conditional posterior distribution, $P(x|y)$. This can be presented in the form of Bayes's equation as;

$$P(x|y) \approx P(y|x) \cdot P(x) . \tag{1}$$

In Eq. (1), $P(y|x)$ is the likelihood of data "y" given "x". In most of the cases, it is not possible to solve the Bayes's equation analytically to obtain the posterior distribution. Instead, the equation is solved numerically using the Monte Carlo simulation method which involves the generation of a large number of random samples. There are many variants of the Monte Carlo simulation method which can be used for generating random samples (Qian et al., 2003). The Markov Chain Monte Carlo (MCMC) simulation method is very popular and hence has been used in this study to derive the posterior distribution. Further information about the applications of the MCMC method can be found in Wang (2008), Engeland and Gottschalk (2002) and Qian et al. (2003).

4.1 Bayesian WMME approach and uncertainty assessment

Figure 2 shows a schematic diagram outlining the steps involved in the Bayesian WMME approach (based on TME) for uncertainty analysis of multiple statistical downscaling methods used in this study. In the above diagram, X_0 is the observed Clutha precipitation, monthly averaged for 1961–1990 period, X_i is the monthly average of the simulated precipitation of the ith downscaling model for the 30-year period of 1961–1990, and Y_i is the monthly average of the simulated precipitation of the ith downscal-

ing model for a future 30-year period of 2070–2099. The subsequent sections provide a concise summary of the steps involved in the formulation of a Bayesian framework for WMME. Full details are available in Tebaldi et al. (2005).

4.1.1 Models for data (Likelihood functions)

The operation of the Bayesian multi-model approach requires the specification of probability distributions to describe the likelihood functions for the observed data as well as for the results of three downscaling methods. In this paper, similar to other studies (e.g. Tebaldi et al., 2005), it is assumed that downscaling models and observed climate data are statistically independent and that three data sets X_i, Y_i and X_0 come from Normal (N) distributions and their means are the true value of the data for the month represented by the data. Accordingly, the likelihood functions for the three data sets are given as (cf. Lopez et al., 2006);

$$X_0 \sim N(\mu, \lambda_0^{-1}) , \tag{2}$$

$$X_i \sim N(\mu, \lambda_i^{-1}) , \tag{3}$$

$$Y_i \sim N(\nu + \beta(X_i - \mu), (\theta\lambda_i)^{-1}) , \tag{4}$$

λ_0 is the observed variability of mean monthly Clutha precipitation as given in Table 4 while μ and ν (assumed to be random variables) are the true values of the present and future Clutha precipitation, The parameter $\lambda_i^{-1} = \sigma_i^2$ is the reciprocal of distributional variance; here it is assumed to be a random variable and a measure of the precision of the ith downscaling model. In the above Eq. (4), the random variable θ is an allowance to have different variances for present and future downscaled output by a model, while random variable β makes the model robust by introducing a correlation between present and future downscaling results.

4.1.2 Specification of parameter priors

The data models or likelihood functions presented above contain the parameters μ, ν, λ_i, θ and β which are treated as random variables and in Bayesian terms are required to have their prior distribution assigned. To keep this Bayesian analysis an objective one, these parameters were assigned uninformative prior densities (assuming no prior knowledge about them). Random variable, λ_i (i=1,...,3) and θ have Gamma (Ga) distribution of the form;

$$\lambda_i \sim \mathrm{Ga}(a,b) \,, \tag{5}$$

$$\theta \sim \mathrm{Ga}(c,d) \,, \tag{6}$$

where $a = b = c = d = 0.001$ with a unity mean and a large variance over the positive real line, thus keeping objectivity in their specification. According to Tebaldi et al. (2005), the prior distribution selected for λ_i is a standard choice for the precision parameters of Gaussian distributions. The prior densities for μ, ν and β are also uninformative being uniform over the real line which is a further attempt to present an objective analysis.

4.1.3 Bayesian analysis

Multiplying Eqs. (2) to (6) by applying Bayes's theorem gives the required joint posterior distribution, which is then solved up to a normalizing constant (sums up to 1) to simplify the computations, giving (cf. Lopez et al., 2006);

$$\prod_{i=1}^{3} \lambda_i^{a-1} e^{-b\lambda_i} \times \lambda_i \theta^{1/2} \exp\left\{ -\frac{\lambda_i}{2} \left[(X_i - \mu)^2 + \theta(Y_i - \nu)^2 \right] \right\}$$
$$\times \theta^{c-1} e^{-d\theta} \times \exp\left\{ -\frac{\lambda_0}{2}(X_0 - \mu)^2 \right\} \,. \tag{7}$$

The full conditional posterior distributions for each parameter as a function of all other parameters are then obtained (Tebaldi et al., 2005), e.g. posterior distribution of precision parameter λ_i (used for model weighting) comes out to be a Gamma function having mean (cf. Lopez et al., 2006);

$$\lambda_i \approx \frac{a + 1}{b + \frac{(X_i - \mu)^2}{2} + \frac{\theta[(Y_i - \nu) - \beta(X_i - \mu)]^2}{2}} \,. \tag{8}$$

The two terms $(X_i - \mu)$ and $(Y_i - \nu)$ in the above expression define the criteria of "bias" and "convergence", respectively as the former term is model error as compared to true value and quantifies the ith model's ability to reproduce the current climate while the latter term is a measure of the distance of the future prediction of ith model from the ensemble mean. Accordingly, a downscaling model showing better skill in reproducing current climate and its future output is more close to the ensemble mean, and will get more weight then the one with inferior results.

4.1.4 MCMC simulation

The joint posterior distribution obtained in the previous step by applying the Bayesian analysis has a complex form. It cannot be solved in a closed form solution. Hence it is not possible to obtain directly the marginal posterior distributions of individual parameters. In such a situation, the MCMC simulation is performed. A large number of random samples are drawn from the posterior distributions for all parameters. These sets of random samples are then used for further statistical analysis.

4.1.5 Uncertainty assessment

The posterior distributions of parameters such as μ and ν were obtained using MCMC simulation in the previous step. These can be utilized for assessment of uncertainties in the estimation of individual parameters or downscaled precipitation change projection in terms of statistical parameters. Therefore, the Bayesian WMME was made to estimate a signal of percent precipitation change for the Clutha watershed by computing the posterior distribution of $\%\Delta P = 100 * (\nu - \mu)/\mu$.

Box plots are usually used to describe and compare the distribution of different groups of data. These plots for posterior distributions of monthly precipitation change ($\%\Delta P$) were produced to present the distribution shape graphically. The difference between the 1st (25% quantile) and 3rd (75% quantile) quartile of a probability distribution is called the "inter-quartile range (IQR)" or mid-spread. It gives an idea about the compactness of the distribution and is represented by a box in the plot. The noted advantages it offers make it a popular application as a robust measure of variability and hence uncertainty (Lee, 1995). Therefore, it has been adopted herein as the uncertainty assessment tool of the downscaled precipitation change.

In this study, a monthly multi-model downscaling ensemble is developed based on TME, an improved version of the model developed by Giorgi and Mearns (2002). In this multi-model ensemble, the three downscaling models are assigned weights based on their precision parameter calculated using a criteria of "bias" and "convergence" (explained in Sect. 4.1.3). TME has been implemented in a computer program developed using the statistics package R which can be downloaded from; http://www.r-project.org to develop the multi-model downscaling ensemble.

