Approximating uncertainty of annual runoff and reservoir yield using stochastic replicates of Global Climate Model data

M. C. Peel[1], R. Srikanthan[2], T. A. McMahon[1] and D. J. Karoly[3]

[1]{Department of Infrastructure Engineering, University of Melbourne, Victoria, 3010, Australia}

[2]{Environment and Research Division, Bureau of Meteorology, Melbourne, Victoria, 3001, Australia}

[3]{School of Earth Sciences and ARC Centre of Excellence for Climate System Science, University of Melbourne, Victoria, 3010, Australia}

Correspondence to: M. C. Peel (mpeel@unimelb.edu.au)

Abstract

Two key sources of uncertainty in projections of future runoff for climate change impact assessments are uncertainty between Global Climate Models (GCMs) and within a GCM. Within-GCM uncertainty is the variability in GCM output that occurs when running a scenario multiple times but each run has slightly different, but equally plausible, initial conditions. The limited number of runs available for each GCM and scenario combination within the Coupled Model Intercomparison Project phase 3 (CMIP3) and phase 5 (CMIP5) datasets, limits the assessment of within-GCM uncertainty. In this second of two companion papers, the primary aim is to present a proof-of-concept approximation of within-GCM uncertainty for monthly precipitation and temperature projections and to assess the impact of within-GCM uncertainty on modelled runoff for climate change impact assessments. A secondary aim is to assess the impact of between-GCM uncertainty on modelled runoff. Here we approximate within-GCM uncertainty by developing non-stationary stochastic replicates of GCM monthly precipitation and temperature data. These replicates are input to an off-line hydrologic model to assess the impact of within-GCM uncertainty on projected annual runoff and reservoir yield. We adopt stochastic replicates of available GCM runs to approximate

within-GCM uncertainty because large ensembles, hundreds of runs, for a given GCM and scenario are unavailable, other than the Climateprediction.net dataset for the Hadley Centre GCM. To-date within-GCM uncertainty has received little attention in the hydrologic climate change impact literature and this analysis provides an approximation of the uncertainty in projected runoff, and reservoir yield, due to within- and between-GCM uncertainty of precipitation and temperature projections. In the companion paper, McMahon et al., (2014) sought to reduce between-GCM uncertainty by removing poorly performing GCMs, resulting in a selection of five better performing GCMs from CMIP3 for use in this paper. Here we present within- and between-GCM uncertainty results in mean annual precipitation (MAP), temperature (MAT) and runoff (MAR), the standard deviation of annual precipitation (SDP) and runoff (SDR) and reservoir yield for five CMIP3 GCMs at 17 world-wide catchments. Based on 100 stochastic replicates of each GCM run at each catchment, within-GCM uncertainty was assessed in relative form as the standard deviation expressed as a percentage of the mean of the 100 replicate values of each variable. The average relative within-GCM uncertainty from the 17 catchments and 5 GCMs for 2015-2044 (A1B) were: MAP 4.2%, SDP 14.2%, MAT 0.7%, MAR 10.1% and SDR 17.6%. The Gould-Dincer Gamma procedure was applied to each annual runoff time-series for hypothetical reservoir capacities of 1xMAR and 3xMAR and the average uncertainty in reservoir yield due to within-GCM uncertainty from the 17 catchments and 5 GCMs were: 25.1% (1xMAR) and 11.9% (3xMAR). Our approximation of within-GCM uncertainty is expected to be an underestimate due to not replicating the GCM trend. However, our results indicate that within-GCM uncertainty is important when interpreting climate change impact assessments. Approximately 95% of values of MAP, SDP, MAT, MAR, SDR and reservoir yield from 1xMAR or 3xMAR capacity reservoirs are expected to fall within twice their respective relative uncertainty (standard deviation/mean). Within-GCM uncertainty has significant implications for interpreting climate change impact assessments that report future changes within our range of uncertainty for a given variable – these projected changes may be due solely to within-GCM uncertainty. Since within-GCM variability is amplified from precipitation to runoff and then to reservoir yield, climate change impact assessments that do not take into account within-GCM uncertainty risk providing water resources management decision makers with a sense of certainty that is unjustified.

1 Introduction

This study is part of a research project that seeks to enhance our understanding of the uncertainty of future annual river flows, leading to more informed decision-making for the sustainable management of scarce water resources. This is the second of two papers examining the uncertainty of streamflow estimates derived from global climate models (GCMs). In the first paper, McMahon et al., (2014) assessed the adequacy of GCMs from phase 3 of the Coupled Model Inter-comparison Project (CMIP3, Meehl et al., 2007) to simulate observed values of mean annual precipitation, standard deviation of annual precipitation, mean annual temperature, monthly patterns of precipitation and temperature, and Köppen climate classification. Five GCMs (HadCM3, MIROCM, MIUB, MPI and MRI, see Table 1 of McMahon et al., 2014 for full GCM names) were selected as better performing GCMs for use in this second paper.

In this paper we address a significant limitation to characterising the uncertainty of future runoff which is the lack of sufficient GCM runs of historical (20C3M) and future projections (e.g., A1B). Modelling historical runoff involves numerous uncertainties (Peel and Blöschl, 2011) including uncertainties in observed input data used to drive the hydrologic model (Andréassian et al., 2004; McMillan et al., 2011), observed data against which the hydrologic model is calibrated (Di Baldassarre and Montanari, 2009; McMillan et al., 2010), the calibration method and objective function adopted (Efstratiadis and Koutsoyiannis, 2010) and the hydrologic model structure itself (Andréassian et al., 2009; Vogel and Sankarasubramanian, 2003). Additional uncertainty is introduced when modelling future runoff through (1) assuming the hydrologic model calibration applies into the future (Chiew et al., 2014), (2) assuming a bias correction for adjusting GCM data developed over the observed period applies into the future and (3) through differences in future climate projections between GCMs and within a GCM. Recent investigations into uncertainty introduced at different stages of the model train, from GCM to hydrologic model, for climate change impact assessments include Bosshard et al. (2013), Dobler et al. (2012), Hingray and Saïd (2014), Kay et al. (2009), Lafaysse et al. (2014), Prudhomme and Davies (2009a; 2009b), Steinschneider et al. (2012), Teng et al. (2012), and Woldemeskel et al. (2014).

The uncertainty between GCM projections of future climate can be assessed through analysis of runs from a wide range of GCMs, such as those available from CMIP3 and being collated within the Coupled Model Intercomparison Project phase 5 (CMIP5). Our selection of five

better performing GCMs from CMIP3 in the companion paper (McMahon et al., 2014) is an attempt to reduce between-GCM uncertainty by removing poorly performing GCMs from the analysis conducted in this paper. Our primary aim in this proof-of-concept paper is to present an approximation of within-GCM uncertainty, which is the variability in GCM output that occurs when running a scenario multiple times but each run has slightly different, but equally plausible, initial conditions. Although the importance of within-GCM uncertainty for climate change impact assessments has been highlighted by Tebaldi and Knutti (2007), Hawkins and Sutton (2009; 2011) and Deser et al (2012; 2014), to-date it has received little attention in the hydrology climate change impact literature. Here we develop an approximation of within-GCM uncertainty and apply it to a climate change impact assessment for future runoff and reservoir yield.

The magnitude of within-GCM uncertainty for a metric like mean annual precipitation can be assessed directly from GCM output if a large enough ensemble of runs from a GCM for a given emission scenario are available. The number of GCM runs required to adequately assess uncertainty depends upon the metric of interest and the level of confidence adopted. For example, for a given level of confidence an extreme value metric will require many more runs than a mean to obtain a reliable estimate. For a more detailed discussion of this issue see Salas (1992). Currently, large ensembles of runs from each GCM and scenario are unavailable. In the CMIP3 dataset most GCMs have a single run of a given scenario from which a direct assessment of within-GCM uncertainty is impossible. In terms of ensemble members CMIP5 is an improvement over CMIP3 in that more runs of each scenario are being reported for each GCM. However, the number of runs per GCM and scenario combination in CMIP5 is still in the order of 3 to 10, rather than the hundreds of runs required for adequate estimation of within-GCM uncertainty of some metrics. The Climate*prediction*.net dataset contains an ensemble of several thousand runs from the Hadley Centre GCM (Frame et al., 2009; Rowlands et al., 2012). However, this dataset can only be used to directly assess within-GCM uncertainty for the Hadley Centre GCM.

Previous assessments of the impact of within-GCM uncertainty on runoff have been limited by the lack of available GCM runs. For example, when investigating sources of uncertainty in the climate change impact on hydrology, Chen et al. (2011) were limited to 5 runs with different initial conditions from the MRI GCM. Similarly, Velázquez et al. (2013) were limited to 5 runs from one GCM and 3 runs from a second in their comparison of the

uncertainty due to hydrologic models and within-GCM uncertainty. Prudhomme and Davies (2009a) sought to overcome the limited number of GCM runs by introducing a seasonal block-resampling technique to estimate natural climate variability via 100 bootstrap replicates of observed and GCM time series. In Prudhomme and Davies (2009b) they applied seasonal block-resampling to a 30-year baseline period and future period to assess whether climate change impacts were significantly different to baseline climate variability. However, seasonal block-resampling is unable to address inter-decadal variability, as noted by Kay et al. (2009), or periods with significant trend as the bootstrap replicates will scramble any inter-decadal variability or trend. Finally, Hingray and Saïd (2014) and Lafaysse et al. (2014) adopted a stochastic approach whereby they generated 100 stochastic replicates from each of six statistical downscaling models for each of 11 runs from 5 GCMs (Hingray and Saïd, 2014), or 12 runs from 6 GCMs (Lafaysse et al., 2014). They used this multi-model ensemble of stochastic replicates to investigate the magnitude of within- and between-GCM uncertainty for the Durance catchment in France.

In this proof-of-concept paper we develop an approximation of within-GCM uncertainty using non-stationary stochastic replicates of GCM monthly precipitation and temperature data that seeks to preserve any inter-decadal variability and trend. Unlike Hingray and Saïd (2014) and Lafaysse et al. (2014) whose replicates were produced by the statistical downscaling model, here we stochastically replicate the original GCM runs prior to downscaling. Estimating uncertainty in a time series metric via stochastic modelling of a time series is standard hydrologic practice (Hipel and McLeod, 1994). A stochastic model is fit to the time series of interest and an ensemble of time series replicates with the same stochastic properties as the original series is generated. The metric of interest is calculated for each ensemble member and the metric uncertainty is estimated from the distribution of metric values. In this paper we stochastically replicate the GCM output data, use an ensemble of stochastic replicates as input to an off-line hydrologic model to estimate an ensemble of future runoff projections, from which we estimate the variability in mean and variance of annual runoff. Finally, the ensemble of future runoff projections is used to investigate the impact of within- and between-GCM uncertainty on future reservoir yield.

In this paper we model runoff in an off-line hydrologic model rather than adopt GCM generated runoff. Arora (2001) demonstrated the quality of GCM runoff mainly depends on the quality of GCM precipitation, with any bias in precipitation amplified in the resulting

runoff. In our companion paper (McMahon et al., 2014) we assessed GCM bias in reproducing observed precipitation conditions and found substantial biases for all GCMs, thus we would expect significant bias in runoff generated by a GCM. Furthermore, Sperna Weiland et al. (2012) found that runoff estimates from an external hydrologic model generally outperformed GCM runoff estimates. However, Sperna Weiland et al. (2012) noted that when the GCM Land Surface Scheme is specifically tuned to reproduce observed runoff and a routing scheme is added then GCM runoff becomes more acceptable. We also use the terms streamflow and runoff interchangeably and adopt depth (in mm) as a measure of flow rather than a volume unit.

Following this introduction, in Section 2 we outline the approximation methodology (Ensemble Empirical Mode Decomposition, stochastic data generation, quantile-quantile bias correction of precipitation and temperature, precipitation-evapotranspiration-runoff modelling, and uncertainty in reservoir yield) and related literature. We test our stochastic within-GCM uncertainty approximation for the largest ensemble of GCM runs in the CMIP3 dataset for a given GCM and scenario in Section 3. In Section 4 results of applying the methodology to output from five GCMs identified in the companion paper (McMahon et al., 2014) are presented and discussed. Conclusions from the analysis and discussion are presented in Section 5. Further details about the precipitation-evapotranspiration-runoff model, source code and example input and output are provided in the Supplementary Material Section.

2 Methodology and related literature

2.1 Overall methodology

The methodology to approximate within-GCM uncertainty and assess the impact of within- and between-GCM uncertainty on future runoff and reservoir yield is shown in Figure 1. Five better performing GCMs were identified in the companion paper for use in this paper through a literature review and assessment of how well CMIP3 GCMs reproduced observed mean annual precipitation, annual temperature and average monthly precipitation and temperature at GCM grid cell scales (McMahon et al., 2014). The five GCMs identified were HadCM3, MIROCM(1), MIUB(1), MPI(1) and MRI(3), where the number in brackets refers to the run number for that GCM in the CMIP3 data set (see McMahon et al., 2014, Table 1 for full

GCM names). As part of the analysis in McMahon et al., (2014) catchment average values of concurrent monthly precipitation and temperature for the 20C3M and A1B emissions scenarios were extracted from each GCM in the CMIP3 data set. The catchment average was calculated for each catchment and GCM combination by determining the proportion of catchment area associated with each GCM grid cell and performing an area weighted average of the GCM data for each month. Catchment average precipitation and temperature from the five better performing GCMs are used throughout this paper.

An ideal assessment of within-GCM uncertainty would involve analysis of between hundreds of runs of a single GCM for a given scenario with each run having slightly different, but equally plausible, initial conditions. Each run in this ideal ensemble would have a different sequence of monthly values and a different overall trend. How different the monthly sequence and overall trend is from one run to the next represents the within-GCM uncertainty. In this paper we do not seek to approximate the overall trend, as this information is best provided by a GCM responding to an emissions scenario. Here we approximate differences in the monthly sequence around the trend by using stochastic data generation. To achieve this we de-trend the catchment average GCM data, stochastically replicate the de-trended series and add the trend to the stochastic data to form a stochastic replicate of the GCM data for the entire period of GCM record. In this way we approximate the uncertainty around the overall trend, but not the uncertainty in the trend. Therefore, the approximation presented here represents an under-estimate of the true within-GCM uncertainty as the trends used are restricted to those available in GCM runs in the CMIP3 data set. This stochastic methodology is a temporary solution for approximating within-GCM uncertainty until sufficient GCM runs become available to directly estimate within-GCM uncertainty from a large ensemble of GCM runs.

The procedure adopted here to approximate within-GCM uncertainty for a catchment consists of the following steps (see Figure 1):

1. De-trend the 20C3M and A1B catchment average GCM monthly precipitation and temperature data using ensemble empirical mode decomposition (EEMD). EEMD also allows any low frequency signals in the time-series to be identified.
2. Generate stochastically at a monthly time-step k replicates (where k is arbitrarily adopted as 100 to demonstrate the proof-of-concept) of precipitation and temperature of ~250 years (~150 from 20CM3 and 100 from A1B) ensuring the cross-correlations and auto-correlations in the precipitation and temperature time-series are preserved

and any significant low frequency signals are taken into account. The length of available 20C3M data differs between GCMs, hence there are approximately 250 years of monthly data to replicate.

3. Add the appropriate trend to the time-series for each replicate of monthly precipitation and temperature.
4. Bias-correct both the precipitation and the temperature time-series using the quantile-quantile procedure.
5. Calibrate the precipitation-evapotranspiration-runoff monthly model (PERM) for each catchment using observed precipitation, temperature and runoff data.
6. Model runoff using PERM and the bias-corrected stochastic replicates of GCM monthly precipitation and temperature.
7. Compute mean annual runoff (MAR), standard deviation of annual runoff (SDR), the lag-1 serial correlation of annual runoff (lag-1) and hypothetical reservoir yield for each replicate.
8. Estimate the within- and between-GCM uncertainty in MAR, SDR and lag-1 and hypothetical reservoir yield based on the 100 replicates.

An advantage of this methodology over a bootstrap based methodology, like Prudhomme and Davies (2009a), is that the entire period of the GCM run is replicated, which can preserve any inter-decadal variability and trend in the replicates. The replicates can be used to drive a hydrologic model for the entire period (~250 years), rather than for short (~30 years) separate periods, since the overall trend has been preserved. Therefore, hydrologic model stores will be representative of prior conditions at the beginning of any future period of runoff impact assessment.

2.2 De-trend GCM data

GCM projections of precipitation and temperature are non-stationary in terms of mean and existing stochastic data generation techniques generally deal with stationary data. In order to apply existing stochastic methods we de-trend the GCM monthly precipitation and temperature data using Ensemble Empirical Mode Decomposition (EEMD).

The original Empirical Mode Decomposition (EMD) algorithm, introduced by Huang et al. (1998), is an adaptive spectral analysis technique that is robust when applied to non-linear and non-stationary data. EMD decomposes a time-series into a set of intrinsic mode functions (IMFs) and a residual. Each IMF is a zero-mean fluctuation in which the frequency and amplitude may vary within a given IMF. Subsequent IMFs represent progressively lower frequency fluctuations. The EMD residual captures any trend in a time-series which may be an unresolved low frequency fluctuation with an average period longer than the period of record or a linear or non-linear trend. The nature of the EMD residual is not assumed prior to running the algorithm, rather it is a data-driven output. More recently, Wu and Huang (2009) proposed Ensemble EMD (EEMD), a noise assisted data analysis procedure as an improvement over the original EMD. In EEMD, an ensemble of EMD trials is obtained by adding white noise of finite amplitude to the time-series prior to each EMD run. The IMFs and residual from each trial are grouped by IMF order into ensembles and the average of each IMF group and the average residual yield the EEMD result. Because the white noise is different for each EMD trial, during averaging the noise cancels out as the ensemble size increases. The purpose of the noise is to change the ordering of local maxima and minima within the time-series, thus generating a different EMD outcome in each trial. Details are given in Wu and Huang (2009) and an application to the Southern Oscillation Index is presented by Peel et al. (2011b) and to Australian monthly rainfall and temperature by Srikanthan et al. (2011).