5 Applications of the three satistical downscaling methods and MME model to study watershed

5.1 SDSM Model

SDSM is a multiple regression based tool for generating future scenarios to assess the impact of climate change. It has the ability to capture the inter-annual variability better than other statistical downscaling approaches, e.g. weather generators, weather typing (Wilby et al., 2002). SDSM is a combination of a stochastic weather generator approach and a transfer function model (Wilby et al., 2002) needing two types of daily data. The 1st type corresponds to local predictands of interest (e.g. temperature, precipitation etc) and the 2nd type corresponds to the data of large scale predictors (NCEP and GCM) of a grid box closest to the study area. Correlation and partial correlation analysis is performed in SDSM between the predictand of interest and predictors to select a set of predictors most relevant for the site in question (Wilby et al., 2002).

Initially, the precipitation data of a selected station from the study watershed (Ophir2) are used in SDSM. The correlation analysis performed within SDSM between the precipitation of this site and the NCEP re-analysis predictors revealed very poor results. Consequently, offline statistical analysis was undertaken to improve the results to an acceptable limit. A cross-correlation analysis between the daily precipitation of Ophir2 and the NCEP predictors was an attempt in this regard. An optimal lag or time shift, required to improve the correlation between each predictor-predictand pair, was identified and the correlation results after this analysis are shown in Fig. 3. Examination of the figure shows that the correlation coefficient values obtained are well below the acceptable limit as indicated in previous studies (e.g. Hessami et al., 2008). Further cross-correlation analysis using the data of other stations revealed conclusions similar to those obtained using the Ophir2 station. Accordingly, these predictors cannot be used directly for reliable downscaling at the station scale.

As the outcome of the station scale analysis was not favorable, an areal approach was adopted. The arithmetic average of the daily data from 23 stations within the

Clutha watershed, with a record length of at least 30 years, was taken. This areal average precipitation time series is referred to as "Clutha precipitation". Cross correlation analysis was again performed between the Clutha precipitation and the NCEP predictors. Table 2 shows the results of this cross correlation analysis in terms of predictand-predictor lag (in days) required to get the maximum correlation between them. After arranging each NCEP predictor against Clutha precipitation on the basis of Table 2, the predictor-predictand correlation improved significantly. This has been highlighted in Fig. 4. Examination of Fig. 4 shows that there are a number of NECP predictors which have correlation coefficient values in the range of 13% to 25% for the Clutha precipitation. This range is considered to be acceptable when dealing with precipitation downscaling (cf. Wilby et al., 2002).

The lagged NCEP predictors and Clutha precipitation obtained in the previous step were then used in SDSM for the final analysis. Screening of the most relevant predictors' set was performed in SDSM on the basis of correlation and partial correlation analysis among the predictand and the individual predictors and a set of 10 predictors were chosen. This is shown in bold text in Table 1. The predictor selection process is consistent with similar studies (Dibike and Coulibaly, 2005). The 10 chosen predictors were used for calibration of the downscaling model.

The Clutha precipitation and NCEP predictors for the period of 1961–2000 were split into two parts. The first part for 1961–1990 was used for model calibration while the remaining data of 1991–2000 were used for model validation (as an independent set of data). SDSM was calibrated for each month of the year using the same set of 10 selected NCEP predictors for the calibration period. Different values of SDSM set-up parameters such as "Variance Inflation" and "Bias Correction" were tested. The best combination of these parameters gave a model with maximum coefficient of determination (R^2) and identical standard deviation in the comparison of observed and simulated data. This model tuning/calibration strategy is in line with the one explained in Dibike and Coulibaly (2005). Model validation was performed by testing it for an independent data set of 1991–2000 which also revealed satisfactory results, as given in Sect. 6.1.

6549

5.2 LARS-WG model

LARS-WG is a weather generation tool. It can be used to synthesize, fill in missing values or generate for an ungauged site (using observed data properties of a neighboring gauged site), the daily climatic parameters such as precipitation, temperature and solar radiation. It takes as an input the long term daily information of the climatic parameter of interest for a site. It can also generate the scenarios of changed climate for a site by perturbing the parameters derived from the observed data to generate synthetic data.

Before using the LARS-WG for downscaling, its performance using the Clutha precipitation for the period 1961–2000 was analyzed. This precipitation time series was used as an input to LARS-WG to generate a synthetic daily precipitation time series with a record length of 500 years. The statistical properties of the synthetic time series were compared to those of the observed data in order to gauge the ability of LARS-WG in reproducing the observed precipitation statistics. The statistical properties include mean monthly precipitation and monthly standard deviations of precipitation which are very similar to those used in previous studies (e.g. Dibike and Coulibaly, 2005).

The downscaling of precipitation using LARS-WG and the output from a GCM is performed with the help of relative monthly change factors for average precipitation and average length of dry/wet spell. In the present study, the daily precipitation time series of HadCM3, representing current climate forcing, for 1961–1989 and a future time series of 2080s based on SRES A2 scenario run are used in calculating the relative change factors. The data extracted from HadCM3 output for the two periods (present and 2080s) were used by LARS-WG to calculate month-wise mean daily precipitation and average length of wet and dry spells and hence the change between the two climate regimes, projected by the GCM. Table 3 gives the month-wise relative change factors for mean daily precipitation, average wet spell days and average dry spell days, as calculated from the HadCM3 daily precipitation data and provided to the LARS-WG. Once the change factors were calculated they were subsequently used in LARS-WG to perturb parameters in order to generate a 30-year daily time series,

6550

representing 2080s.

5.3 GEP Model

As discussed in Sect. 3, studies have shown that soft computing non-linear regression methods such as ANNs are efficient downscaling models that can capture the complex relationship between the highly variable large scale predictors and the required predictands. However, a recent study by Coulibaly (2004) has noted that non-linear models developed using symbolic regression and genetic programming are powerful downscaling models for temperature. The symbolic regression is a form of nonparametric regression in which the function relating predictors and predictand variables is not specified a priori, but the function is constrained to contain a number of mathematical or logical expressions to be chosen from a large pre-selected set of mathematical expressions (symbols) and predictor variables. In the symbolic regression, the genetic programming which mimics the Darwin evolution theory is used to obtain the optimum set of symbols and predictor variables. In this study, the GeneXproTools 4.0 software, a powerful soft computing package is utilized for the first time to perform symbolic regression operations using gene expression programming (GEP) to develop a precipitation downscaling model. The GEP is a variant of Genetic programming and further details can be found in Ferreira (2001, 2006).

The software package was provided with the daily input of the same lagged large scale predictors as were used previously in SDSM and Clutha precipitation as local predictand in the form of two data sets. The first set which covers the period of 1961–1990 is used for training/calibration and the second set which covers the period of 1991–2000 for testing/validation. The results obtained using GEP show a marked improvement over SDSM in developing a predictor-predictand relationship in terms of the obtained coefficient of determination (R^2) (Observed and simulated daily time series R^2 is "0.48"). In terms of monthly standard deviation, the model developed by GEP underestimates for all months of the year.

5.4 MME model

The values of mean monthly precipitation for present (X_i) and GCM projected (Y_i) were obtained from downscaling models' output for the selected watershed. The natural variability of observed monthly data, λ_0, given in Table 4 were also provided as input data. MCMC simulation is performed with a Gibbs sampler to draw random samples from the posterior distributions of parameters (i.e. "μ", "v" and "λ_is") and was made to run through 7 50 000 iterations, saving every 50th sample to avoid correlation between the successive values. A number of initial samples were discarded (as burn-in period) to include only the final stable output into the analysis. The final set comprised of 5000 samples of each parameter of interest, the selection of the total number of iterations, "save every" and burn-in period being consistent with other studies (Tebaldi et al., 2005). The means of the posterior distributions of λ_is, computed through WMME, for three downscaling models were computed and then the relative weight (ω_i) to be assigned to each model was computed. Using this methodology, the relative weights calculated for each model are given in Table 5.

6 Results and discussion

6.1 SDSM results

Figure 5 shows comparisons of the observed and SDSM estimated month-wise mean daily precipitation and its standard deviation, respectively. Examination of Fig. 5 shows that the calibrated models have reproduced the seasonal values quite well. Although the model has slightly underestimated the mean monthly precipitation for the months of April, June, July and September (<1 mm) and almost equally overestimated for the months of January, February, March, May, August, October, November and December. Hence the observed and the estimated annual precipitation are equal.

Further examination of Fig. 5, in terms of monthly standard deviation results, reveals

that only for a few months (June, July, September and December), SDSM simulated standard deviation is below that of the observed data. For all other months of the year, the simulated and observed standard deviations are in a good agreement.

Downscaling models are often regarded as less skillful to model the standard deviation (or variance) of the observed precipitation with great accuracy (Wilby et al., 2004). However, the SDSM model for this study has reproduced the observed standard deviation exceptionally well.

After accomplishing a satisfying calibration, the multiple regression model was validated using the data of 1991–2000 (as an independent set of data outside the period for which the model is calibrated) and the results obtained are shown in Fig. 6. Examination of this figure reveals that the model is successfully validated.

Next the corresponding large scale predictors of HadCM3 were used to downscale the future climate data of the selected watershed. The calibrated model was run for 10 corresponding GCM predictors for the current climate (1961–1990) and future (2080s). The results obtained on the basis of these runs are shown in Fig. 7.

The future picture sketched by HadCM3 (Fig. 7) shows a mixed trend as for months of March, September and December, it projects a decrease in mean daily precipitation as compared to an increase for the rest of the year. Overall the precipitation in this watershed is projected by HadCM3 to be increased by 1 mm or more by the end of this century (2080s).

6.2   LARS-WG results

Visual analysis of LARS-WG results in terms of monthly mean precipitation and standard deviation are presented in Fig. 8 which shows that means of daily precipitation for each month are very well modeled by LARS-WG except for the summer months (January, February and December) where they are overestimated. In terms of standard deviation, LARS-WG has shown an average performance as it mostly underestimates the observed data but overall it shows much better results for the months of February to May and July than for the rest of the year.

6553

The monthly change factors used by LARS-WG to generate a 30 year data record representing the 2080s time slice are given in Table 3. A comparison between the LARS-WG generated future data of 2080s and the observed data of 1961–2000 period in terms of mean daily precipitation for each month is presented in Fig. 9 to reveal the change in the precipitation regime of the Clutha watershed between the two analyzed time periods. For most of the months, GCM is projecting an increase in mean daily precipitation except for the months of January, March and December. This increase varies between 0.2–1.5 mm with May being the month with highest increase.

6.3   GEP results

Graphical presentation of the GEP model training results is provided in Fig. 10. In the comparison of observed and simulated data, the GEP model underestimates the mean daily precipitation for the months of January, March, April and December while the opposite is true for the months of May to November. In terms of standard deviation of precipitation, the GEP model's results (Fig. 10) are less satisfying as the model consistently underestimates throughout the year. Results obtained using the GCM predictors to run the GEP model are presented in Fig. 11. These results show a considerable increase in precipitation throughout the year for the period of 2080s. The projected increase ranges from 0.3–3.5 mm, which is larger as compared to the SDSM downscaling results.

6.4   MME results

Skill comparison of downscaling models on the basis of "bias" and "convergence" criteria is one of the main objectives of this study. Table 5 provides this comparison numerically. The performance of all models is variable for different months of the year so they have been given weightage accordingly. For example, in the ensemble of three downscaling models, SDSM has a 30% or higher weightage for most of the year, an exception being the month of June. The weightage assigned to LARS-WG is always

6554

more than 10% but varies between 10–60% throughout the year. The GEP model is the weakest of the three for most months of the year on the basis of the weightage assigned to it. But the months of June and July are both exceptional cases, where it is the best performing model as it has been given the highest weightage. Figure 12a and b is pictorial representation comparing the skill of three downscaling models on the basis of their precision (λ_i) value. This figure actually contains the box plots of the monthly posterior distributions of the model specific precision parameter (λ_i) for all three downscaling models. The vertical position of a box (representing any of the three models) in comparison to the other two boxes is indicative of the precision of that model in comparison to the others. A model having large λ_i indicates it is more skillful in matching the true climate response than one with smaller λ_i. Also, a model with higher box-plot position will get more weight than one with lower position. Using this comparison measure, it can be seen that the vertical position of all three boxes is highly variable for different months of the year. On most occasions, the models differ considerably in their skill. An exception is the month of May where the three models show quite similar performance. In most instances, the model with highest weightage was either SDSM or LARS-WG. In this way, the MME has taken into consideration the strength and the weakness of each model and produced a downscaled output which would be more reliable than either of the individual models.

Figure 13 shows the box plots of monthly posterior distributions of the percent precipitation change (%ΔP) providing an overall picture about the nature of the distributions for each month and the level of uncertainty in the monthly precipitation change projections after downscaling. The white thick line in each box is the median value. Examining Fig. 13 in terms of IQR as a measure of uncertainty, a variable trend can be seen of monthly precipitation change distributions. This shows different degrees of uncertainty for different months which is a good representation of magnitudes of precipitation change among 12 months of the year. The months of May, August and December have very compact (tight) distributions as compared to the boxes for other months. For these three months, the small IQR indicates very small uncertainty in their

6555

positive precipitation change signal. On the other hand, February and April are exceptional cases where the quantiles are wide apart giving large IQR. The large uncertainty thus resulting can be attributed to the GCM and the downscaling models' bias for these months. For the months of January, March, April, July and September to November a lucid change signal is obtained as boxes have some part on both sides of the 0-line.

The results shown and discussed above further strengthen the idea of employing Bayesian WMME for multiple downscaling models instead of relying on the output of an individual model.

7 Conclusions

The objective of the study is to present the application of a Bayesian framework to develop WMME of multiple downscaling models and efficiently quantify uncertainty in the downscaling of monthly precipitation of a watershed. The WMME is based on the Bayesian framework developed by Tebaldi et al. (2005). Three well reputed downscaling models namely SDSM, LARS-WG and GEP were used. Downscaling experiments were carried out for precipitation data of the Clutha watershed in South Island, New Zealand. The two time slices analyzed for precipitation change were the baseline period of 1961–1990 and the future period of 2080s. The large scale data of HadCM3 model has been used for baseline period and future period of 2080s for SRES A2 scenario. By the application of Bayesian framework, the monthly outputs from three downscaling models and observed data of the watershed were efficiently combined. On the basis of the strengths of individual models in comparison to fellow downscaling models in terms of "bias" and "convergence" criteria, each model was assigned a relative weight to be used in the WMME. This new application for multiple downscaling models has proven to be very efficient in terms of quantifying uncertainty contained in downscaled results and gives new directions for more work on multi-model downscaling ensembles. The work presented in this paper strongly supports the use of multi-model ensemble downscaling for providing the required data for hydrological impact

6556

assessment.