For this analysis the relevant features of EEMD are the residual, which represents the time-series trend, and any low frequency signal in the GCM data. Some GCMs reproduce features of the El Niño–Southern Oscillation (ENSO) and associated low frequency variability (van Oldenborgh et al., 2005). For GCMs with an ENSO signal in precipitation we would like to maintain this information in the stochastic replicates. To identify low frequency signals in GCM data we follow Wu and Huang (2004) and compare each set of EEMD results against a white noise model. Low frequency IMFs (average period > 2 years) with more variance than expected from a white noise model are considered a low frequency signal. The white noise model is an ensemble of 200 EMD results from white noise series of the same length and variance as the GCM series.

EEMD was applied to 20C3M and A1B precipitation and temperature data and the residual (trend) identified. For temperature data all IMFs are summed together to form a de-trended

time-series ready for stochastic replication. For precipitation data, where a low frequency signal is not present all IMFs are summed together to form a de-trended time-series ready for stochastic replication. Where a low frequency precipitation signal is identified, all IMFs with an average period ≤ 2 years are summed to form a high frequency component and all IMFs with an average period > 2 years are summed to form a low frequency component.

In this EEMD analysis we use rational spline EMD (Pegram et al., 2008) with tension parameter = 0.5 and reflective spline end condition (Peel et al., 2009). Each EEMD analysis has 200 ensemble members, the standard deviation of white noise added to the series represents 0.4 of the original series standard deviation and non-orthogonal IMFs, IMF pairs with Orthogonality Index > 0.1, are automatically combined following Peel et al. (2011a).

2.3 Stochastic data generation

In this step we approximate uncertainty around the GCM trend by generating stochastic replicates of de-trended GCM catchment average time-series of concurrent monthly precipitation and temperature. In order to preserve any cross correlation between the precipitation and temperature series and their auto-correlations, the Matalas (1967) multi-site stochastic data generation procedure was adopted. In order to preserve any low frequency precipitation information the generation procedure also needs to be able to simulate both high and low frequency time-series. To achieve this we adapt the method of McMahon et al. (2008) who used EMD to decompose six-monthly precipitation data into intra- and inter-decadal components, replicated each component separately, and then combined the component replicates to form the six-monthly precipitation replicate. In this way their stochastic replicates were able to reproduce observed multi-year dry periods. Replicating intra- and inter-decadal components separately was possible in McMahon et al. (2008) as IMFs from EMD, and EEMD, are orthogonal to each other. In this paper we use EEMD to identify any low frequency component (> 2 years) in the precipitation and utilise the orthogonal nature of EEMD IMFs to replicate the high and low frequency components separately before combining the generated component replicates and the trend to form the overall replicate.

The first step in the data generation process is to remove the trends (one for precipitation and one for temperature) identified through EEMD analysis in Section 2.2 from the monthly precipitation and temperature time-series. If the GCM precipitation does not contain a low

frequency component then there are two separate time-series to replicate concurrently: (1) the de-trended temperature (sum of EEMD IMFs), and (2) the de-trended precipitation (sum of EEMD IMFs). If GCM precipitation does contain a low frequency component then the de-trended precipitation is divided into a high frequency component (sum of EEMD IMFs with average period ≤ 2 years) and a low frequency component (sum of EEMD IMFs with average period > 2 years), resulting in three time-series to replicate concurrently. Next, for each calendar month, these time-series are standardised to zero mean and unit variance. Data generation then takes place and the resulting standardised values are rescaled by the calendar monthly means and variances. Finally, the respective trends are added to the rescaled precipitation and temperature data to form the final replicates.

For sites without a low frequency precipitation component, the following auto-regressive lag-one (AR1) model is appropriate:

$$[X_t]_{2\times1} = [A]_{2\times2}[X_{t-1}]_{2\times1} + [B]_{2\times2}[\varepsilon_t]_{2\times1} \quad (1)$$

where $[X_t]_{2\times1} = \begin{bmatrix} P \\ T \end{bmatrix}_t$, $[X_{t-1}]_{2\times1} = \begin{bmatrix} P \\ T \end{bmatrix}_{t-1}$, $[A]_{2\times2}$ and $[B]_{2\times2}$ are 2×2 matrices of constant coefficients to preserve the cross-correlations between the standardised monthly P_t and T_t and their auto-correlations, $[\varepsilon_t]$ is a matrix of random skewed variates with a mean = 0, and variance = 1.

To take into account the skewness in a time-series, ε_t is defined by the Wilson-Hilferty transformation (Wilson and Hilferty, 1931) and replaced in Eq. (1) by:

$$\varepsilon_t = \left[\frac{2}{\gamma} \left(1 + \frac{\gamma\xi}{6} - \frac{\gamma^2}{36} \right)^3 - \frac{2}{\gamma} \right] \quad (2)$$

where ε_t is a random skewed variate (zero mean and unit variance) to account for the skewness in the standardised data defined by the coefficient of skewness γ, ξ is a random normal variate with zero mean and unit variance.

The matrices $[A]$ and $[B]$ are determined from (Matalas, 1967):

$$[A] = [M_1][M_0]^{-1} \quad (3)$$

$$[B][B]^T = [M_0] - [M_1][M_0]^{-1}[M_1]^T \quad (4)$$

where $[M_0]$ and $[M_1]$ are the lag zero and lag-one cross correlation matrices respectively. The elements of $[M_0]$ and $[M_1]$ corresponding to variables i and j are given by:

$$m_0^{ij} = \frac{1}{n}\sum_{i=1}^{n} X_i X_j \tag{5}$$

$$m_1^{ij} = \frac{1}{n-1}\sum_{i=1}^{n-1} X_i X_{j-1} \tag{6}$$

The matrix $[A]$ can be obtained from Eq. (3). The matrix $[B][B]^T$ in Eq. (4) is symmetric and should be positive semi-definite for solving $[B]$, where $[B]$ can be estimated by the Cholesky decomposition in which matrix $[B]$ is lower triangular (Hipel and McLeod, 1994). The elements b_{ij} of $[B]$ are obtained from the following recursive relationships:

$$b_{ij} = 0, j > i \tag{7}$$

$$b_{11} = \sqrt{c_{11}} \tag{8}$$

where c_{ij} is the element of the matrix $[B]$ in Eq. (4). The remaining element in the first column of $[B]$ is given by:

$$b_{12} = \frac{c_{12}}{b_{11}} \tag{9}$$

The second diagonal element is obtained from:

$$b_{22} = c_{22} - b_{21}^2 \tag{10}$$

Once matrices $[A]$ and $[B]$ are determined, 100 replicates of standardised skewed values are generated using Equations (1) and (2), then rescaled by the standard deviation and the mean of the monthly calendar values and trend added to obtain the stochastic replicates of monthly precipitation and temperature data. In general, the generated monthly data will not preserve the annual characteristics especially for precipitation as it is highly variable. The generated monthly temperature data were found to preserve the annual characteristics, while the generated precipitation did not. Hence, the generated monthly precipitation data were nested in an annual model (Srikanthan 2004). The details of the nesting are described later.

For sites with a low frequency precipitation component, an AR(2) model is used to incorporate the low frequency component. A general multi-site AR(2) model takes the form:

$$[X_t]_{3\times1} = [A]_{3\times3}[X_{t-1}]_{3\times1} + [B]_{3\times3}[X_{t-2}]_{3\times1} + [C]_{3\times3}[\varepsilon_t]_{3\times1} \quad (11)$$

where $[X_t]_{3\times1} = \begin{bmatrix} P_H \\ P_L \\ T \end{bmatrix}_t$, $[X_{t-1}]_{3\times1} = \begin{bmatrix} P_H \\ P_L \\ T \end{bmatrix}_{t-1}$, $[X_{t-2}]_{3\times1} = \begin{bmatrix} P_H \\ P_L \\ T \end{bmatrix}_{t-2}$, and the elements of $[A]_{3\times3}$ and $[B]_{3\times3}$ are 3×3 matrices of constant coefficients to preserve the cross-correlation between P_H, P_L and T at time t and their auto-correlations.

Due to problems with inverting matrices that are not positive semi-definite and only the low frequency precipitation is AR(2), a contemporaneous form of the model in Equations (12) and (13) is used (Hipel and McLeod, 1994).

$$[A]_{3\times3} = \begin{bmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{bmatrix} \quad (12)$$

$$[B]_{3\times3} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & b_{22} & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (13)$$

where a_{11} is the AR(1) parameter of the high frequency precipitation IMFs, a_{22}, b_{22} are the AR(2) parameters of the low frequency precipitation IMFs, and a_{33} is the AR(1) parameter of the temperature IMFs.

Matrix $[C]$ is determined from the following equation (Salas et al., 1980):

$$[C][C]^T = [M_0] - [A][M_1]^T - [B][M_2]^T \quad (14)$$

where $[M_0]$, $[M_1]$ and $[M_2]$ are the lag-zero, lag-one and lag-two correlation matrices respectively. The elements of $[M_0]$ and $[M_1]$ corresponding to variables i and j are given by Equations (5) and (6) respectively. The elements of $[M_2]$ are given by:

$$m_2^{ij} = \frac{1}{n-2}\sum_{k=1}^{n-2} X_k^i X_{k-2}^j \quad (15)$$

and the elements of matrices $[A]$ and $[B]$ in Eq. (14) are given by:

$$a_{11} = m_1^{11} \tag{16}$$

$$a_{22} = \frac{m_1^{22}\left(1 - m_2^{22}\right)}{\left[1 - \left(m_1^{22}\right)^2\right]} \tag{17}$$

$$a_{33} = m_1^{33} \tag{18}$$

$$b_{22} = \frac{\left[m_2^{22} - \left(m_1^{22}\right)^2\right]}{\left[1 - \left(m_1^{22}\right)^2\right]} \tag{19}$$

For catchments with a low frequency precipitation component, matrices $[M_0]$, $[M_1]$ and $[M_2]$ are calculated using Equations (5), (6), and (15). The elements of the matrices $[A]$ and $[B]$ are calculated using Equations (16) to (19) and those of matrix $[C]$ are calculated from Eq. (14). One hundred replicates of standardised variates are generated from Eq. (11) then the skewness is incorporated using Eq. (2). The mean and standard deviation are reintroduced and, finally, the trends added to obtain the 100 stochastic replications of monthly precipitation and temperature.

As mentioned above, to ensure the generated monthly precipitation data preserved the annual characteristics, the generated monthly precipitation data were nested in an annual AR(1) model.

$$\frac{Y_i - \mu}{\sigma} = \rho \frac{Y_{i-1} - \mu}{\sigma} + \left(1 - \rho^2\right)^{0.5} \frac{\tilde{Y}_t - \mu_g}{\sigma_g} \tag{20}$$

where Y_i is the adjusted annual precipitation for year i, μ the mean of the observed annual data, σ the standard deviation of the observed annual data, ρ the lag one autocorrelation coefficient of the observed annual data, $\tilde{Y}_t$ the annual precipitation for year i obtained by aggregating the generated monthly data, μ_g the mean of the generated annual data, and σ_g the standard deviation of the generated annual data. The generated monthly data are then multiplied by the ratio $\frac{Y_i}{\tilde{Y}_t}$.

The stochastic model was tested by applying the above procedure to monthly precipitation and temperature data for 20CM3 from the MIROCM GCM after the data were subjected to EEMD analysis. Table 1 summarises the performance of the stochastic procedure to replicate

annual data for six catchments covering a range of climate types world-wide. Five of the catchments were modelled by an AR(1) process whereas station 6304 required an AR(2) model because it exhibited a low frequency precipitation component. Table 2 summarises the performance of the stochastic procedure to replicate monthly data for station 6304. Overall, the stochastic model performed satisfactorily at the monthly and annual time scales. As a general rule one would expect the value of the input parameters (GCM in this study) to be within ± 2x standard deviation (SD) of the mean of the generated values. This is achieved for all variables and catchments except in two cases: (1) annual lag-1 auto-correlation in catchment 6304 where the low frequency precipitation IMFs have high auto-correlation, which we assume bias the standardised variates and, therefore, the generated series; and (2) the standard deviation of January precipitation in catchment 6304. There was some variation between the generated and historical coefficient of skewness (results not shown) but in terms of the level of modelling required for this project, these differences are acceptable. The monthly precipitation and temperature values were satisfactorily generated as represented by station 6304 in Table 2, which was the most difficult catchment to replicate due to the high and low frequency precipitation components. We conclude from this assessment that the AR(1) and AR(2) stochastic models are able to preserve the monthly and annual precipitation and temperature characteristics satisfactorily for the purposes of this study.

2.4 Quantile-quantile bias correction of P and T

Prior to using GCM, or stochastic replicates of GCM, data in a climate change impact assessment, any bias between GCM and observed conditions needs to be corrected. The extent of bias in the GCM precipitation and temperature data is reported in the complementary paper (McMahon et al., 2014). For example, the mean annual precipitation (MAP) data for MIUB(3) compared with CRU MAP data at the GCM grid scale exhibits a slope of 0.69 on logarithmic scales, which indicates the GCM over-estimates low MAP and under-estimates high MAP. Mean annual temperatures are much less biased and require only a small amount of bias correction.

Ehret et al. (2012) presents a detailed review of bias correction and discusses the associated assumptions and implications of applying bias correction to GCM or regional climate model data. Many procedures are available for bias correction, with techniques falling into two categories: dynamical downscaling and statistical downscaling. Dynamical downscaling procedures are sophisticated and resource intensive (Tisseuil et al., 2010) and are impractical

for applying to globally distributed catchments and a range of GCMs as proposed in this study. In keeping with the proof-of-concept nature of this paper we adopt a simple empirical-statistical downscaling and error correction approach that is appropriate for bias correcting catchment average monthly (not daily) GCM outputs for input into a lumped (not spatially distributed) hydrologic model. We did not adopt the delta change method, also known as simply daily scaling (Chiew, 2010), where the observed series is scaled by the relative difference between future and baseline conditions, as delta change would not make full use of the re-ordering of precipitation and temperature events provided by the stochastic replicates. Rather, we adopted quantile-quantile or quantile mapping as discussed in Themeβl et al., (2012) and Bárdossy and Pegram (2011). The basis of the quantile-quantile bias correction is a comparison of the empirical cumulative distribution functions (ECDF) of the observed data and the GCM data for a common period. Here the common period is the observed catchment record and the concurrent period of GCM data from the 20C3M scenario. The difference between observed and GCM ECDFs for a given value provides the bias correction. Here we also adopt the frequency adaptation method discussed in Themeβl et al., (2012) for when the GCM series has a higher frequency of zero values than the observed series. The issue of new extremes, values outside the range of the GCM and observed data during the period in which the bias correction is established, was also investigated by Themeβl et al., (2012). We adopt option QMv1a of Themeβl et al., (2012), which takes the bias correction at the highest (lowest) quantile and applies that correction to all new upper (lower) extremes. In our analysis we establish and apply a bias correction for each calendar month (12 corrections in all), rather than a single correction for the whole of record at each catchment.

An assumption of using this bias correction is that the correction applies into the future under different conditions. This assumption is supported by Teutschbein and Seibert (2013) who found the quantile-quantile method performed best out of six alternate bias corrections in differential split sample tests for non-stationary conditions. The quantile-quantile bias correction is applied to the precipitation and temperature stochastic replicates (100 replicates of ~250 years covering the 20CM3 and A1B scenarios) for each GCM run (5) at 17 global catchments. These 8,500 (100 x 5 x 17) sets of monthly precipitation and temperature are used as input to our hydrologic model PERM to estimate monthly runoff.

2.5 Monthly Precipitation-Evapotranspiration-Runoff Model (PERM)

In order to convert GCM monthly precipitation and temperature into runoff, the PERM model was developed specifically to meet the requirements for hydrologic modelling in this project. PERM is a simple lumped, not spatially distributed, conceptual precipitation-runoff model run on a monthly time-step with five parameters to be optimised. The time-step was dictated by the availability of monthly streamflow data and concurrent precipitation and temperature data. Further details about the precipitation-evapotranspiration-runoff model, source code and example input and output are provided in the Supplementary Material Section.

2.5.1 Model structure

The structure of PERM is shown in Figure 2 with the parameters to be calibrated highlighted in bold. As observed in Figure 2 monthly precipitation is either added to the interception store (if the monthly mean daily temperature is > 0°C) or accumulated in a snow pack (if the monthly mean daily temperature is ≤ 0°C). The contents of the interception store are reduced by evaporation. Excess precipitation is subject to an infiltration function in which a surface runoff component (designated as *PAreaF*) is dependent on the contents of the soil moisture store. When precipitation is accumulated as snow, there is no evaporation for that month from the snow pack or evapotranspiration from the soil moisture store. The snow pack continues to accumulate as long as the monthly mean daily air temperature is ≤ 0°C. The snow pack begins to melt when the monthly mean daily temperature is > 0°C. Snowmelt is partitioned into two components, a runoff and soil moisture infiltration component based on the parameter *Melt*. When the maximum capacity of the soil moisture store is exceeded, saturation excess runoff (*SMF*) occurs. Evapotranspiration from the soil moisture store is estimated either as a linear function of the available soil moisture or as a linear function of monthly mean daily temperature. The algorithms representing these soil moisture or energy limiting conditions are given in Figure 2. Baseflow from the soil moisture store is simulated as a linear recession of the water content in the store.

Two other hydrologic processes – impervious area runoff and deep recharge – were considered for inclusion in PERM. The inclusion of impervious area was considered unnecessary. With respect to deep seepage, the reviews of Petheram et al. (2002), Scanlon (2006) and Crosbie et al. (2010) suggest the maximum effect could be, on average, equivalent to 5% of the long-term average annual precipitation. From the results of these reviews and

taking into account the model time-step, the available data, and the fact that the parameters in PERM are calibrated, it was concluded that incorporating deep seepage would yield little benefit to the modelling exercise.