References

Bayes, T.: An essay towards solving a problem in the doctrine of chance, Philos. T. R. Soc. Lond., 53, 370–418, 1763.

Christensen, J. H., Carter, T. R., Rummukainen, M., and Amanatidis, G.: Evaluating the performance and utility of regional climate models: The prudence project, Climatic Change, 81, 1–6, 2007.

Coelho, C. A. S., Stephenson, D. B., Doblas-Reyes, F. J., Balmaseda, M., Guetter, A., and van Oldenborgh, G. J.: A Bayesian approach for multi-model downscaling: Seasonal forecasting of regional rainfall and river flows in South America, Meteorol. Appl., 13(1), 73–82, 2006.

Conway, D. and Jones, P. D.: The use of weather types and air flow indices for GCM downscaling, J. Hydrol., 212–213(1–4), 348–361, 1998.

Coulibaly, P.: Downscaling daily extreme temperatures with genetic programming, Geophys. Res. Lett., 31, L16203, doi:10.1029/2004GL020075, 2004.

Dibike, Y. B. and Coulibaly, P.: Hydrologic impact of climate change in the Saguenay watershed, comparison of downscaling methods and hydrologic models, J. Hydrol., 307(1–4), 145–163, 2005.

Doblas-Reyes, F. J., Deque, M., and Piedelievre, J. P.: Multi-model spread and probabilistic seasonal forecasts in PROVOST, Q. J. Roy. Meteor. Soc., 126(567), 2069–2087, 2000.

Doblas-Reyes, F. J., Hagedorn, R., and Palmer, T. N.: The rationale behind the success of multi-model ensembles in seasonal forecasting – II. Calibration and combination, Tellus A, 57(3), 234–252, 2005.

Engeland, K. and Gottschalk, L.: Bayesian estimation of parameters in a regional hydrological model, Hydrol. Earth Syst. Sci., 6, 883–898, 2002, http://www.hydrol-earth-syst-sci.net/6/883/2002/.

6557

Fowler, H. J., Blenkinsop, S., and Tebaldi, C.: Linking climate change modelling to impacts studies, Recent advances in downscaling techniques for hydrological modeling, Int. J. Climatol., 27, 1547–1578, 2007.

Giorgi, F. and Mearns, L. O.: Calculation of average, uncertainty range, and reliability of regional climate changes from AOGCM simulations via the "Reliability ensemble averaging" (REA) method, J. Climate, 15, 1141–1158, 2002.

Hagedorn, R., Doblas-Reyes, F. J., and Palmer, T. N.: The rationale behind the success of multi-model ensembles in seasonal forecasting – I. Basic concept, Tellus A, 57, 219–233, 2005.

Hessami, M., Gachon, P., Ouarda, T. B. M. J., and St-Hilaire, A.: Automated regression-based statistical downscaling tool, Environ. Modell. Softw., 23, 813–834, 2008.

Hewitson, B. C. and Crane, R. G.: Climate downscaling: Techniques and application, Clim. Res., 7, 85–95, 1996.

Hewitson, B. C. and Crane, R. G.: Consensus between GCM climate change projections w ith empirical downscaling: Precipitation downscaling over South Africa, Int. J. Climatol., 26, 1315–1337, 2006.

Heyen, H., Zorita, E., and Von Storch, H.: Statistical downscaling of monthly mean North Atlantic air-pressure to sea level anomalies in the Baltic sea, Tellus A, 48, 312–323, 1996.

Khan, M. S., Coulibaly, P., and Dibike, Y.: Uncertainty analysis of statistical downscaling methods, J. Hydrol., 319, 357–382, 2006.

Ladd, J. W. and Driscoll, D. M.: A comparison of objective and subjective means of weather typing: An example from West Texas, J. Appl. Meteorol., 19, 691–704, 1980.

X. Liu, Coulibaly, P., and Evora, N.: Comparison of data-driven methods for downscaling ensemble weather forecasts, Hydrol. Earth Syst. Sci., 12, 615–624, 2008, http://www.hydrol-earth-syst-sci.net/12/615/2008/.

Lopez, A., Tebaldi, C., New, M., Stainforth, D., Allen, M., and Kettleborough, J.: Two approaches to quantifying uncertainty in global temperature changes, J. Climate, 19, 4785–4796, 2006.

McKerchar, A. I. and Henderson, R. D.: Shifts in flood and low-flow regimes in New Zealand due to interdecadal climate variations, Hydrolog. Sci. J., 48, 637–654, 2003.

Mitchell, T. D. and Hulme, M.: Predicting regional climate change: Living with uncertainty, Prog. Phys. Geog., 23, 57–78, 1999.

Min, S. K. and Hense, A.: A Bayesian assessment of climate change using multimodel ensembles. Part I: Global mean surface temperature, J. Climate, 19, 3237–3256, 2006.

6558

Min, S. K. and Hense, A.: Hierarchical evaluation of IPCC AR4 coupled climate models with systematic consideration of model uncertainties, Clim. Dynam., 29, 853–868, 2007.

Mpelasoka, F. S., Mullan, A. B., and Heerdegen, R. G.: New Zealand climate change information derived by multivariate statistical and artificial neural networks approaches, Int. J. Climatol., 21, 1415–1433, 2001.

Pasini, A., Lore, M., and Ameli, F.: Neural network modelling for the analysis of forcings/temperatures relationships at different scales in the climate system, Ecol. Model., 191, 58–67, 2006.

Qian, S. S., Stow, C. A., and Borsuk, M. E.: On Monte Carlo methods for Bayesian inference, Ecol. Model., 159, 269–277, 2003.

Ruosteenoja, K., Tuomenvirta, H., and Jylhä, K.: GCM-based regional temperature and precipitation change estimates for Europe under four SRES scenarios applying a super-ensemble pattern-scaling method, Climatic Change, 81, 193–208, 2007.

Semenov, M. A. and Barrow, E. M.: LARS-WG A Stochastic Weather Generator for Use in Climate Impact Studies, Developed by Mikhail A. Semenov, Version 3.0, User Manual, 2002.

Semenov, M. A. and Barrow, E. M.: Use of a stochastic weather generator in the development of climate change scenarios, Climatic Change, 35, 397–414, 1997.

Semenov, M. A., Brooks, R. J., Barrow, E. M., and Richardson, C. W.: Comparison of the WGEN and LARS-WG stochastic weather generators for diverse climates, Clim. Res., 10, 95–107, 1998.

SRES (Special Report on Emission Scenarios): Intergovernmental Panel on Climate Change (IPCC), http://www.grida.no/climate/ipcc/emission/index.htm, 2000.

Tebaldi, C., Mearns, L. O., Nychka, D., and Smith, R. L.: Regional probabilities of precipitation change: A Bayesian analysis of multimodel simulations, Geophys. Res. Lett., 31, 1–5, 2004.

Tebaldi, C., Smith, R. L., Nychka, D., and Mearns, L. O.: Quantifying uncertainty in projections of regional climate change: A Bayesian approach to the analysis of multimodel ensembles, J. Climate, 18, 1524–1540, 2005.

Tripathi, S., Srinivas, V. V., and Nanjundiah, R. S.: Downscaling of precipitation for climate change scenarios: A support vector machine approach, J. Hydrol., 330, 621–640, 2006.

Wang, Q. J.: A Bayesian method for multi-site stochastic data generation: Dealing with non-concurrent and missing data, variable transformation and parameter uncertainty, Environ. Modell. Softw., 23, 412–421, 2008.

Wilby, R. L., Charles, S. P., Zorita, E., Timbal, B., Whetton, P., and Mearns, L. O.: Guidelines

6559

for use of climate scenarios developed from statistical downscaling methods, Supporting material of the Intergovernmental Panel on Climate Change, available from the DDC of IPCC TGCIA, 27, 2004.

Wilby, R. L., Dawson, C. W., and Barrow, E. M.: SDSM-a decision support tool for the assessment of regional climate change impacts, Environ. Modell. Softw., 17, 145–157, 2002.

Wilby, R. L. and Harris, I.: A framework for assessing uncertainties in climate change impacts: Low-flow scenarios for the River Thames, UK, Water Resour. Res., 42, W02419, doi:10.1029/2005WR004065, 2006.

Wilby, R. L., Hassan, H., and Hanaki, K.: Statistical downscaling of hydrometeorological variables using general circulation model output, J. Hydrol., 205, 1–19, 1998.

Xu, C. Y.: From GCMs to river flow: A review of downscaling methods and hydrologic modelling approaches, Prog. Phys. Geog., 23, 229–249, 1999.

Table 1. Name and description of all NCEP predictors on HadCM3 grid (bold texted predictors were used for model calibration).

Sr. no.	Predictor	Description	Sr. no.	Predictor	Description
1	**ncepmslpaz**	**Mean sea level pressure**	14	ncepp500az	500 hPa geopotential height
2	ncepp5_faz	500 hPa airflow strength	15	**ncepp850az**	**850 hPa geopotential height**
3	**ncepp5_uaz**	**500 hPa zonal velocity**	16	ncepp__faz	Surface airflow strength
4	ncepp5_vaz	500 hPa meridional velocity	17	ncepp__uaz	Surface zonal velocity
5	**ncepp5_zaz**	**500 hPa vorticity**	18	ncepp__vaz	Surface meridional velocity
6	ncepp5thaz	500 hPa wind direction	19	**ncepp__zaz**	**Surface vorticity**
7	**ncepp5zhaz**	**500 hPa divergence**	20	ncepp_thaz	Surface wind direction
8	ncepp8_faz	850 hPa airflow strength	21	ncepp_zhaz	Surface divergence
9	**ncepp8_uaz**	**850 hPa zonal velocity**	22	**ncepr500az**	**Relative humidity at 500 hPa**
10	**ncepp8_vaz**	**850 hPa meridional velocity**	23	ncepr850az	Relative humidity at 850 hPa
11	**ncepp8_zaz**	**850 hPa vorticity**	24	nceprhumaz	Near surface relative humidity
12	ncepp8thaz	850 hPa wind direction	25	ncepshumaz	Surface specific humidity
13	ncepp8zhaz	850 hPa divergence	26	nceptempaz	Mean temperature at 2 m

Table 2. Results of the cross correlation analysis.

Predictor	Description	Optimal lag with Clutha precipitation
ncepmslpaz	Mean sea level pressure	+1
ncepp5_faz	500 hPa airflow strength	+1
ncepp5_uaz	500 hPa zonal velocity	+1
ncepp5_vaz	500 hPa meridional velocity	+1
ncepp5thaz	500 hPa wind direction	−1
ncepp5zhaz	500 hPa divergence	+1
ncepp8_faz	850 hPa airflow strength	+1
ncepp8_uaz	850 hPa zonal velocity	+1
ncepp8_vaz	850 hPa meridional velocity	+1
ncepp8_zaz	850 hPa vorticity	+1
ncepp8thaz	850 hPa wind direction	+1
ncepp8zhaz	850 hPa divergence	+1
ncepp500az	500 hPa geopotential height	+3
All other predictors showed an optimal lag of zero.		

Table 3. Relative change in monthly statistics derived from HadCM3 daily output for present and A2 future scenario.

Month	Mean daily	Wet spell length	Dry spell length
Jan	1.01	0.90	1.13
Feb	1.13	0.99	1.21
Mar	1.01	0.82	1.05
Apr	0.98	0.99	1.14
May	1.30	1.17	1.09
Jun	1.16	1.10	1.13
Jul	1.20	1.47	1.11
Aug	1.10	1.03	1.08
Sep	1.10	1.15	1.28
Oct	1.14	1.38	1.06
Nov	1.07	0.92	1.07
Dec	0.96	0.98	1.06

Table 4. Natural variability of observed data calculated as described in Tebaldi et al. (2005).

Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
0.25	0.54	0.56	0.48	2.49	1.79	1.74	2.20	0.51	1.25	0.53	1.17

Table 5. Month-wise relative weighting of the three downscaling models.

Month	Relative Weights (%)		
	SDSM	LAR-WG	GEP
Jan	45.78	45.41	8.81
Feb	44.83	23.94	31.22
Mar	37.25	54.72	38.73
Apr	44.70	11.95	43.35
May	33.54	39.30	27.16
Jun	3.85	46.48	49.67
Jul	39.33	20.42	40.25
Aug	45.59	48.48	5.93
Sep	33.32	44.61	22.07
Oct	45.92	34.63	19.45
Nov	40.24	41.86	17.90
Dec	51.29	36.18	12.52