2.5.2 Model calibration

PERM is run on a monthly time-step and calibrated against observed annual runoff. Details of the calibration are set out in the Supplementary Material Section. In summary, an objective function, defined as the sum of squared differences between the estimated and observed annual runoff, was minimised with penalties applied to the objective function to ensure the calibrated model approximately reproduced the mean and coefficient of variation of observed annual runoff. An automatic pattern search optimisation method was used to calibrate the model (Hooke and Jeeves, 1961; Monro, 1971) with 10 different parameter sets used as starting points to increase the likelihood of finding the global optimum of parameter values. A K-fold cross-validation method (where K = 3) described by Efron and Tibshirani (1993) was used to validate the calibrated model. PERM was calibrated for 699 catchments world-wide using the observed monthly precipitation, temperature and runoff data described in Peel et al. (2010).

2.5.3 Model performance and catchment selection

An objective of our study is to examine the within-GCM uncertainty in runoff estimated from GCM projections of precipitation and temperature. To examine this uncertainty we need to minimise any uncertainty in future runoff due to poor hydrologic model calibration. Therefore, a sub-set of the 699 catchments was selected for further analysis that exhibited minimum error as a result of the calibration process. Several criteria were used to assess the adequacy of the PERM calibration for selecting the catchments. These criteria included: the annual Nash-Sutcliffe efficiency (NSE) (Nash and Sutcliffe, 1970) between observed and modelled runoffs was > 0.8, NSE values based on 3-fold independent testing were > 0.60, monthly NSE values were > 0.6, and the mean and the coefficient of variation of annual runoff were estimated to be within ± 5% and ± 10% respectively of the observed values. From a practical point of view, catchments less than 1,000 km^2 were excluded as were several that were spatially very close. Using these selection criteria, 17 catchments from the initial 699 data set were selected for later analyses. Figure 3 shows the location of all 699 catchments

and the sub-set of 17 catchments used in later analyses, while Table 3 provides information about the selected catchments.

Details of the modelling performance of PERM for the 699 catchments are presented in the Supplementary Material Section. For the 17 selected catchments the difference between the average modelled and observed MAR is -0.2% and coefficient of variation of annual runoff is -4.4%. At the annual time-step, the average NSE and R^2 between modelled and observed runoff are both 0.88, and the monthly NSE is 0.72. PERM is well calibrated for these 17 catchments and, therefore, uncertainty in runoff due to poor model calibration is minimised using these catchments.

A key assumption of using PERM, or any hydrologic model, to model future runoff is that the calibrated parameters are appropriate for the future climatic conditions. Where future climatic conditions are similar to the observed calibration period, then this assumption is likely to hold. If climatic conditions differ from the calibration period, then there is no evidence to support this assumption. However, in terms of the analysis conducted in the next section this assumption is a pragmatic one that may well affect the bias of future runoffs but should have less impact on the range of uncertainty.

2.6 Uncertainty in reservoir yield

The 17 catchments modelled by PERM are unregulated catchments and do not have an existing reservoir on which to base our analysis. Therefore, we need to assume a hypothetical reservoir for each catchment. Many procedures exist to estimate reservoir yield from a hypothetical storage (see McMahon and Adeloye, 2005). For the purposes of this analysis we require a method that is simple to apply as there are 100 replicates of future runoff generated by PERM from 5 GCMs for the 17 selected catchments. Here we adopt the Gould-Dincer Gamma (G-DG) procedure for estimating reservoir yield, which is defined as (McMahon and Adeloye, 2005; Petheram et al., 2008):

$$D = \mu - \frac{\sigma^2}{\mu\tau\gamma^2}\left(\frac{1-\rho^3}{\left(1-\rho^2\right)^{1.5}}\right)^{-2}\left[\left\{1+\frac{\gamma z_p}{6}\left(\frac{1-\rho^3}{\left(1-\rho^2\right)^{1.5}}\right)-\frac{\gamma^2}{36}\left(\frac{1-\rho^3}{\left(1-\rho^2\right)^{1.5}}\right)^2\right\}^3-1\right]^2\left(\frac{1+\rho}{1-\rho}\right) \qquad (21)$$

where D is the annual yield or draft (mm year^{-1}), μ and σ are the mean and the standard deviation, respectively, of annual runoff (mm year^{-1}) into the reservoir storage, τ is the

hypothetical storage capacity specified as a ratio of mean annual runoff, γ is the coefficient of skewness of annual runoff, ρ is the lag-1 serial correlation of annual runoff, and z_p is the standardized normal variate (zero mean and unit variance) at p the probability of failure. In our analysis, we adopted 95% reliability of supply, thus $(1-p)$ = 0.05. In our analysis we specify two hypothetical storage sizes, τ = 1 and τ = 3, which are storage capacities equal to one and three times the mean annual runoff. The G-DG procedure assumes there is no net evaporation loss from the storage (see McMahon et al., (2007) for a detailed description of the procedure), which in this analysis is not considered critical as we are mainly considering relative changes in yield. In this analysis the uncertainty in reservoir yield is estimated via the G-DG procedure using annual runoff parameters (μ, σ, γ, and ρ) estimated from the monthly runoff time-series generated by PERM (see steps 7 and 8 of Figure 1).

3 Testing the stochastic within-GCM uncertainty approximation

Testing our stochastic approximation of within-GCM uncertainty requires multiple runs from a single GCM for a given scenario from which to estimate within-GCM uncertainty and compare against our stochastic results. In the CMIP3 data set the CCSM GCM has the most runs (seven) for the 20C3M and A1B scenarios. In this section we test the ability of our stochastic methodology to approximate within-GCM uncertainty for the CCSM GCM using the seven available runs for the period 1870-2100 (20C3M and A1B emissions scenarios).

A comparison of within-GCM uncertainty based on seven runs from the CCSM GCM and the stochastic approximation of within-GCM uncertainty for (a) annual precipitation and (b) annual temperature for the Herbert River at Gleneagle is shown in Figure 4. The CCSM runs and stochastic replicates presented in Figure 4 are not bias corrected. In each plot the maximum, median and minimum annual value for a given year are shown for the seven CCSM runs and are compared with the maximum, median and minimum of the 700 (7 x 100) stochastic replicates of the CCSM runs for annual precipitation and temperature. For both precipitation and temperature the median of the 700 stochastic replicates overlies the median of the 7 CCSM runs and the difference between the maximum and minimum lines around the median for the two datasets is totally consistent given only seven CCSM runs and 700 stochastic replicates. A comparison of the standard deviation of all annual values calculated for the seven CCSM runs (precipitation = 110 mm, temperature = 1.21 ^{o}C) and the 700

stochastic replicates (precipitation = 111 mm, temperature = 1.20 ^{o}C) confirms the stochastic replicates are replicating the CCSM GCM runs well in terms of overall trend and variability around the trend. These results confirm the credibility of the stochastic methodology for approximating the within-GCM uncertainty when limited GCM runs are available.

4 Results and Discussion

In this section we present and discuss results from the methodology described in the previous section to approximate within-GCM uncertainty of precipitation and temperature from five GCMs and assess the consequent impact of these uncertainties on estimated runoff and reservoir yield at 17 catchments for two 30-year periods – 1965-1994 (20C3M emissions scenario) and 2015-2044 (A1B emissions scenario).

To assist interpretation of within- and between-GCM uncertainty results for 17 catchments and 5 GCMs presented in later tables and figures we present results for an example catchment, the Herbert River at Gleneagle in Australia, in Figure 5. Box-plots of MAP (Figure 5a) and MAT (Figure 5b) are presented for two 30-year periods for each GCM. These box-plots represent our approximation of within-GCM uncertainty of MAP and MAT. The box represents the inter-quartile range of MAP (MAT) from the 100 bias-corrected stochastic replicates of GCM precipitation (temperature). The median MAP (MAT) is represented by the bar across the box and the box-plot whiskers represent the maximum and minimum MAP (MAT) from the 100 replicates. The range of within-GCM uncertainty of MAP (Figure 5a) is similar for all GCMs except MIROCM(1), where the inter-quartile and maximum/minimum range are approximately 50% larger. The range of within-GCM uncertainty of MAT (Figure 5b) is similar for all GCMs.

The box-plots in Figure 5 can also be used to assess between-GCM uncertainty through differences between GCMs in the range of within-GCM uncertainty and differences in the direction of change between 30-year period box-plots. All GCMs have an increasing trend in MAP over time for this catchment except HadCM3 (Figure 5a), whereas all GCMs show a similar increasing trend in MAT over time (Figure 5b).

Also shown in Figure 5 is a ‘Raw’ symbol plotted next to each box-plot. These MAP and MAT values are calculated from bias-corrected original CMIP3 GCM runs and are the only values of MAP and MAT available for this combination of catchment, GCM and scenario if

stochastic replication is not used. In a traditional climate change impact assessment, without stochastic replication, the 'Raw' values are all that are available for analysis and the magnitude of uncertainty associated with them is unknown. Figure 5 shows that the range of within-GCM uncertainty associated with 'Raw' values of MAT is smaller than for MAP.

Figure 5 can also be used to check whether our stochastic methodology is performing well at this catchment. Our stochastic methodology generates statistically similar replicates of each 20C3M and A1B GCM run from which we calculate MAP and MAT over two 30-year periods to obtain our box plots. If our methodology is performing well we would expect the 'Raw' values from the original GCM runs to fall within our box-plot range, which they do in all cases. In should be noted that the true within-GCM uncertainty range for MAP and MAT will be larger than what is shown by our box-plots since we have only replicated the uncertainty around the GCM trend and not the uncertainty in the trend itself.

In Table 4 within-GCM uncertainty results are presented for the five GCMs over the period 1965-1994 at the 17 catchments for six variables – MAP, standard deviation of annual precipitation (SDP), MAT, MAR, SDR and lag-1. Here the six variables have been calculated for each stochastic replicate and the results are presented as the mean ± the standard deviation of the 100 replicate estimates. The mean values in Table 4 show the range of hydroclimatic conditions represented by the 17 catchments, while the standard deviation around each mean represents our approximation of within-GCM uncertainty of that variable. The mean values differ between the five GCMs for a given catchment for at least two reasons: (1) the stochastic variability in the mean value of a sample of 100 replicates; and (2) each GCM has a different trend for each catchment during this period.

Although Table 4 provides absolute values of within-GCM uncertainty for each combination of catchment, GCM and variable, it is difficult to draw conclusions from this table. Therefore, in Tables 5, 6 and 8 we express within-GCM uncertainty in relative form as the standard deviation of the 100 replicate estimates as a percentage of the mean of the 100 replicate estimates. If the 100 replicate values are normally distributed, then approximately 95% of the values will be within ± 2 standard deviations of the mean. The assumption of normality was tested for six variables – MAP, SDP, MAT, MAR, SDR and reservoir yield for a reservoir equal to 3x MAR – estimated from 100 replicates of HadCM3 at each of the 17 catchments for the period 1965-1994. Normality was assessed using the Anderson-Darling normality test (Anderson and Darling, 1954). Of the 102 normality tests (17 catchments x 6 variables) 11

(10.8%) were not normally distributed at the 5% level of significance (not shown), which is more than expected from random chance. The distribution of non-normal results was MAP(2), SDP(0), MAT(1), MAR(4), SDR(2) and reservoir yield(2). Based on our analysis of HadCM3 replicates an expectation that roughly 95% of MAP, SDP and MAT values will be within ± 2 standard deviations of the mean appears reasonable. Whereas for MAR, SDR and reservoir yield this expectation is less justified and within-GCM uncertainty is less likely to be symmetrically distributed around the mean.

4.1 Annual precipitation and temperature

In this sub-section we present and discuss within- and between-GCM uncertainty results for annual precipitation (MAP and SDP) and temperature (MAT). In Table 5 a summary is presented of the within-GCM uncertainty results shown in Table 4. The uncertainty results in Table 5 are in relative form (standard deviation as a percentage of the mean), except for lag-1 where the standard deviation is used, and are the average uncertainty across the 17 catchments for each GCM. Averaging relative uncertainty values across the catchments allows differences in within-GCM uncertainty between GCMs to be examined and the average uncertainty across all GCMs for a given variable of interest to be estimated. For MAP within-GCM uncertainty varies between GCMs from 3.4% to 4.6% and the average across the five GCMs is 4.1%. Given a normal distribution of MAP values across the 100 replicates this translates into 95% of MAP values being within ±7% to ±9% of the replicate mean. Although the average within-GCM uncertainty for SDP (14.3%) is 3 – 4 times higher than for MAP (4.1%), the difference between GCMs is narrower for SDP (13.9% – 14.6%). Within-GCM uncertainty of inter-annual variability of precipitation is high with approximately 95% of SDP values being within ±28% to ±29% of the mean SDP. In contrast, within-GCM uncertainty for MAT is very small (1%, 95% within ±2% of mean MAT) and is very consistent between GCMs. Across the five GCMs MIUB(1) has the least within-GCM uncertainty for MAP and SDP and the highest for MAT, while MIROCM(1) has the highest for MAP and SDP.

A similar set of results to those in Table 5 are presented in Table 6 for the 30-year period 2015-2044 (A1B). On average across the five GCMs within-GCM uncertainty of MAP, SDP and MAT are similar between the two periods. However, differences in uncertainty between GCMs are apparent across the two periods. For example, for MAP the maximum and minimum values in Tables 5 and 6 are from MIROCM(1) (4.6% and 4.9%) and MIUB(1) (3.4% and 3.4%) respectively. It should be noted that the uncertainties in Table 6 are for the

projected values of precipitation, temperature and runoff in 2015-2044, and not the uncertainties in their changes between the earlier and the later periods. Often, regional climate change projections present uncertainties in the projected changes in variables. However, here we present the uncertainties in the projected values, as these are important for the projected runoff and reservoir yield.

Within-GCM uncertainty results for MAP, SDP and MAT from Table 4 are also summarised in Figures 6 to 8 where results from the 17 catchments are plotted for each GCM. Figure 6 shows within-GCM uncertainty in MAP varies from below 20 mm year^{-1} to more than 90 mm year^{-1}. For each GCM there is a weak positive relationship between uncertainty and MAP. The strongest relationship is for MIUB(1), which also has the lowest uncertainty. In Figure 7, within-GCM uncertainty results for SDP are in contrast to the MAP results of Figure 6. Although the range in uncertainty of SDP is only slightly less than for MAP, ranging from approximately 10 mm year^{-1} to above 70 mm year^{-1}, there is a very strong positive relationship between uncertainty in SDP and SDP for each GCM shown. The reason for the difference in relationship strengths in Figures 6 and 7 is due to the uncertainty in MAP being strongly positively related to SDP (not shown). This is to be expected as higher inter-annual variability, represented here by SDP, increases the uncertainty in a 30-year estimate of MAP. Thus in Figure 6 the weak relationship between MAP and the uncertainty in MAP is due to a combination of stochastic variability and the SDP at each catchment. Whereas the strong relationship observed in Figure 7 between SDP and uncertainty in SDP is solely a function of stochastic variability. Within-GCM uncertainty in mean annual temperature is small, varying from less than 0.05°C to 0.18°C, and decreases with increasing MAT for all GCMs, although the relationships between uncertainty and mean annual temperature are weak (Figure 8).

4.2 Annual runoff

In this sub-section we present and discuss within- and between-GCM uncertainty results for annual runoff (MAR, SDR and lag-1). For each GCM and catchment 100 bias-corrected stochastic replicates of monthly P and T were used as input to PERM and 100 series of monthly runoff generated. In Figure 9 box-plots of MAR (Figure 9a) and SDR (Figure 9b) are presented for the Herbert River at Gleneagle in Australia as an example of the runoff results for 17 catchments and 5 GCMs. For each GCM a box-plot of the 100 values calculated from PERM is presented for the two 30-year periods. Consistent with results for MAP shown in Figure 5a the range of within-GCM uncertainty of MAR is similar for all GCMs except

MIROCM(1), where the inter-quartile and maximum/minimum range are approximately 50% larger. For MAP all GCMs, except HadCM3, had an increasing trend in MAP over time (Figure 5a). However, trends in MAP are not consistently transferred to MAR. As expected HadCM3 has a decreasing trend in MAR and MIROCM(1) shows an increasing trend in MAR. However, MIUB(1), MPI(1) and MRI(3) show little change in MAR between periods. In Figure 9b the range of within-GCM uncertainty of SDR is similar across all the GCMs, except MIROCM(1) which again has the highest inter-quartile and maximum/minimum range. The pattern of trend in SDR between the two periods is complex. MIROCM(1) is the only GCM to have an increase in median SDR and inter-quartile range over time, while HadCM3, MIROCM(1) and MRI(3) all have an increase in maximum/minimum range over time. Also shown in Figure 9 are the 'Raw' values of MAR and SDR calculated from PERM runs of bias-corrected original GCM data, which are the only values of MAR and SDR available for this combination of catchment, GCM and scenario if stochastic replication is not used. As was the case for MAP and MAT, the 'Raw' MAR and SDR values for each GCM and 30-year period fall within the box-plot range, indicating that our stochastic replication methodology is performing satisfactorily. Again, the true within-GCM uncertainty range for MAR and SDR is expected to be larger than what is shown by our box-plots since we only replicated the noise around the GCM trend and not the GCM trend itself.