Fig. 1. Watershed boundary of the Clutha River above Balclutha on a terrain map of South Island, New Zealand (Source: NIWA).
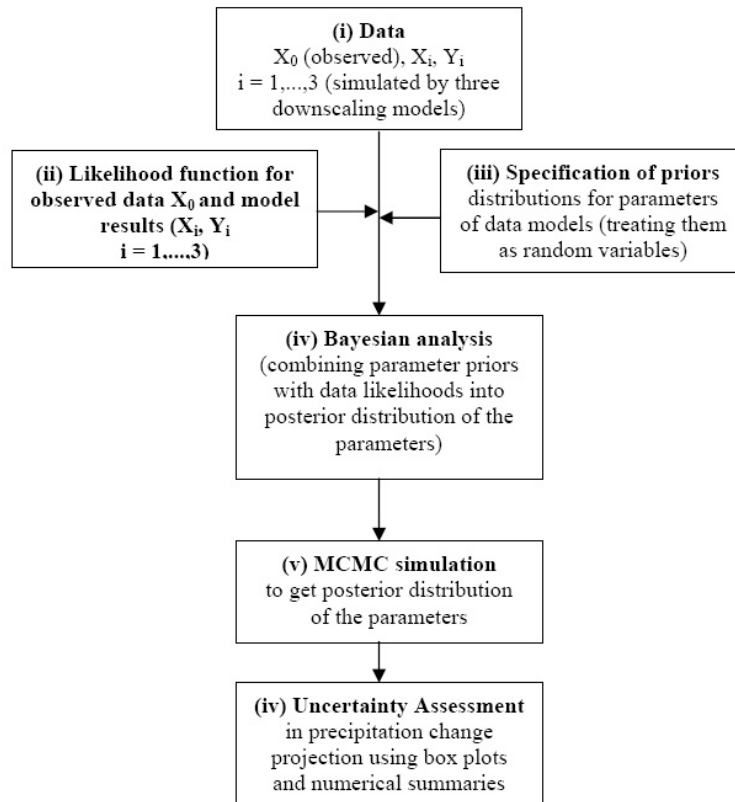
Fig. 2. Schematic diagram outlining the steps involved in the Bayesian WMME approach.
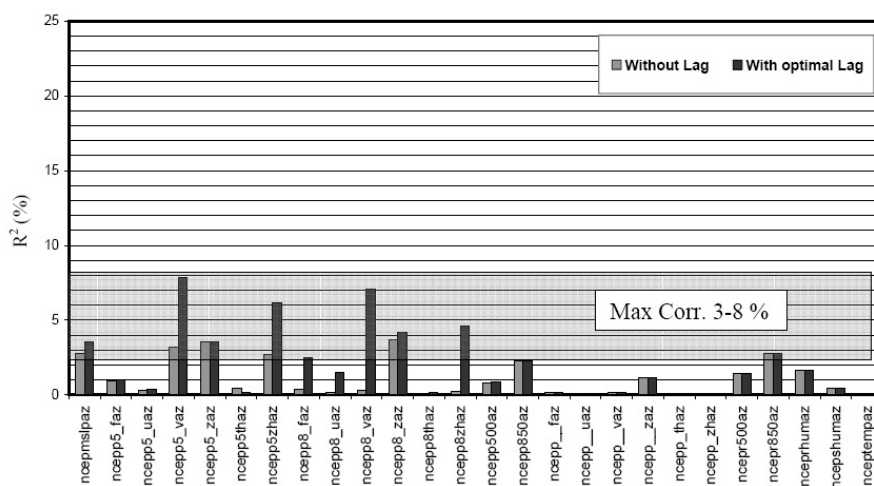
Fig. 3. Correlation between Ophir2 precipitation data with NCEP predictors and maximum range of correlation.
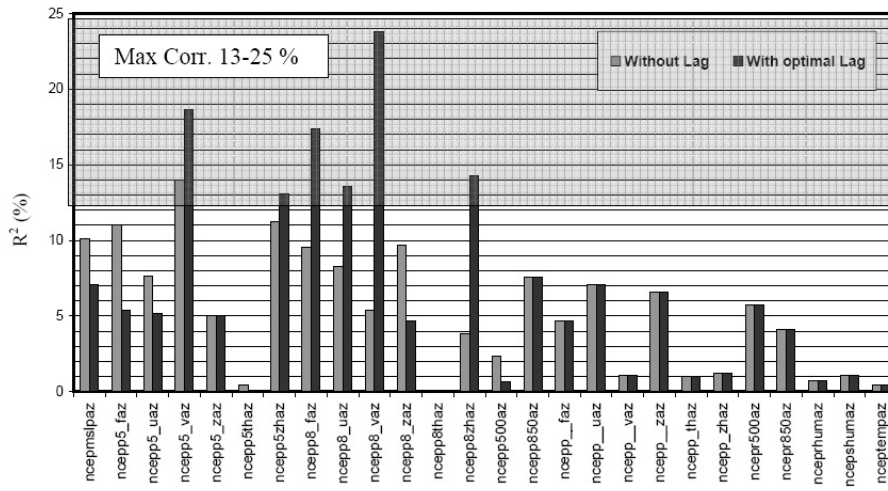
Fig. 4. Correlation between average Clutha precipitation data with NCEP predictors and maximum range of correlation.
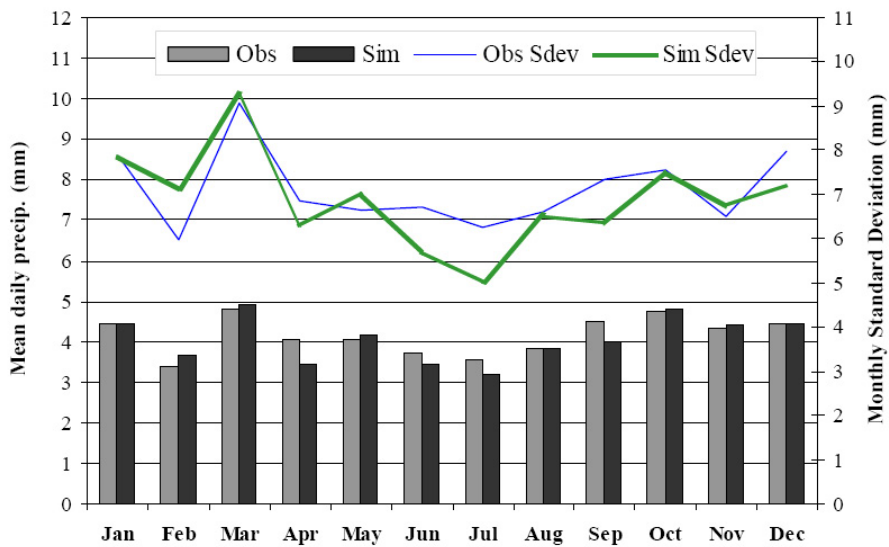
Fig. 5. SDSM model calibration for mean daily precipitation of 1961–1990.
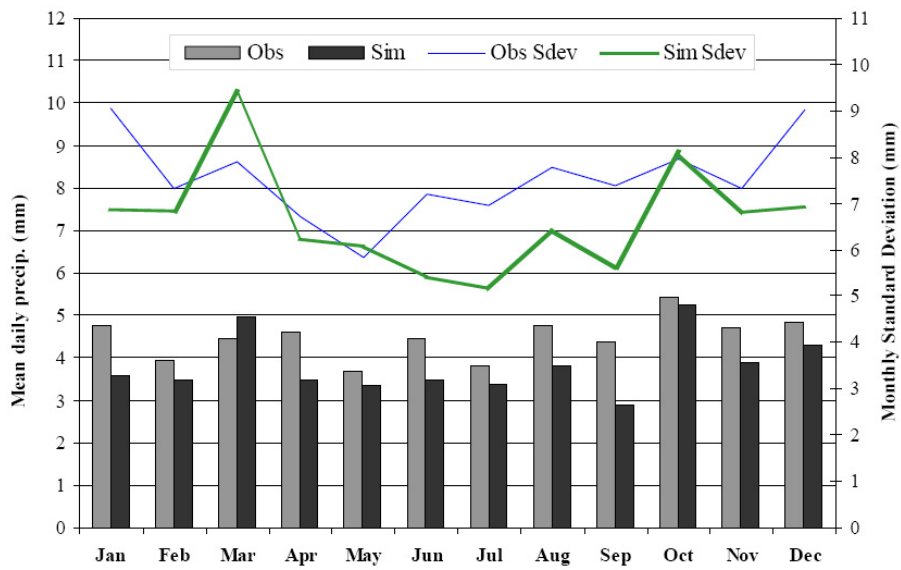