Within-GCM uncertainty results for MAR and SDR averaged across the 17 catchments for each of the 5 GCMs and expressed in relative form (standard deviation as a percentage of the mean) are shown in Table 5 for the 30-year period 1965-1994 (20C3M). Also shown in Table 5 is the standard deviation and the lag-1 serial correlation of runoff. In Table 5 within-GCM uncertainty of MAR varies between GCMs from 8.1% to 10.9% and the average across the GCMs is 9.7%. Although the 100 MAR values may not be normally distributed we would expect roughly 95% of the MAR values to be within ±16% to ±22% of the replicate mean MAR. The average within-GCM uncertainty of MAR (9.7%) is over double that of MAP (4.1%), which demonstrates how uncertainty in precipitation is magnified in runoff through the precipitation-runoff relationship. In Table 5 the average within-GCM uncertainty for SDR (17.4%) is approximately 80% higher than for MAR (9.7%) and the difference between GCMs is narrower for SDR (16.4% – 18.4%) than for MAR. Within-GCM uncertainty of inter-annual variability of runoff is high with approximately 95% of SDR values being within ±33% to ±37% of the mean SDR. Although the within-GCM uncertainty of SDR is high (17.4%), it is only ~20% higher than the uncertainty for SDP (14.3%). In Table 5 there is little

difference between GCMs in within-GCM uncertainty of lag-1 serial correlation with standard deviation values varying from 0.17 to 0.18 and the average across the GCMs is 0.18.

Within-GCM uncertainty of MAR, SDR and lag-1 serial correlation for the 30 year period 2015-2044 (A1B) in Table 6 is similar to the results shown in Table 5. However, differences in uncertainty between GCMs are apparent across the two periods. For example, for MAR the minimum values in Tables 5 and 6 are from MIUB(1) (8.1% and 8.0%), whereas the second highest MAR in Table 5 and the maximum value in Table 6 are from HadCM3 (10.8% and 13.0%).

Within-GCM uncertainty results for MAR and SDR from the 17 catchments and 5 GCMs in Table 4 are summarised in Figures 10 and 11. Figure 10 shows the within-GCM uncertainty in MAR varies from below 10 mm year^{-1} to more than 80 mm year^{-1}. For each GCM there is a positive relationship between uncertainty and MAR. The strongest relationship is for MIUB(1), which also has the lowest uncertainty. In Figure 11, the within-GCM uncertainty results for SDR are in contrast to the MAR results of Figure 10. Although the range in uncertainty of SDR is slightly less than for MAR, ranging from under 10 mm year^{-1} to above 60 mm year^{-1}, there is a much stronger positive relationship between the uncertainty in SDR and SDR for each GCM shown. Although the uncertainty relationships for precipitation (Figures 6 and 7) and runoff (Figures 10 and 11) are broadly similar, modelling runoff through PERM has modified the uncertainty relationships of precipitation relative to runoff. The relationships in Figure 10 for MAR are stronger than those for MAP in Figure 6, while the relationships for SDR in Figure 11 are weaker than those for SDP in Figure 7. The within-GCM uncertainty of lag-1 serial correlation of annual runoff is approximately constant at 0.18 and shows little difference between GCMs (not shown).

4.3 Reservoir yield

In this sub-section we present and discuss within- and between-GCM uncertainty results for reservoir yield. The impact of within-GCM uncertainty on reservoir yield is shown in Figure 12 for the Herbert River at Gleneagle. The Gould-Dincer Gamma method was used to estimate average annual yield from a hypothetical reservoir of capacity equal to 3x MAR with 95% reliability of draft using annual runoff statistics from PERM modelled runoff. Each box-plot is based on 100 values of average annual reservoir yield for the two 30-year periods estimated from PERM runs using stochastic replicates of precipitation and temperature for

each GCM. In Figure 12 the minimum average annual yield is zero in two cases – HadCM3 (2015-2044) and MIROCM(1) (1965-1994). In these two cases the Gould-Dincer Gamma method returned a physically impossible negative draft, which indicates that a positive draft cannot be supplied with 95% reliability from the hypothetical reservoir for at least one replicate. For a given GCM an increasing trend in average reservoir yield between periods is due to either an increasing trend in MAR (Figure 9a) and or a decreasing trend in SDR (Figure 9b). The within-GCM uncertainty of average annual yield for MIROCM(1) is 70% larger than for the other GCMs for this catchment, which is consistent with the MAP, MAR and SDR results. The 'Raw' values of average reservoir yield again fall within the box-plot range, indicating that our stochastic replication methodology is performing satisfactorily.

Table 7 lists reservoir yield results for the 17 catchments, 5 GCMs and two hypothetical reservoir capacities (1x MAR and 3x MAR) for the 30-year period 2015-2044 (A1B). The average reservoir yield is the average of the 100 PERM runs and the uncertainty is the standard deviation of the 100 PERM runs. Differences between GCMs in average reservoir yield at a given catchment largely reflect differences in MAR and SDR trends during this period. If the Gould-Dincer Gamma method returned a physically impossible negative draft, the yield for that run was set to zero. In Table 7 if more than half the yields (>50) were set to zero then results for that GCM and catchment combination were not reported (N/R).

The yield uncertainties in Table 7 due to within-GCM uncertainty are expressed as a percentage of the mean yield and averaged across the 17 catchments in Table 8. For the larger storage, τ = 3, the average uncertainty across the 5 GCMs is 11.9%, which is approximately half the average uncertainty (25.1%) for the smaller storage (τ = 1). The GCM with the highest uncertainty in reservoir yield is MIROCM(1), which is consistent with the high uncertainty in MAP, MAR and SDR for this GCM. MIUB(1) is the only GCM to have below average uncertainty for both reservoir sizes. The uncertainty data in Table 7 for τ = 3 are plotted in Figure 13 against mean reservoir yield. Uncertainty of reservoir yield is only weakly positively related to reservoir yield. For a given reservoir yield MIROCM(1) and HadCM3 generally have higher uncertainty than the other GCMs. The range of uncertainty is approximately 15 mm year^{-1} for low yield reservoirs (100 mm year^{-1}) through 40 mm year^{-1} for reservoirs that yield 1,000 mm year^{-1}.

The main driver of uncertainty in reservoir yield is the variability of annual reservoir inflows. Figure 14 shows the relationship between uncertainty in reservoir yield, for τ = 3, against the

variability of reservoir inflows expressed as the standard deviation of annual runoff for the 5 GCMs over the period 2015-2044 (A1B). The relationship between uncertainty in reservoir yield and annual runoff variability is strongly positive with all GCMs having an $R^2 \geq 0.78$, except for MIROCM(1). Uncertainty in yield is driven more strongly by inflow variability than by inflow mean (not shown, but very similar to Figure 13 as reservoir yield and mean annual runoff are highly correlated).

5 Conclusions and Implications

Climate change impact assessments for future hydrology are subject to significant uncertainties. The contribution of within-GCM uncertainty to total uncertainty has not been well quantified due to the limited number of GCM runs available for each GCM and scenario combination. In this paper we developed a methodology to approximate within-GCM uncertainty of precipitation and temperature projections using non-stationary stochastic data generation. Our methodology is a contribution toward quantifying within-GCM uncertainty and provides an objective approach for communicating the uncertainty in climate change impact assessments in a quantitative manner. In a proof-of-concept application of our procedure we estimated the impact of within-GCM uncertainty on annual runoff and reservoir yield, which can inform water resources engineers and management decision makers about the uncertainty in climate change impacts in the short to medium term planning horizon. For the research community our stochastic data generation methodology provides a way to assess within-GCM uncertainty on a temporary basis until the number of GCM runs for a given GCM and scenario combination becomes adequate to estimate within-GCM uncertainty directly from GCM runs.

In our proof-of-concept application we de-trended GCM projections of monthly precipitation and temperature from five better performing CMIP3 GCMs (HadCM3, MIROCM, MIUB MPI and MRI) identified in the companion paper (McMahon et al., 2014). We stochastically replicated the de-trended series 100 times, combined the replicates with their respective trends and applied a bias-correction to the replicates. Within-GCM uncertainty of precipitation and temperature were assessed using the stochastic replicates from each GCM for two periods: (1) 1965-1994 (20C3M scenario); and (2) 2015-2044 (A1B scenario) at 17 catchments distributed around the world. At each catchment within-GCM uncertainty was estimated as the standard deviation of the replicate values divided by the mean replicate value. The uncertainty value

for a given GCM was taken as the average of the 17 catchment values for that GCM. Within-GCM uncertainty of mean annual precipitation varied from 3.4% – 4.9% between GCMs over the two periods and averaged approximately 4.1% across the five GCMs. For the standard deviation of annual precipitation the average within-GCM uncertainty (14.3%) was 3 – 4 times larger than for mean annual precipitation, while within-GCM uncertainty of mean annual temperature was smaller (1%).

The stochastic replicates were input to a calibrated hydrologic model (PERM) to estimate future projections of annual runoff. The impact of within-GCM uncertainty on mean annual runoff varied from 8.0% – 13.0% between GCMs over the two periods and averaged approximately 9.9% across the five GCMs. The uncertainty in the standard deviation of annual runoff varied from 16.0% – 20.1% between GCMs and averaged approximately 17.5% across the five GCMs. The within-GCM uncertainty in precipitation and temperature is amplified in the runoff through the precipitation-runoff relationship. Summary statistics for the two periods were estimated from each annual runoff series (100 per catchment) and used in the Gould-Dincer Gamma method to estimate reservoir yield from two hypothetical reservoir capacities (1x and 3x mean annual runoff) for 95% reliability of supply. For the period 2015-2044 the uncertainty in reservoir yield due to within-GCM uncertainty varied from 18.6% (9.3%) to 33.2% (14.3%) for the 1x (3x) mean annual runoff capacity reservoir and averaged approximately 25% (12%) across the 5 GCMs. The main driver of uncertainty in reservoir yield was the variability of annual runoff inflows.

In this analysis between-GCM uncertainty was limited to small differences in within-GCM uncertainty for a given variable and differences in trend between the two 30-year periods analysed. The reason why differences between GCMs are not larger here is due to the application of bias correction. The quantile-quantile bias correction forces the mean and variance of the GCM precipitation and temperature data over the observed period to match the observed mean and variance. Thus, a significant source of between-GCM uncertainty, their bias in mean and variance, has been removed.

A significant implication of our results is that within-GCM uncertainty is important when interpreting climate change impact assessments. Although the variables calculated from the stochastic replicates and hydrologic modelling of the replicates are not strictly normally distributed, a rough guide to the magnitude of within-GCM uncertainty is to double the values reported above (±2x standard deviation / mean). Thus for the five GCMs analysed here during

the period 2015-2044 the within-GCM uncertainty around a value of mean annual precipitation is approximately ±7% to ±10%. For the standard deviation of annual precipitation the uncertainty is approximately ±27% to ±29%, while for mean annual temperature the uncertainty is approximately ±1.2% to ±1.6%. Compared to precipitation, runoff uncertainties are larger with approximate uncertainties of mean and standard deviation of annual runoff being ±16% to ±26% and ±32% to ±40% respectively. The uncertainty around reservoir yield for a 1x mean annual runoff reservoir is approximately ±37% to ±66% and ±18% to ±28% for a 3x mean annual runoff reservoir.

The amplification of precipitation and temperature within-GCM uncertainty in runoff and reservoir yield has significant implications for interpreting climate change impact assessments of these variables. For example, in Figures 5 (MAP, MAT), 9 (MAR, SDR) and 12 (3xMAR yield) we presented within-GCM uncertainty box-plots for two periods for each GCM and the 'Raw' value associated with each box-plot that would be the only value available in a traditional climate change impact assessment for the Herbert River at Gleneagle. For MAT (Figure 5b), low within-GCM uncertainty allows conclusions based on analysis of the 'Raw' values to be consistent with conclusions drawn from the box-plots – the 'Raw' values indicate MAT is projected to increase by ~5% between the two periods and the box-plots don't overlap between the two periods for any of the GCMs. However, for variables where within-GCM uncertainty is higher, conclusions drawn from a traditional climate change impact assessment would be misleading. For example, in Figure 5a MIUB(1) shows the largest increase in 'Raw' MAP between the two periods (14.7%), yet the box-plots for this GCM clearly overlap and the increase in median MAP is only 4.3%. Similarly in Figure 9a MIUB(1) shows the largest increase in 'Raw' MAR between the two periods (34.1%), yet the box-plots for this GCM overlap and the increase in median MAR is only 0.3%. Finally, in Figure 12 MIUB(1) shows the largest increase in 'Raw' reservoir yield between the two periods (41.3%), yet the box-plots for this GCM overlap and the increase in median yield is only 0.6%. A traditional, without stochastic replication, climate change impact assessment reporting future increases in MAP, MAR and reservoir yield of 14.7%, 34.1% and 41.3% respectively would initially seem significant, yet our approximation of within-GCM uncertainty suggests these increases could be solely due to within-GCM uncertainty.

Finally, we expect our results are an under-estimate of the true within-GCM uncertainty due to our stochastic method only approximating the uncertainty around the overall GCM trend

and not the uncertainty in the GCM trend itself. To obtain a true estimate of within-GCM uncertainty requires analysis of many (≥100) GCM runs of a given scenario. Until considerably more GCM runs of a scenario become available the methodology presented here provides an interim objective technique for estimating the influence of within-GCM uncertainty on climate change impact assessments that is suitable for existing, or future, GCM scenario runs. Climate change impact assessments based on projections that do not take into account within-GCM uncertainty risk providing water resources management decision makers with a sense of certainty that is unjustified.

Acknowledgements

This research was financially supported by Australian Research Council Grant LP100100756 and FT120100130, Melbourne Water and the Australian Bureau of Meteorology. Lionel Siriwardena, Sugata Narsey and Dr Ian Smith assisted with extraction and analysis of CMIP3 GCM data. We acknowledge the modeling groups, the Program for Climate Model Diagnosis and Intercomparison (PCMDI) and the WCRP's Working Group on Coupled Modelling (WGCM) for their roles in making available the WCRP CMIP3 multi-model dataset. Support of this dataset is provided by the Office of Science, U.S. Department of Energy. We also acknowledge the contribution of two anonymous reviewers whose comments and suggestions improved the paper.

References

Anderson, T. W., and Darling, D. A.: A test of goodness of fit, J. Am. Stat. Assoc., 49, 765-769, 1954.

Andréassian, V., Perrin, C., and Michel, C.: Impact of imperfect potential evapotranspiration knowledge on the efficiency and parameters of watershed models, J. Hydrol., 286, 19-35, 2004.

Andréassian, V., Perrin, C., Berthet, L., Le Moine, N., Lerat, J., Loumagne, C., Oudin, L., Mathevet, T., Ramos, M.-H., and Valéry, A.: HESS Opinions "Crash tests for a standardized evaluation of hydrological models", Hydrol. Earth Syst. Sc., 13, 1757-1764, 2009.

Arora, V.: Streamflow simulations for continental-scale river basins in a global atmospheric general circulation model. Adv. Water Resourc., 24, 775-791, 2001.

Bárdossy, A., and Pegram, G.: Downscaling precipitation using regional climate models and circulation patterns toward hydrology, Water Resour. Res., 47, W04505, doi:10.1029/2010WR009689, 2011.

Bosshard, T., Carambia, M., Goergen, K., Kotlarski, S., Krahe, P., Zappa, M., and Schär, C.: Quantifying uncertainty sources in an ensemble of hydrological climate-impact projections, Water Resour. Res., 49, 1523–1536, doi:10.1029/2011WR011533, 2013.

Chen, J., Brissette, F. P., Poulin, A., and Leconte, R.: Overall uncertainty study of the hydrological impacts of climate change for a Canadian watershed, Water Resour. Res., 47, W12509, doi:10.1029/2011WR010602, 2011.

Chiew, F. H. S.: Lumped conceptual rainfall-runoff models and simple water balance methods: Overview and applications in ungauged and data limited regions, Geography Compass, 4/3, 206-225, 2010.

Chiew, F. H. S., Potter, N. J., Vaze, J., Petheram, C., Zhang, L., Teng, J., and Post, D. A.: Observed hydrologic non-stationarity in far south-eastern Australia: implications for modelling and prediction, Stoch. Environ. Res. Risk. Assess., 28, 3-15, 2014.

Crosbie, R. S., Jolly, I. D., Leaney, F. W., and Petheram, C.: Can the dataset of field based recharge estimates in Australia be used to predict recharge in data-poor areas? Hydrol. Earth Syst. Sc., 14, 2023-2038, 2010.

Deser, C., Phillips, A., Bourdette, V., and Teng, H.: Uncertainty in climate change projections: the role of internal variability, Clim. Dynam., 38, 527–546, 2012.

Deser, C., Phillips, A. S., Alexander, M. A., and Smoliak, B. V.: Projecting North American climate over the next 50 years: Uncertainty due to internal variability, J. Climate, 27, 2271-2296, 2014.

Di Baldassarre, G., and Montanari, A.: Uncertainty in river discharge observations: A quantitative analysis, Hydrol. Earth Syst. Sc., 13, 913-921, 2009.

Dobler, C., Hagemann, S., Wilby, R. L., and Stötter, J.: Quantifying different sources of uncertainty in hydrological projections in an Alpine watershed, Hydrol. Earth Syst. Sc., 16, 4343-4360, 2012.

Efron, B., and Tibshiran,i R. J.: An Introduction to the Bootstrap. Chapman & Hall, New York, 436, 1993.

Efstratiadis, A., and Koutsoyiannis, D.: One decade of multi-objective calibration approaches in hydrological modelling: A review, Hydrological Sciences Journal, 55(1), 58-78, 2010.

Ehret, U., Zehe, E., Wulfmeyer, V., Warrach-Sagi, K., and Liebert, J.: HESS Opinions "Should we apply bias correction to global and regional climate model data?", Hydrol. Earth Syst. Sci., 16, 3391-3404, 2012.

Frame, D. J., Aina, T., Christensen, C. M., Faull, N. E., Knight, S. H. E., Piani, C., Rosier, S. M., Yamazaki, K., Yamazaki, Y. H., and Allen, M. R.: The climate*prediction*.net BBC climate change experiment: design of the coupled model ensemble. Phil. Trans. R. Soc. A, 367, 855-870, 2009.

Hawkins, E., and Sutton, R.: The potential to narrow uncertainty in regional climate predictions, B. Am. Meteorol. Soc., 90(8), 1095–1107, 2009.