Fig. 6. SDSM model validation for mean daily precipitation of 1991–2000.
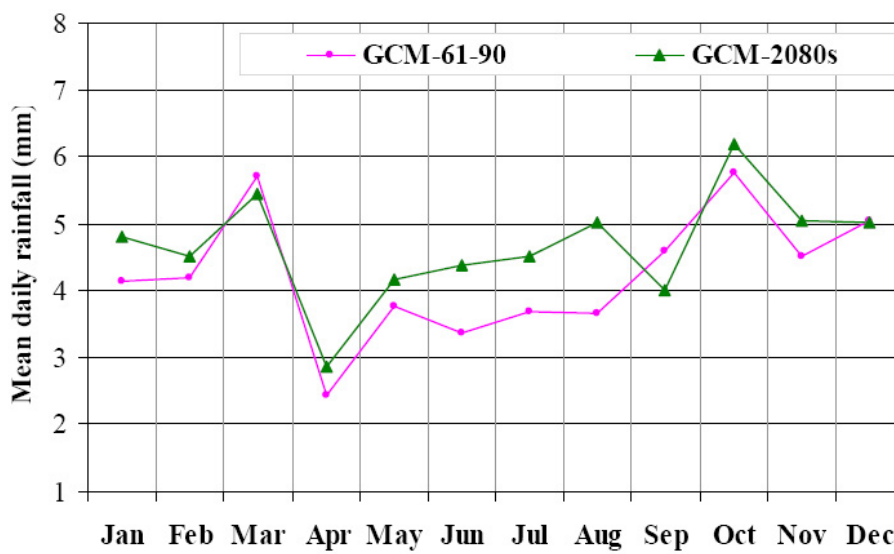
Fig. 7. SDSM: Downscaling results using HadCM3 predictors.
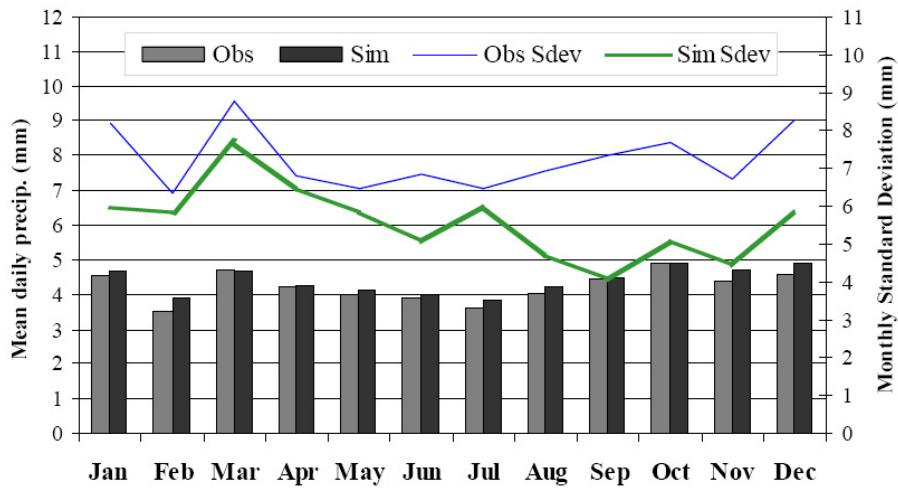
Fig. 8. Comparison of Clutha precipitation and LARS-WG simulated data for 1961–2000.
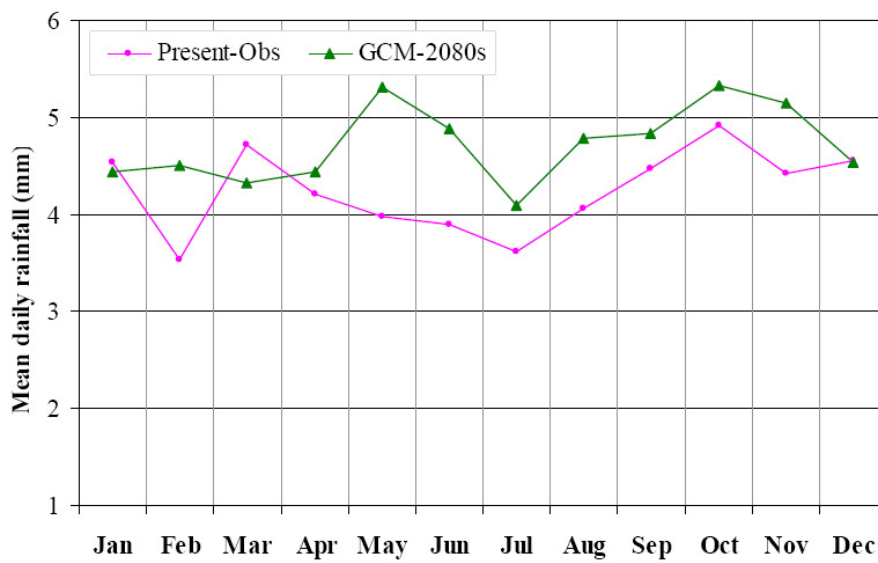
Fig. 9. LARS-WG: Comparison of present and HadCM3 predicted future Clutha precipitation.
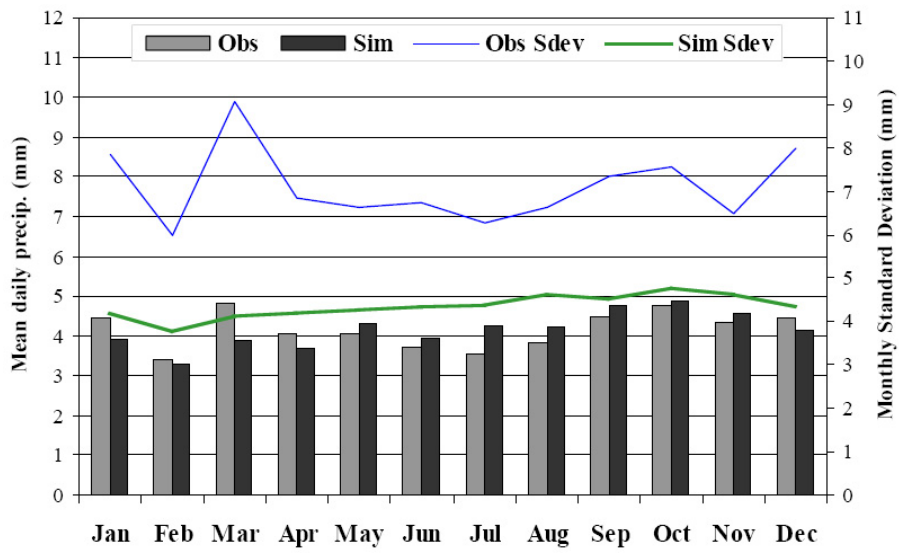
Fig. 10. GEP model training results in terms of mean daily precipitation of 1961–1990.
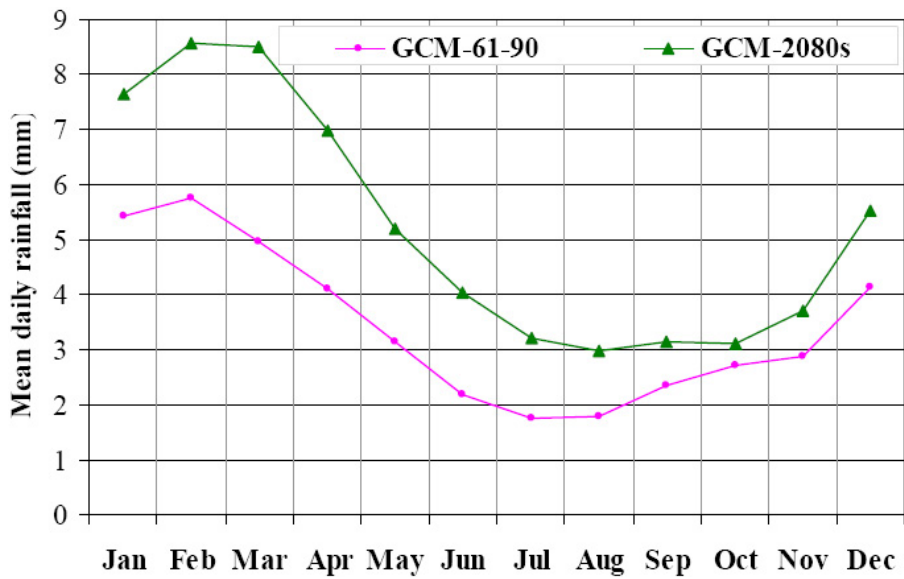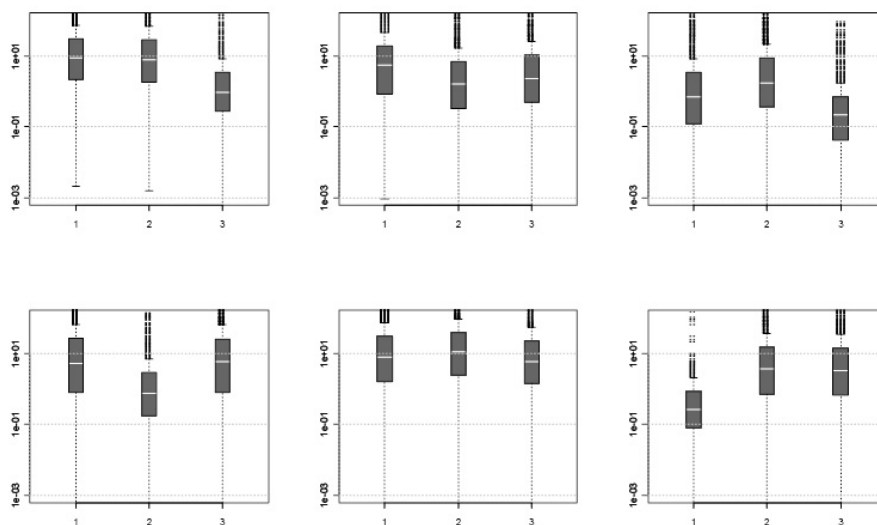
Fig. 11. GEP model: Downscaling using HadCM3 data.