Hawkins, E., and Sutton, R.: The potential to narrow uncertainty in projections of regional precipitation change, Clim. Dynam., 37, 407–418, 2011.

Hingray, B., and Saïd, M.: Partitioning internal variability and model uncertainty components in a multimember multimodel ensemble of climate projections, J. Climate, 27, 6779-6798, 2014.

Hipel, K. W., and McLeod, A. I.: Time series modelling of water resources and environmental systems. Developments in Water Science, 54 (Elsevier), 1013 pp, 1994.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N. C., Tung, C. C., and Liu, H. H.: The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, Proceedings Royal Society London A, 454, 903-995, 1998.

Hooke, R., and Jeeves, T. A.: Direct search solution of numerical and statistical problems, Journal Association Computing Machines, 8(2), 212-229, 1961.

Kay, A. L., Davies, H. N., Bell, V. A., and Jones, R. G.: Comparison of uncertainty sources for climate change impacts: flood frequency in England, Climatic Change, 92, 41-63, 2009.

Lafaysse, M., Hingray, B., Mezghani, A., Gailhard, J., and Terray, L.: Internal variability and model uncertainty components in future hydrometeorological projections: The Alpine Durance basin, Water Resour. Res., 50, 3317–3341, doi:10.1002/2013WR014897, 2014.

McMahon, T. A., and Adeloye, A. J.: Water Resources Yield. (Water Resources Publications, CO, USA), 220pp, 2005.

McMahon, T. A., Kiem, A. S., Peel, M. C., Jordan, P. W., and Pegram, G. G. S.: A new approach to stochastically generating six-monthly rainfall sequences based on Empirical Model Decomposition, J. Hydrometeorol., 9, 1377-1389, 2008.

McMahon, T. A., Pegram, G. G. S., Vogel, R. M., and Peel, M. C.: Review of Gould Dincer reservoir storage-yield-reliability estimates, Adv. Water Resourc., 30, 1873–1882, 2007.

McMahon, T. A., Peel, M. C., and Karoly, D. J.: Assessment of precipitation and temperature data from CMIP3 Global Climate Models for hydrologic simulation, Submitted to Hydrol. Earth Syst. Sc., 2014.

McMillan, H., Freer, J., Pappenberger, F., Krueger, T., and Clark, M.: Impacts of uncertain river flow data on rainfall-runoff model calibration and discharge predictions, Hydrol. Process., 24, 1270-1284, 2010.

McMillan, H., Jackson, B., Clark, M., Kavetski, D., and Woods, R.: Rainfall uncertainty in hydrological modelling: An evaluation of multiplicative error models, J. Hydrol., 400, 83-94, 2011.

Matalas, N. C.: Mathematical assessment of synthetic hydrology, Water Resour. Res., 3, 937–945, 1967.

Meehl, G. A., Covey, C., Delworth, T., Latif, M., McAvaney, B., Mitchell, J. F. B., Stouffer, R. J., and Taylor, K. E.: The WCRP CMIP3 multi-model dataset: A new era in climate change research, B. Am. Meteorol. Soc., 88, 1383-1394, 2007.

Monro, J. C.: Direct search optimisation in mathematical modelling and a watershed model application. National oceanic Atmospheric Administration, National Weather Service, US Dept. of Commerce, NOAA, Silver Spring, MD, Tech. Memo. NWS HYDRO-12, 52, 1971.

Nash, J. E., and Sutcliffe, J. V.: River flow forecasting through conceptual models, 1. A discussion of principles, J. Hydrol., 10, 282-290, 1970.

Peel, M. C., and Blöschl, G.: Hydrological modelling in a changing world, Prog. Phys. Geog., 35(2), 249-261, 2011.

Peel, M. C., Finlayson, B. L., and McMahon, T. A.: Updated world map of the Köppen-Geiger climate classification, Hydrol. Earth Syst. Sc., 11, 1633-1644, 2007.

Peel, M. C., McMahon, T. A., and Finlayson, B. L.: Vegetation impact on mean annual evapotranspiration at a global catchment scale, Water Resour. Res., 46, W09508, doi:10.1029/2009/2009WR008233, 2010.

Peel, M. C., McMahon, T. A., and Pegram, G. G. S.: Assessing the performance of rational spline based Empirical Mode Decomposition using a global annual precipitation dataset, Proceedings Royal Society London Series A, 465, 1919-1937, 2009.

Peel, M. C., McMahon, T. A., Srikanthan, R., and Tan, K. S.: Ensemble Empirical Mode Decomposition: Testing and objective automation. Proceedings of the 33rd Hydrology and Water Resources Symposium, Brisbane, Engineers Australia, pp: 702-709, 2011a.

Peel, M. C., Srikanthan, R., McMahon, T. A., and Karoly, D. J.: Ensemble Empirical Mode Decomposition of monthly climatic indices relevant to Australian hydroclimatology. In Chan, F., Marinova, D. and Anderssen, R.S. (eds) MODSIM2011, 19th International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, December 2011, pp. 3615-3621, 2011b.

Pegram, G. G. S., Peel, M. C., and McMahon, T. A.: Empirical mode decomposition using rational splines: an application to rainfall time series, Proceedings Royal Society London Series A, 464, 1483-1501, 2008.

Petheram, C., Walker, G., Grayson, R., Thierfelder, T., and Zhang, L.: Towards a framework for predicting impacts of land-use on recharge: 1. A review of recharge studies in Australia, Australian Journal of Soil Science, 40, 397-417, 2002.

Petheram, C., McMahon, T. A., and Peel, M. C.: Flow characteristics of rivers in northern Australia: Implications for development, J. Hydrol., 357, 93-111, 2008.

Prudhomme, C., and Davies, H.: Assessing uncertainties in climate change impact analyses on the river flow regimes in the UK. Part 1: baseline climate, Climatic Change, 93, 177-195, 2009a.

Prudhomme, C., and Davies, H.: Assessing uncertainties in climate change impact analyses on the river flow regimes in the UK. Part 2: future climate, Climatic Change, 93, 197-222, 2009b.

Rowlands, D. J., Frame, D. J., Ackerley, D., Aina, T., Booth, B. B. B., Christensen, C., Collins, M., Faull, N., Forest, C. E., Grandey, B. S., Gryspeerdt, E., Highwood, E. J., Ingram, W. J., Knight, S., Lopez, A., Massey, N., McNamara, F., Meinshausen, N., Piani, C., Rosier, S. M., Sanderson, B. M., Smith, L. A., Stone, D. A., Thurston, M., Yamazaki, K., Yamazaki, Y. H., and Allen, M. R.: Broad range of 2050 warming from an observationally constrained large climate model ensemble. Nature Geoscience, 5, 256-260, 2012.

Salas, J. D.: Analysis and Modeling of Hydrologic Time Series, Handbook of Hydrology, Chap. 19, pp. 19.1-19.72, edited by D. R. Maidment, 1992.

Salas, J. D., Delleur, J. W., Yevjevich, V., and Lane, W. L.: Applied modelling of hydrologic series. Water Resources Publications, 1980.

Scanlon, B. R., Keese, K. E., Flint, A. L., Flint, L. E., Gaye, C. B., Edmunds, W. M., and Simmers, I.: Global synthesis of groundwater recharge in semiarid and arid regions, Hydrol. Process., 20, 3335-3370, 2006.

Sperna Weiland, F. C., van Beek, L. P. H., Kwadijk, J. C. J., and Bierkens, M. F. P.: On the suitability of GCM runoff fields for river discharge modelling: A case study using model output from HadGEM2 and ECHAM5, J. Hydrometeorol., 13, 140-154, 2012.

Srikanthan, R.: Stochastic generation of daily rainfall data using a nested model. 57th Canadian Water Resources Association Annual Congress, 16-18 June, Montreal, Canada, 2004.

Srikanthan, R., Peel, M. C., McMahon, T. A., and Karoly, D. J.: Ensemble empirical mode decomposition of Australian monthly rainfall and temperature data. In Chan, F., Marinova, D. and Anderssen, R.S. (eds) MODSIM2011, 19th International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, December 2011, pp. 3643-3649, 2011.

Steinschneider, S., Polebitski, A., Brown, C., and Letcher, B. H.: Toward a statistical framework to quantify the uncertainties of hydrologic response under climate change, Water Resour. Res., 48, W11525, doi:10.1029/2011WR011318, 2012.

Teutschbein, C., and Seibert, J.: Is bias correction of regional climate model (RCM) simulations possible for non-stationary conditions? Hydrol. Earth Syst. Sc., 17, 5061-5077, 2013.

Tebaldi, C., and Knutti, R.: The Use of the Multi-Model Ensemble in Probabilistic Climate Projections, Philosophical Transactions: Mathematical, Physical and Engineering Sciences, 365, 2053-2075, 2007.

Teng, J., Vaze, J., Chiew, F. H. S., Wang, B., and Perraud, J.-M.: Estimating the relative uncertainties sourced from GCMs and hydrological models in modeling climate change impact on runoff, J. Hydrometeorol., 13, 122-139, 2012.

Themeβl, M. J., Gobiet, A., and Heinrich, G.: Empirical-statistical downscaling and error correction of regional climate models and its impact on the climate change signal, Climatic Change, 112, 449-468, 2012.

Tisseuil, C., Vrac, M., Lek, S., and Wade, A. J.: Statistical downscaling of river flows, J. Hydrol., 385, 270-291, 2010.

van Oldenborgh, G. J., Philip, S. Y., and Collins, M.: El Niño in a changing climate: a multi-model study, Ocean Sci., 1, 81-95, 2005.

Vogel, R. M., and Sankarasubramanian, A.: Validation of a watershed model without calibration. Water Resour. Res., 39(10), 1292, 2003.

Wilson, E. B., and Hilferty, M. M.: The distribution of chi-square, Proceedings National Academy of Science, 17, 684-688, 1931.

Woldemeskel, F. M., Sharma, A., Sivakumar, B., and Mehrotra, R.: A framework to quantify GCM uncertainties for use in impact assessment studies, J. Hydrol., 519, 1453-1465, 2014.

Wu, Z., and Huang, N. E.: A study of the characteristics of white noise using the empirical mode decomposition method, Proceedings Royal Society London A, 460, 1597-1611, 2004.

Wu, Z., and Huang, N. E.: Ensemble empirical mode decomposition: A noise-assisted data analysis method, Advances in Adaptive Data Analysis, 1(1), 1-41, 2009.

Table 1. Comparison between MIROCM(1) GCM estimates of mean and standard deviation of annual precipitation and mean annual temperature of 20C3M data and stochastically generated values for six world-wide catchments. Generated values are based on 100 replicates, each 151 years long.

Reference number		1237	3284	4133	5410	6043	6304
River		Faleme	Duck	Huasco	Prut	Donnelly	Pioneer
Station		Gourbassy	Columbia	Algodones	Chernovtzsy	Nannup Road Br	Pleystowe Mill
Lat°		13.23	35.62	-28.73	48.26	-34.33	-21.15
Longº		-11.38	-87.03	-70.5	25.95	115.77	149.05
Catchment area (km2)†		14698	3140	7269	6874	755	1375
Köppen climate zone*		Aw	Cfa	BWk	Dfb	Csb	Cwa
MAP (mm yr^{-1})	GCM	1120	1111	126	763	620	1039
	Gen+	1120±20	1111±13	126±3.3	763±8.3	620±8.8	1039±99
SDP (mm yr^{-1})	GCM	212	155	39.1	94.7	94.4	274
	Gen	213±13	156±9	39.1±2.6	94.9±5.6	95.6±6.3	302±71
Precipitation lag-1	GCM	0.13	0.07	0.00	0.03	0.16	0.16
	Gen	0.17±0.08	0.10±0.07	0.01±0.08	0.09±0.08	0.24±0.07	0.77±0.07
MAT (°C)	GCM	26.6	16.6	15.0	8.87	14.8	22.5
	Gen	26.6±0.03	16.6±0.04	15.0±0.03	8.87±0.06	14.8±0.02	22.5±0.04

† catchment area from digital elevation model (see Peel et al., 2010) is within 5% of reported catchment area.

* see Peel et al. (2007)

+ Gen: shows the mean value ± standard deviation of 100 replicates; MAP: mean annual precipitation; SDP: standard deviation of annual precipitation; MAT: mean annual temperature

Table 2. Comparison between MIROCM(1) GCM estimates of mean and standard deviation of monthly precipitation and mean monthly temperature of 20C3M data and stochastically generated values for catchment 6304. Generated values are based on 100 replicates, each 151 years long.

Month	MMP (mm month^{-1})*		SDMP (mm month^{-1})		MMT (°C)	
	GCM	Gen+	GCM	Gen	GCM	Gen
1	163	158±10	103	86.3±7.9	26.3	26.3±0.1
2	172	168±10	93.3	80.8±6.9	26.2	26.2±0.1
3	150	147±9.3	79.4	72.2±5.5	25.1	25.1±0.1
4	75.1	73.8±11	45.0	49.1±4.6	23.0	23.0±0.1
5	52.3	53.1±11	30.5	40.0±5.3	20.5	20.5±0.1
6	48.3	50.4±9.9	36.7	41.1±5.0	18.5	18.5±0.1
7	31.5	38.1±9.2	29.0	39.4±5.8	17.5	17.5±0.1
8	34.4	38.0±9.3	32.8	36.8±5.4	18.7	18.7±0.1
9	33.1	37.5±9.7	28.6	35.6±5.7	20.8	20.8±0.1
10	54.8	56.2±9.8	38.6	44.7±4.4	23.0	23.0±0.1
11	96.4	93.9±8.8	59.0	58.9±4.4	24.6	24.6±0.1
12	127	125±10	83.2	75.0±6.7	26.0	26.0±0.1

*MMP: mean monthly precipitation; SDMP: standard deviation of monthly precipitation; MMT: mean monthly temperature.

+ Gen: shows the mean value ± standard deviation of 100 replicates.

1 Table 3. Details of the 17 selected catchments.

Ref. No.	Country	River	Station	Lat°	Longº	Catchment area (km^2)	Length of record (years)	Köppen climate zone*
1202	Mali	Bafing	Daka Saydou	12.15	-10.2	15500	21	Aw
1325	Benin	Oueme	Pont de Beterou	9.2	2.27	10326	16	Aw
1333	Zimbabwe	Sabi	Condo D/S G/W	-19.22	32.02	11000	26	Cwb
2270	China	Songhuajiang	Haerbin	45.77	126.58	391000	34	Dwa/Dwb
2274	India	Tapi	Kathore	21.28	72.95	61575	18	Aw/BSh
2288	China	Wujiang	Gongtan	28.9	108.35	58300	37	Cfa
3195	USA	Kiamichi	Belzoni	34.2	-95.48	3686	45	Cfa
3279	USA	Black	Kingstree	33.66	-79.84	3243	54	Cfa
3543	USA	Umpqua	Elkton	43.58	-123.55	9539	19	Csb
4014	Colombia	Magdalena	Puerto Berrio	6.5	-74.38	74410	31	Af/Aw/Cfb
4019	Guyana	Cuyuni	Kamaria Falls	6.43	-58.82	53354	30	Af/Aw
4145	Chile	Lumaco	Lumaco	-38.15	-72.9	1054	40	Csb
4179	Brazil	Rio Jaguaribe	Iguatu	-6.35	-39.3	21770	32	BSh
5255	United Kingdom	Clyde	Blairston	55.8	-4.07	1704	24	Cfb
6058	Australia	Herbert	Gleneagle	-18.2	145.33	5236	80	Cwa
6103	Australia	Nymboida	Nymboida	-29.98	152.72	1660	73	Cfa
6279	Australia	Ovens	Wangaratta	-36.36	146.35	5410	41	Cfb

2 *see Peel et al. (2007).

3

1 Table 4. Variation and uncertainty in key hydrologic statistics for five highest ranking GCMs
2 and 17 catchments, based on 100 replicates of de-trended 20C3M for the period 1965-1994.

Hydrologic variable		HadCM3	MIROCM(1)	MIUB(1)	MPI(1)	MRI(3)
1202 Bafing at Daku Saydou (Mali)						
Annual P	MAP	1733±45*	1700±46	1743±43	1743±39	1693±41
	SDP	195±27	201±28	182±21	206±30	194±26
Annual T	MAT	23.7±0.05	23.6±0.04	23.7±0.05	23.7±0.06	23.8±0.05
Annual R	MAR	620±33	589±33	595±30	587±27	566±29
	SDR	143±21	147±22	126±16	143±21	135±19
	Lag-1	0.14±0.17	0.11±0.19	0.12±0.19	0.05±0.19	0.05±0.19
1325 Oueme at Pont de Beterou (Benin)						
Annual P	MAP	1292±30	1358±41	1257±37	1309±28	1254±30
	SDP	169±21	193±28	169±24	159±22	157±21
Annual T	MAT	26.3±0.04	26.2±0.04	26.3±0.05	26.3±0.05	26.4±0.04
Annual R	MAR	226±18	272±24	189±20	202±16	161±14
	SDR	99±14	115±18	96±14	86±13	80±12
	Lag-1	-0.03±0.19	0.05±0.18	0.16±0.19	-0.09±0.16	-0.07±0.18
1333 Sabi at Condo D/S G/W (Zimbabwe)						
Annual P	MAP	823±38	799±35	855±26	785±41	854±44
	SDP	204±27	185±24	149±19	185±24	214±28
Annual T	MAT	19.3±0.09	19.4±0.06	19.3±0.06	19.4±0.07	19.4±0.06
Annual R	MAR	146±17	120±14	130±10	115±15	128±17
	SDR	93±19	75±17	59±9	72±15	87±21
	Lag-1	-0.07±0.17	-0.01±0.19	-0.09±0.17	0.01±0.19	0.06±0.18
2270 Songhuajiang at Haerbin (China)						
Annual P	MAP	492±20	477±18	510±16	495±16	514±17
	SDP	84±14	72±13	72±10	84±12	79±11
Annual T	MAT	1.5±0.13	1.7±0.11	1.7±0.18	1.8±0.14	1.8±0.11
Annual R	MAR	91±10	87±10	99±9	90±9	100±9
	SDR	39±7	37±8	37±6	41±6	39±6
	Lag-1	0.25±0.16	0.20±0.19	0.19±0.18	0.19±0.16	0.19±0.17
2274 Tapi at Kathore (India)						
Annual P	MAP	887±38	865±40	854±38	815±42	891±46
	SDP	204±27	190±23	175±22	206±24	216±34
Annual T	MAT	26.9±0.07	26.9±0.07	26.7±0.10	26.9±0.10	26.7±0.08
Annual R	MAR	259±25	269±28	259±25	241±27	282±32
	SDR	129±19	131±19	117±16	135±18	143±23
	Lag-1	0.02±0.18	0.04±0.19	0.09±0.17	-0.04±0.17	0.21±0.21
2288 Wujiang at Gongtan (China)						
Annual P	MAP	1136±31	1125±26	1186±27	1126±31	1172±29
	SDP	159±23	141±19	139±20	147±20	166±25
Annual T	MAT	15.7±0.09	15.8±0.09	15.7±0.08	15.8±0.09	15.7±0.08
Annual R	MAR	589±31	576±27	640±27	578±32	625±28
	SDR	137±23	126±18	122±18	127±19	143±22
	Lag-1	0.17±0.16	0.22±0.17	0.00±0.18	0.16±0.17	0.07±0.18
3195 Kiamichi at Belzoni (United States)						
Annual P	MAP	1237±53	1302±60	1260±42	1318±65	1308±49

	SDP	267±38	260±38	242±33	284±43	269±36
Annual T	MAT	17.1±0.12	16.8±0.13	16.7±0.14	16.6±0.13	16.9±0.11
Annual R	MAR	407±37	466±47	419±35	483±53	480±41
	SDR	199±30	213±34	187±29	227±38	221±31
	Lag-1	0.02±0.17	0.06±0.19	-0.02±0.17	0.08±0.21	0.01±0.17
3279 Black at Kingstree (United States)						
Annual P	MAP	1110±60	1147±43	1164±29	1127±47	1187±37
	SDP	236±34	198±26	177±23	231±33	208±29
Annual T	MAT	17.5±0.12	17.3±0.11	17.3±0.11	17.4±0.12	17.5±0.11
Annual R	MAR	255±37	276±29	272±21	273±31	294±25
	SDR	140±26	128±22	112±18	143±24	130±19
	Lag-1	0.24±0.17	0.15±0.20	0.05±0.21	0.14±0.18	0.04±0.19
3543 Umpqua at Elkton (United States)						
Annual P	MAP	1085±44	1088±43	1025±36	1134±42	1142±36
	SDP	211±29	224±29	197±27	218±27	217±31
Annual T	MAT	10.3±0.12	10.8±0.09	10.5±0.10	10.7±0.08	10.6±0.11
Annual R	MAR	752±38	734±40	679±35	784±37	801±33
	SDR	189±26	213±27	183±26	203±26	191±26
	Lag-1	0.02±0.19	-0.01±0.19	-0.02±0.20	-0.01±0.19	-0.02±0.17
4014 Magdalena at Puerto Berrio (Colombia)						
Annual P	MAP	1962±46	1843±57	1800±53	1905±37	1901±40
	SDP	226±32	189±30	145±25	206±31	176±24
Annual T	MAT	20.7±0.08	20.7±0.05	20.7±0.07	20.7±0.07	20.8±0.08
Annual R	MAR	1085±30	1007±37	978±34	1047±24	1044±26
	SDR	138±20	115±20	85±17	121±17	103±16
	Lag-1	0.15±0.18	0.42±0.17	0.50±0.16	0.06±0.16	0.26±0.18
4019 Cuyuni at Kamaria Falls (Guyana)						
Annual P	MAP	1566±89	1457±90	1454±31	1413±49	1442±42
	SDP	306±43	258±46	185±26	218±33	244±31
Annual T	MAT	25.1±0.11	25.3±0.13	25.1±0.08	25.2±0.07	25.2±0.06
Annual R	MAR	818±85	697±87	686±31	650±48	673±40
	SDR	293±39	236±49	164±22	203±30	227±30
	Lag-1	0.33±0.17	0.53±0.19	-0.03±0.17	0.21±0.17	-0.04±0.17
4145 Lumaco at Lumaco (Chile)						
Annual P	MAP	1084±47	1037±58	1096±50	1088±43	1052±43
	SDP	217±28	204±29	210±29	224±35	210±29
Annual T	MAT	11.5±0.07	11.6±0.10	11.5±0.08	11.5±0.08	11.6±0.09
Annual R	MAR	595±38	553±47	613±41	584±37	564±35
	SDR	171±24	161±25	175±24	188±31	167±24
	Lag-1	0.12±0.18	0.39±0.17	0.23±0.18	0.06±0.20	0.09±0.17
4179 Rio Jaguaribe at Iguata (Brazil)						
Annual P	MAP	801±96	808±65	675±34	744±38	694±46
	SDP	315±53	266±44	180±23	236±34	260±40
Annual T	MAT	26.6±0.12	26.6±0.12	26.5±0.09	26.5±0.14	26.9±0.08
Annual R	MAR	51.5±22	48.7±16	45.3±9.0	64.7±14	56.4±16
	SDR	81.0±34	62.0±24	50.1±14	76.9±22	81.6±30
	Lag-1	0.20±0.19	0.16±0.19	0.00±0.18	-0.02±0.17	-0.05±0.17
5255 Clyde at Blairston (United Kingdom)						
Annual P	MAP	1025±29	1032±29	989±27	1026±28	978±30

	SDP	154±20	140±17	147±20	134±16	140±20
Annual T	MAT	8.4±0.09	8.5±0.10	8.7±0.10	8.4±0.10	8.7±0.10
Annual R	MAR	762±28	774±29	717±27	784±31	704±30
	SDR	149±20	141±17	149±19	145±21	141±21
	Lag-1	-0.05±0.16	0.05±0.19	0.00±0.19	-0.02±0.17	0.04±0.18
6058 Herbert at Gleneagle (Australia)						
Annual P	MAP	854±44	908±73	900±42	873±50	868±44
	SDP	245±42	323±52	253±40	267±50	251±37
Annual T	MAT	21.5±0.09	21.4±0.08	21.5±0.11	21.5±0.09	21.6±0.07
Annual R	MAR	184±28	230±44	210±29	202±29	201±29
	SDR	150±34	206±45	161±33	166±39	158±31
	Lag-1	-0.06±0.17	0.03±0.16	-0.08±0.17	0.01±0.19	-0.03±0.19
6103 Nymboida at Nymboida (Australia)						
Annual P	MAP	1477±83	1432±88	1479±69	1497±92	1474±90
	SDP	414±72	436±66	365±55	437±62	412±66
Annual T	MAT	17.1±0.07	17.0±0.06	17.2±0.08	17.0±0.08	17.2±0.06
Annual R	MAR	493±59	475±59	480±48	510±64	485±61
	SDR	268±61	282±59	236±47	289±55	270±59
	Lag-1	0.11±0.18	0.15±0.18	0.03±0.20	0.20±0.14	0.11±0.19
6279 Ovens at Wangaratta (Australia)						
Annual P	MAP	1137±58	823±42	1057±36	1074±48	1040±42
	SDP	243±33	181±25	206±30	257±33	214±33
Annual T	MAT	12.6±0.08	12.4±0.10	12.7±0.11	12.5±0.10	13.0±0.08
Annual R	MAR	215±24	106±14	186±15	195±21	182±18
	SDR	105±17	60±10	84±15	106±15	88±14
	Lag-1	0.14±0.17	0.29±0.17	0.00±0.18	0.01±0.16	0.07±0.18

1 * Mean value ± standard deviation based on 100 replicates.

2

Table 5. Relative within-GCM uncertainty of mean and standard deviation of annual precipitation, mean annual temperature and mean, standard deviation and lag-1 serial correlation of annual runoff. Relative uncertainty is the standard deviation of the 100 replicate estimates as a percentage of the mean replicate estimate for each GCM during the period 1965-1994 (20C3M). The average of the 17 catchment relative uncertainty values is presented, except for lag-1 annual runoff which is the average of the 17 standard deviations.

Variable	HadCM3	MIROCM(1)	MIUB(1)	MPI(1)	MRI(3)	Average
MAP	4.6%	4.6%	3.4%	3.9%	3.8%	4.1%
SDP	14.4%	14.6%	13.9%	14.2%	14.2%	14.3%
MAT	1.0%	0.9%	1.1%	1.0%	0.8%	1.0%
MAR	10.8%	10.9%	8.1%	9.3%	9.3%	9.7%
SDR	17.9%	18.4%	16.4%	16.8%	17.3%	17.4%
Lag-1	0.17	0.18	0.18	0.18	0.18	0.18

Table 6. Relative within-GCM uncertainty of mean and standard deviation of annual precipitation, mean annual temperature and mean, standard deviation and lag-1 serial correlation of annual runoff. Relative uncertainty is the standard deviation of the 100 replicate estimates as a percentage of the mean replicate estimate for each GCM during the period 2015-2044 (A1B). The average of the 17 catchment relative uncertainty values is presented, except for lag-1 annual runoff which is the average of the 17 standard deviations.

Variable	HadCM3	MIROCM(1)	MIUB(1)	MPI(1)	MRI(3)	Average
MAP	4.9%	4.9%	3.4%	3.8%	3.8%	4.2%
SDP	14.7%	14.4%	14.4%	13.6%	13.7%	14.2%
MAT	0.8%	0.6%	0.7%	0.8%	0.7%	0.7%
MAR	13.0%	11.7%	8.0%	8.9%	9.0%	10.1%
SDR	20.1%	18.8%	16.6%	16.0%	16.5%	17.6%
Lag-1	0.18	0.17	0.18	0.18	0.18	0.18

Table 7. Average and within-GCM uncertainty of reservoir yield (mm year^{-1}) for 17 catchments using the Gould-Dincer Gamma reservoir storage model for two reservoir sizes (τ = 1x MAR and τ = 3x MAR) and 95% reliability of draft. Average and uncertainty (standard deviation) are calculated from 100 bias-corrected replicates of precipitation and temperature passed through the PERM model for each GCM over the period 2015-2044 (A1B).

Ref. No.	Observed MAP (mm yr^{-1})	Observed MAR (mm yr^{-1})	Observed CvR‡	Relative reservoir size	HadCM3 (mm yr^{-1})	MIROCM(1) (mm yr^{-1})	MIUB(1) (mm yr^{-1})	MPI(1) (mm yr^{-1})	MRI(3) (mm yr^{-1})
1202	1672	569	0.26	1xMAR	609±40	531±39	620±34	516±34	477±33
				3xMAR	632±33	560±34	641±28	537±31	497±25
1325	1253	194	0.54	1xMAR	208±23	238±29	194±40	188±20	121±19
				3xMAR	229±18	268±25	228±28	205±16	143±14
1333	809	126	0.75	1xMAR	90.7±23	75.9±23	112±15	56.6±19	58.0±27
				3xMAR	119±15	105±13	125±10	82.8±11	93.4±15
2270	495	93.9	0.46	1xMAR	68.3±21	74.0±15	89.3±13	70.1±18	82.2±16
				3xMAR	88.6±11	89.8±9	100±8	87.6±10	95.7±10
2274	801	225	0.53	1xMAR	221±40	236±39	244±33	178±51	316±68
				3xMAR	259±27	277±28	278±25	226±29	371±43
2288	1158	617	0.20	1xMAR	557±45	429±44	601±28	541±36	550±37
				3xMAR	587±33	459±34	615±25	563±31	577±29
3195	1234	416	0.45	1xMAR	243±53	197±77	336±42	362±66	391±58
				3xMAR	303±36	273±42	380±30	439±46	443±46
3279	1122	260	0.54	1xMAR	173±69	104±40	220±37	236±69	259±42
				3xMAR	242±53	152±24	249±24	303±41	300±30
3543	1078	721	0.35	1xMAR	599±43	605±44	560±36	688±38	762±39
				3xMAR	630±34	639±39	588±33	715±35	786±35
4014	1900	1043	0.14	1xMAR	1046±37	1058±37	1204±63	1072±23	1122±28
				3xMAR	1059±35	1078±34	1247±41	1081±23	1131±27
4019	1397	636	0.29	1xMAR	283±162	147±109	651±31	341±65	622±52
				3xMAR	475±90	243±88	669±29	400±38	657±42
4145	1050	570	0.33	1xMAR	434±52	229±99	382±79	417±55	399±46
				3xMAR	476±38	318±54	441±46	464±40	433±34
4179	636	38.9	1.37	1xMAR	N/R*	N/R	7.8±9	7.6±11	N/R
				3xMAR	N/R	N/R	30.8±9	37.9±14	19.9±13
5255	1007	749	0.17	1xMAR	729±27	765±28	706±32	723±24	714±31
				3xMAR	742±24	777±27	722±30	735±22	727±29
6058	873	201	0.88	1xMAR	59.8±41	52.6±61	116±48	83.5±55	89.3±52
				3xMAR	130±31	169±52	181±32	161±38	164±31
6103	1455	482	0.59	1xMAR	251±109	494±146	391±85	264±112	273±102
				3xMAR	370±58	653±96	476±54	394±64	390±61
6279	1119	206	0.55	1xMAR	62.0±35	27.6±22	162±20	131±34	125±27
				3xMAR	107±20	54.7±17	184±15	169±21	155±19

‡ Coefficient of variation of annual runoff, * N/R indicates more than half the replicates had negative draft estimate set to zero, so statistics are not reported.

Table 8. Relative within-GCM uncertainty of reservoir yield for hypothetical reservoirs of 1x and 3x MAR. The average of 17 catchment relative uncertainty values is presented. Relative uncertainty is the standard deviation of the 100 replicate estimates as a percentage of the mean replicate estimate for each GCM over the period 2015-2044 (A1B).

Variable	HadCM3	MIROCM(1)	MIUB(1)	MPI(1)	MRI(3)	Average
1xMAR	25.9%	33.2%	19.6%	28.3%	18.6%	25.1%
3xMAR	11.6%	14.3%	9.3%	11.6%	12.7%	11.9%