Fig. 12a. Monthly posterior distribution of model specific precision parameter (λ_i) for three models 1=SDSM, 2=LARS-WG, 3=GEP. Top three plots for January, February, March bottom three for April, May, June.
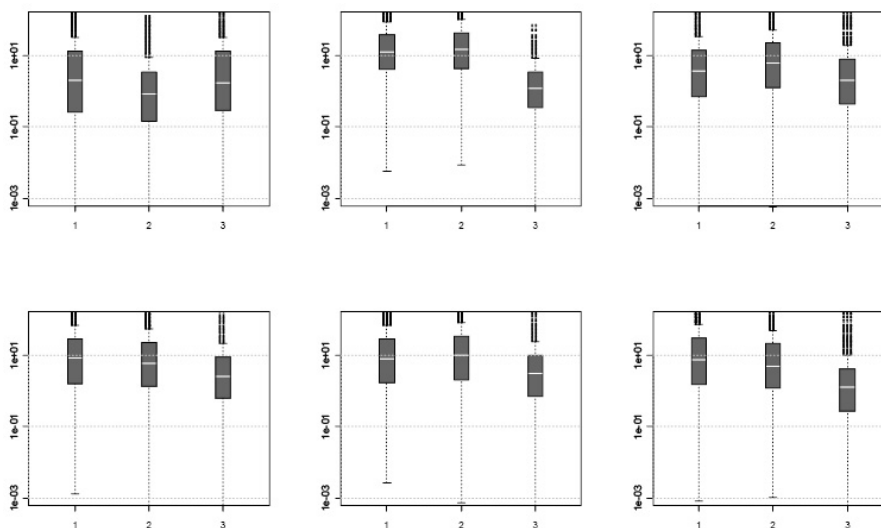
Fig. 12b. Monthly posterior distribution of model specific precision parameter (λ_i) for three models 1=SDSM, 2=LARS-WG, 3=GEP. Top three plots for July, August, September, bottom three for October, November, December.
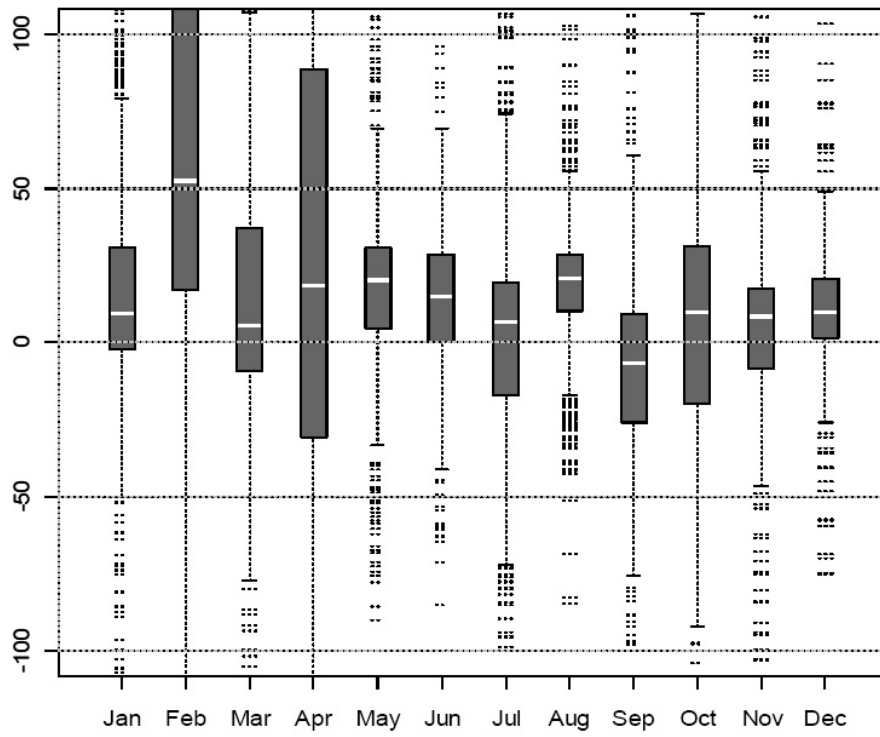
Fig. 13. Month-wise posterior distribution of percent precipitation change on the basis of Bayesian WMME.

6579