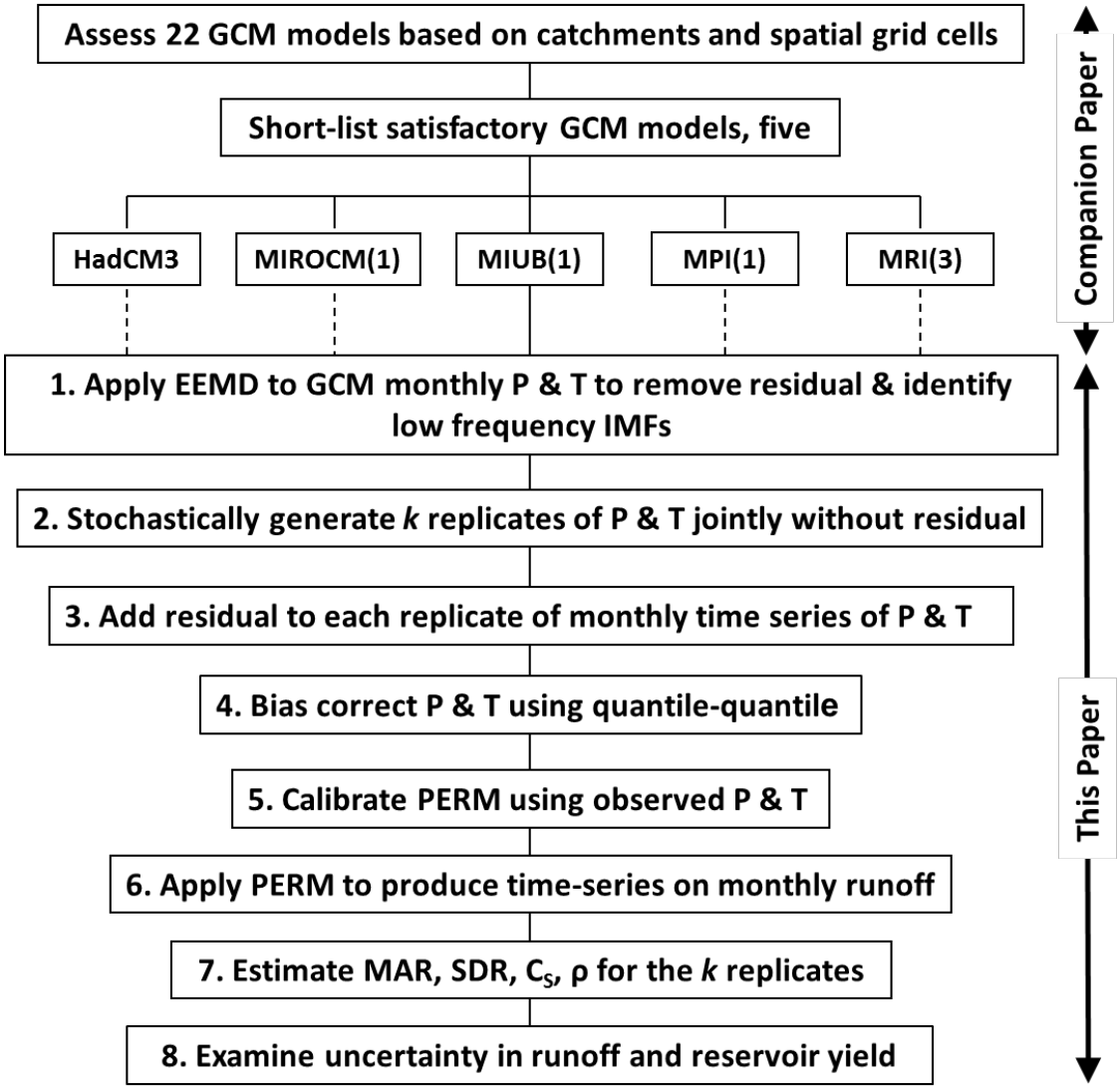
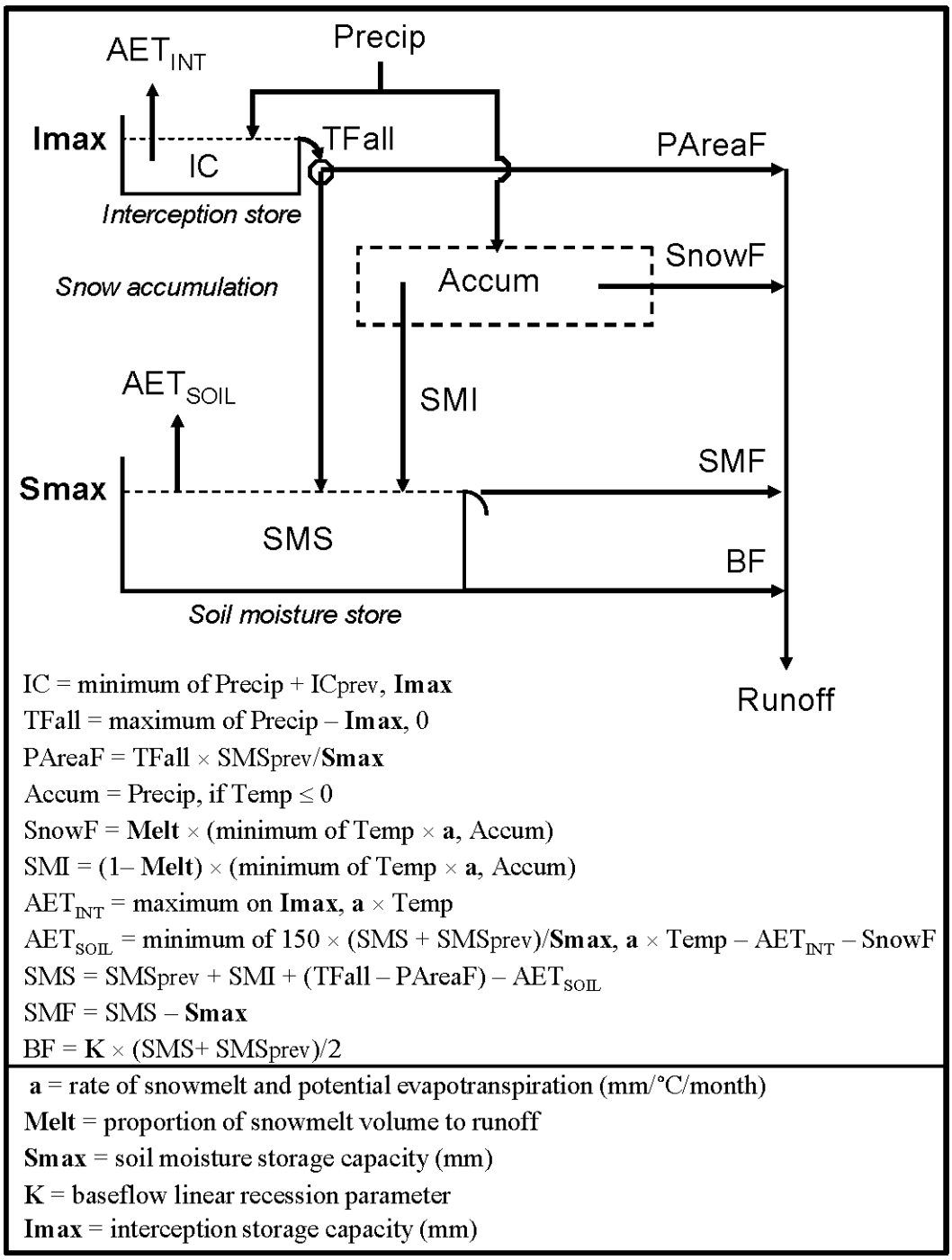


Figure 1. Outline of process to approximate within-GCM uncertainty of future runoff and reservoir yield. The companion paper is McMahon et al. (2014).



IC = minimum of Precip + ICprev, **Imax**
TFall = maximum of Precip – **Imax**, 0
PAreaF = TFall × SMSprev/**Smax**
Accum = Precip, if Temp ≤ 0
SnowF = **Melt** × (minimum of Temp × **a**, Accum)
SMI = (1– **Melt**) × (minimum of Temp × **a**, Accum)
AET_{INT} = maximum on **Imax**, **a** × Temp
AET_{SOIL} = minimum of 150 × (SMS + SMSprev)/**Smax**, **a** × Temp – AET_{INT} – SnowF
SMS = SMSprev + SMI + (TFall – PAreaF) – AET_{SOIL}
SMF = SMS – **Smax**
BF = **K** × (SMS+ SMSprev)/2

a = rate of snowmelt and potential evapotranspiration (mm/°C/month)
Melt = proportion of snowmelt volume to runoff
Smax = soil moisture storage capacity (mm)
K = baseflow linear recession parameter
Imax = interception storage capacity (mm)

Figure 2. Structure of the monthly conceptual precipitation-evapotranspiration-runoff model (PERM) where the five calibration parameters are highlighted in bold.

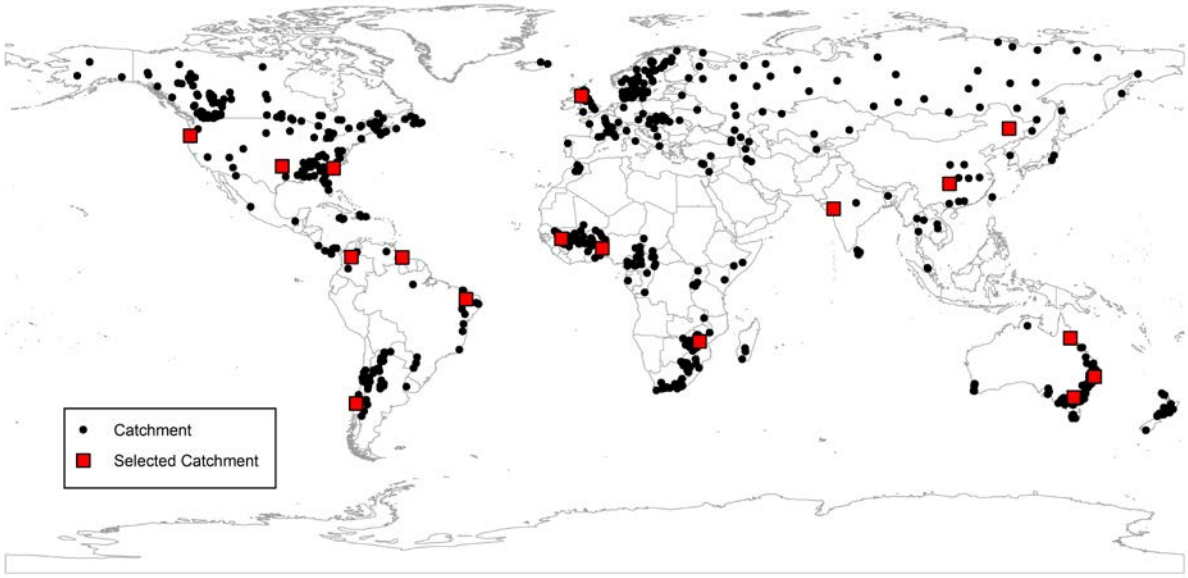


Figure 3. Locations of the initial 699 catchments and the final sub-set of 17 catchments.

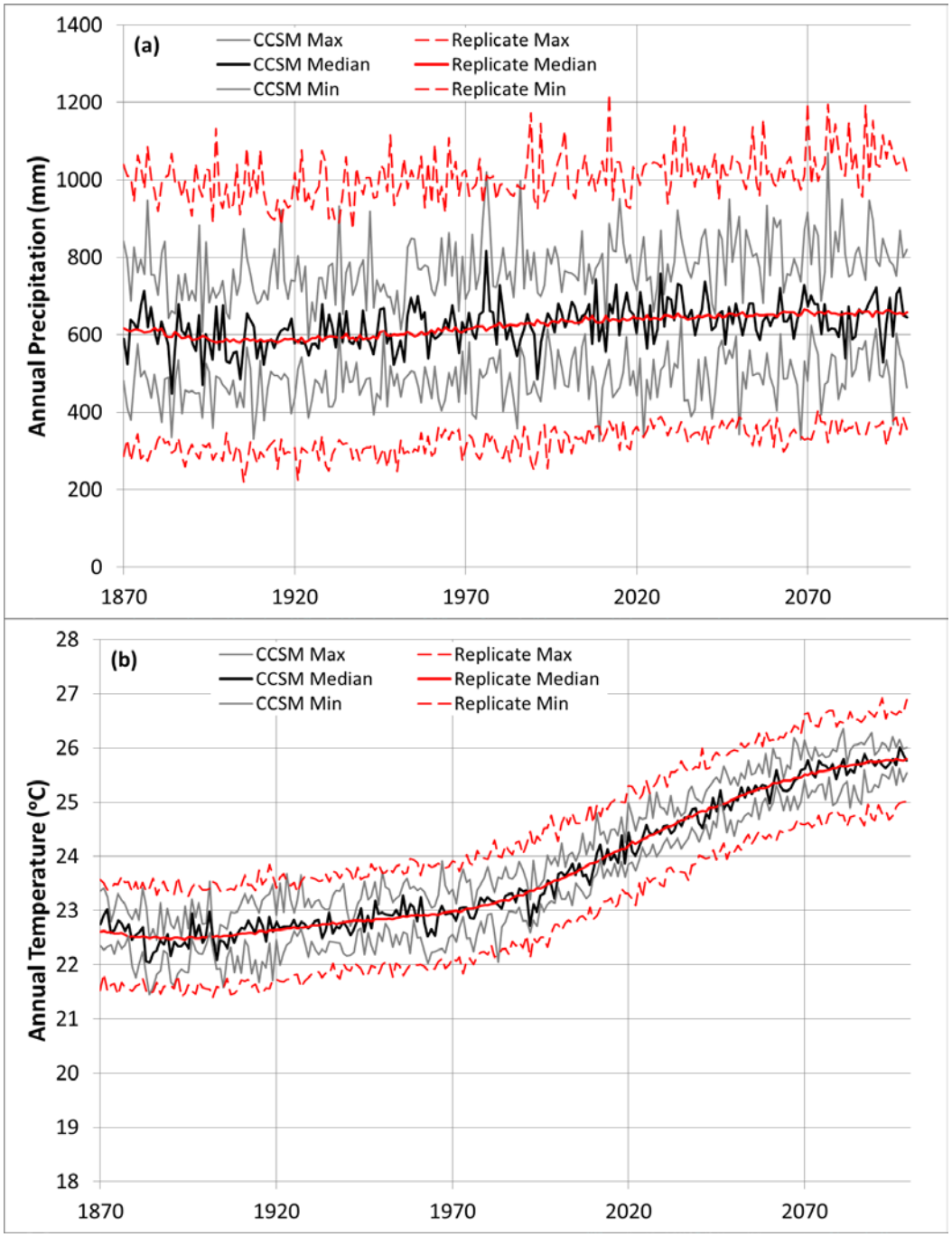


Figure 4. Within-GCM uncertainty for the Herbert River at Gleneagle based on seven runs from the CCSM GCM compared with the stochastic approximation of within-GCM uncertainty for un-bias corrected (a) annual precipitation and (b) annual temperature. In each plot the maximum, median and minimum annual value for a given year are shown for the seven CCSM runs compared with the maximum, median and minimum of the 700 (7 x 100) stochastic replicates of the CCSM runs.

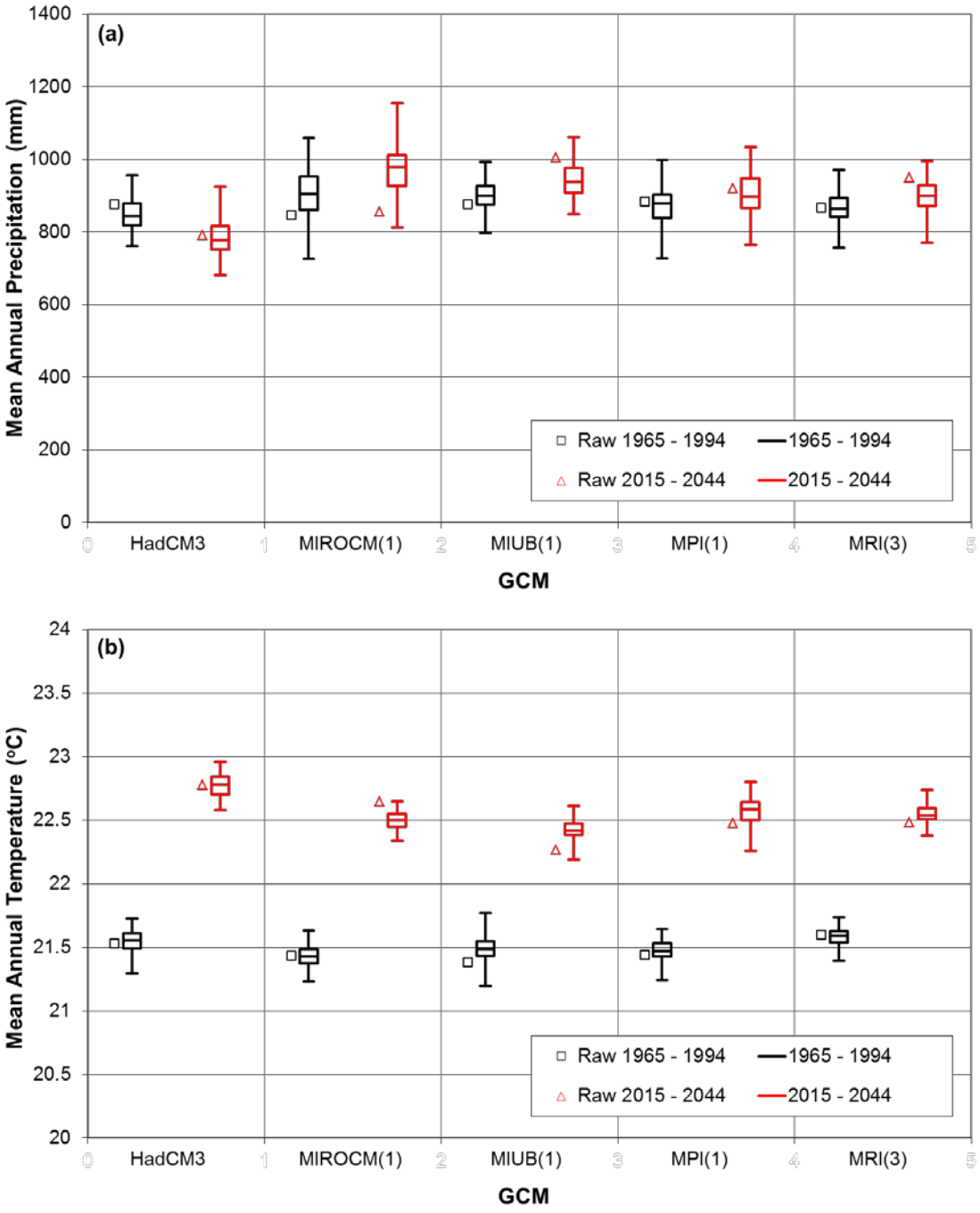


Figure 5. Box-plots of 30-year mean annual (a) precipitation and (b) temperature for the periods 1965-1994 (20C3M) and 2015-2044 (A1B) for five GCMs. Each box-plot is based on 100 quantile-quantile bias-corrected stochastic replicates of GCM data for catchment 6058 – Herbert River at Gleneagle (Australia). The box represents the inter-quartile range and the whiskers extend to the maximum and minimum values. The 'Raw' value next to each box-

plot represents the bias-corrected mean annual value from the GCM run that the stochastic replicates are based on.

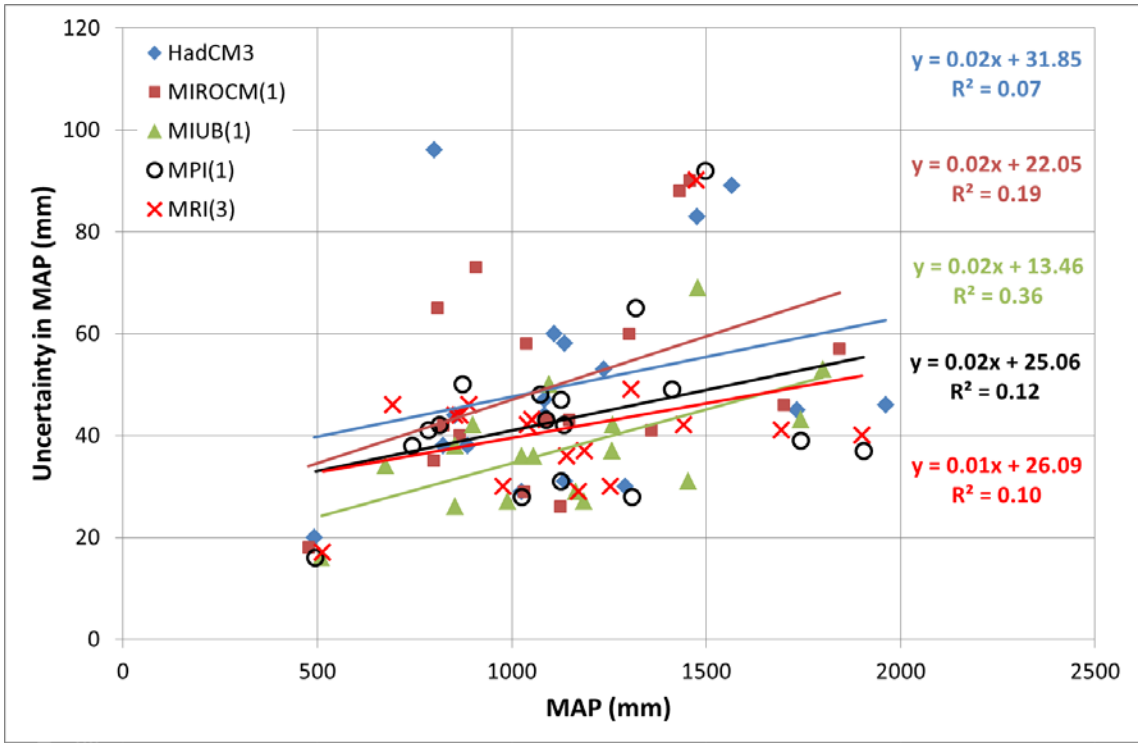


Figure 6. Within-GCM uncertainty (mm year^{-1}) in mean annual precipitation versus mean annual precipitation based on 100 replicates of monthly precipitation (1965-1994, 20C3M) for five GCMs.

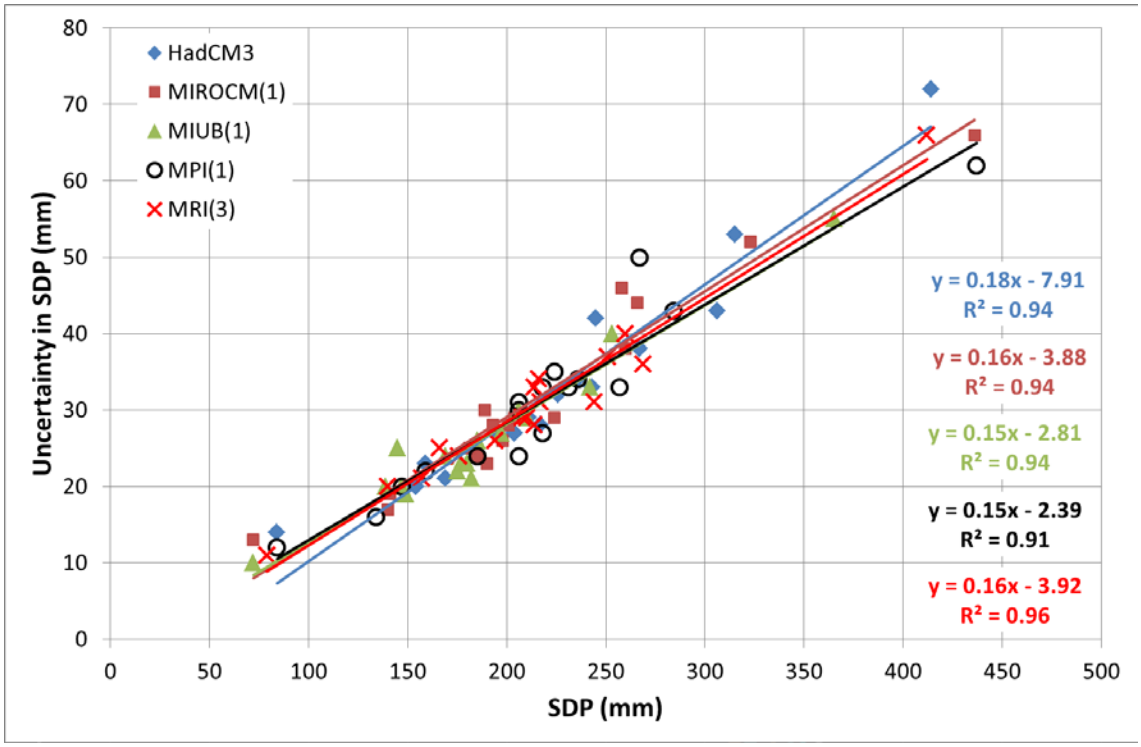


Figure 7. Within-GCM uncertainty (mm year^{-1}) in standard deviation of annual precipitation versus the standard deviation of annual precipitation based on 100 replicates of monthly precipitation (1965-1994, 20C3M) for five GCMs.

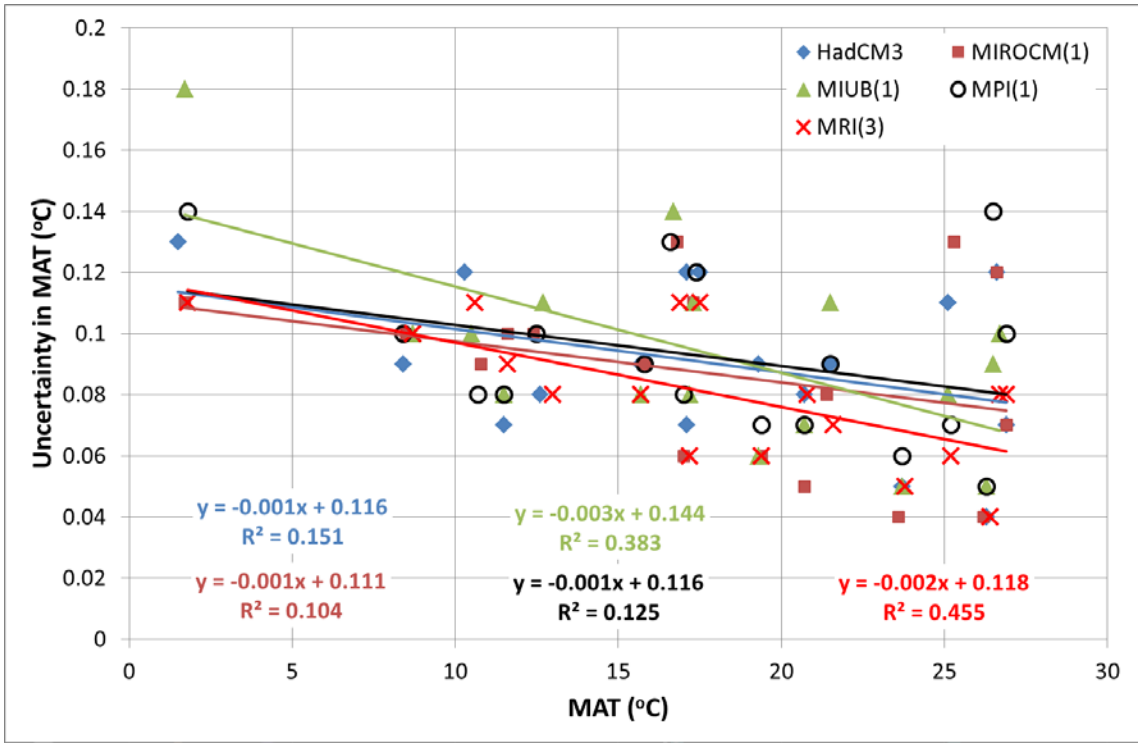
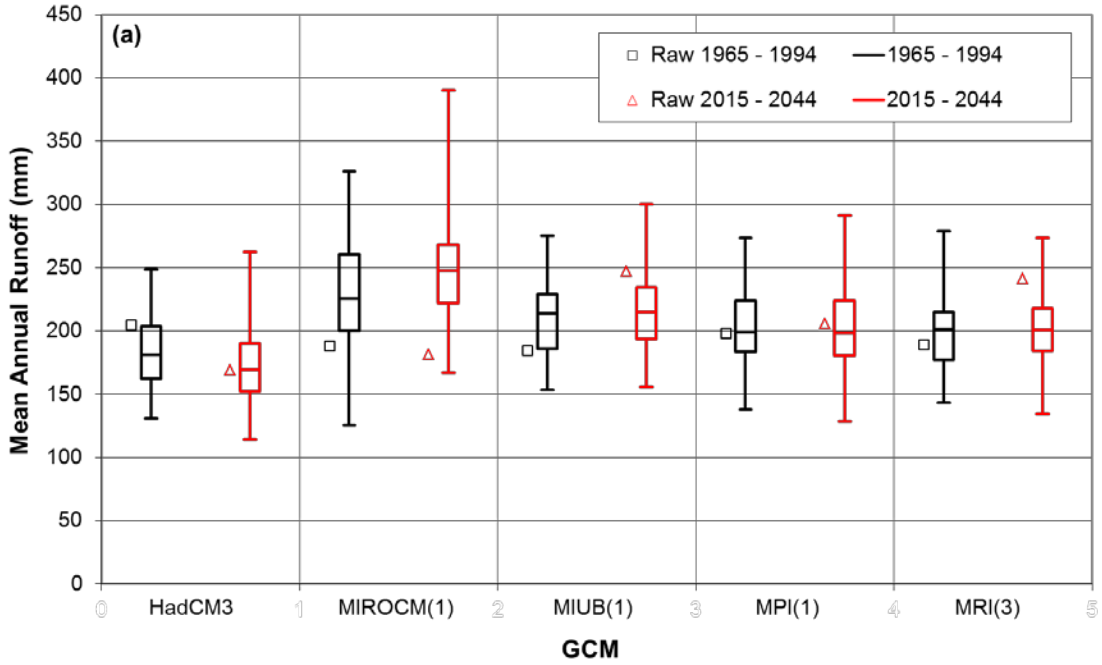


Figure 8. Within-GCM uncertainty (°C) in mean annual temperature versus mean annual temperature based on 100 replicates of monthly temperature (1965-1994, 20C3M) for five GCMs.



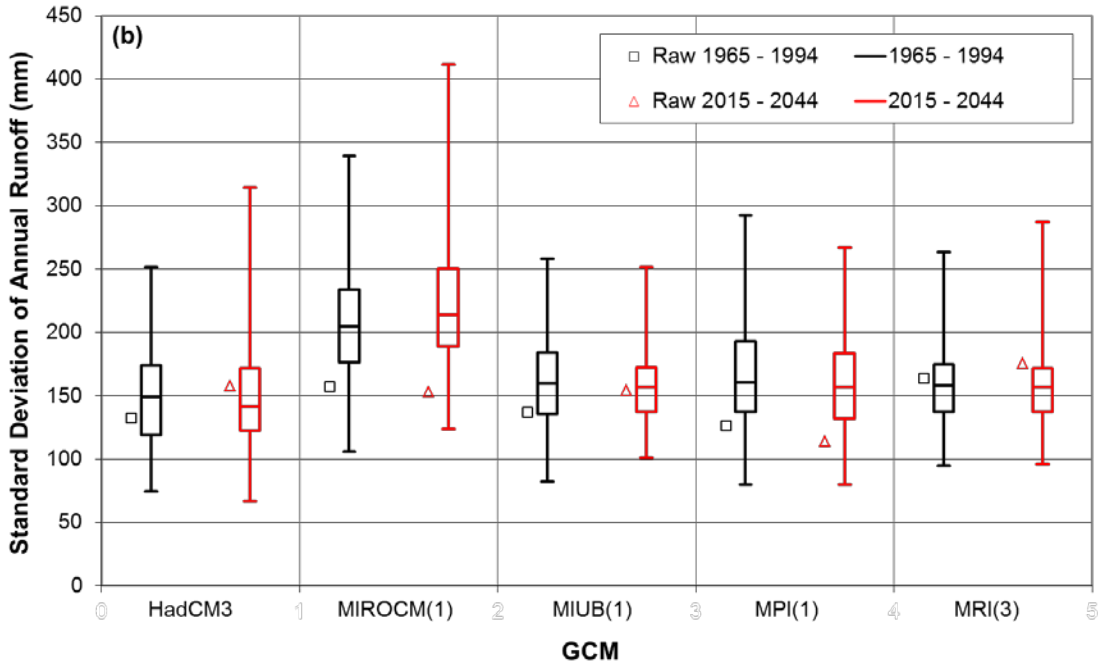


Figure 9. Box-plots of 30-year (a) mean annual runoff and (b) standard deviation of annual runoff for the periods 1965-1994 (20C3M) and 2015-2044 (A1B) for five GCMs. Each box-plot is based on 100 quantile-quantile bias-corrected stochastic replicates of GCM data that have been input to the PERM model of catchment 6058 – Herbert River at Gleneagle (Australia). The box represents the inter-quartile range and the whiskers extend to the maximum and minimum values. The 'Raw' value next to each box-plot represents the PERM output when using the bias-corrected GCM runs that the stochastic replicates are based on.

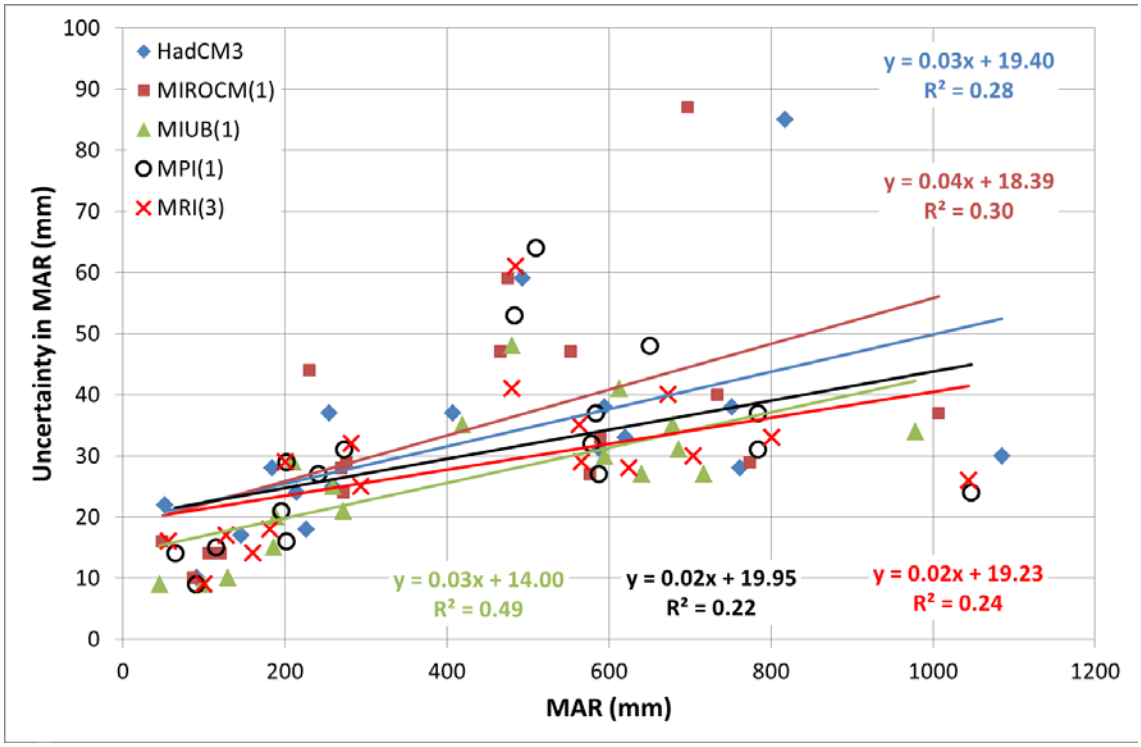


Figure 10. Within-GCM uncertainty (mm year^{-1}) in mean annual runoff versus mean annual runoff based on 100 PERM runs using replicates of monthly precipitation and temperature (1965-1994, 20C3M) as input from five GCMs.

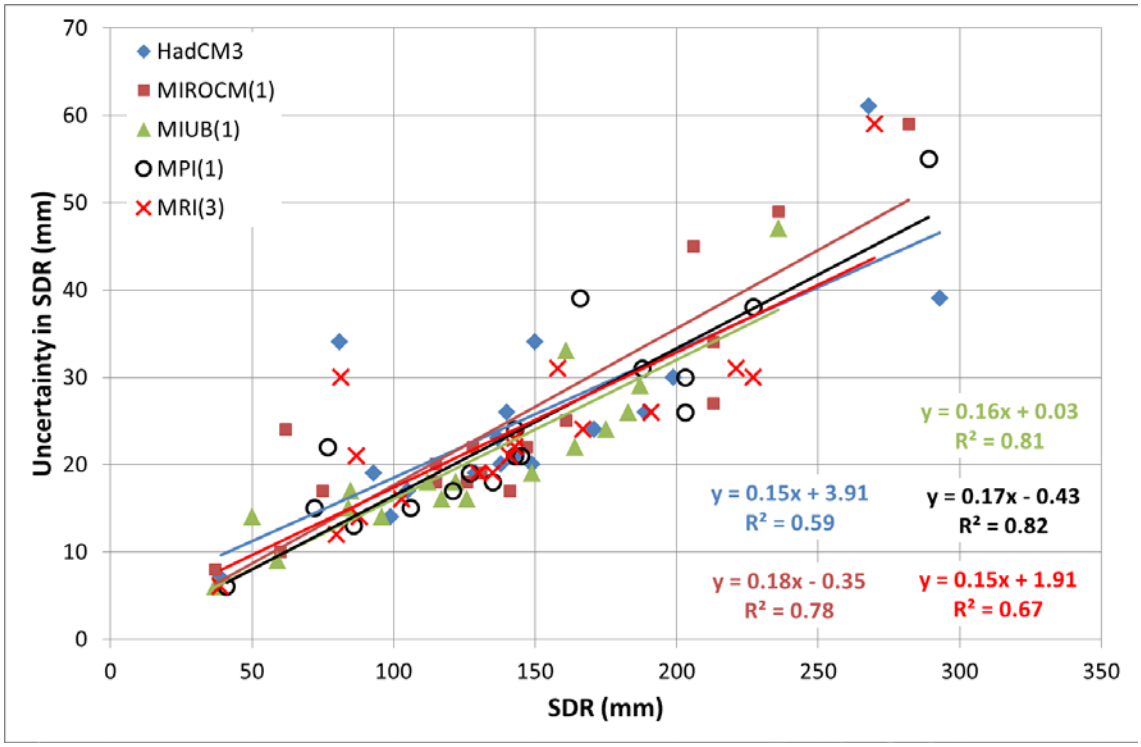


Figure 11. Within-GCM uncertainty (mm year^{-1}) in standard deviation of annual runoff versus standard deviation of annual runoff based on 100 PERM runs using replicates of monthly precipitation and temperature (1965-1994, 20C3M) as input from five GCMs.

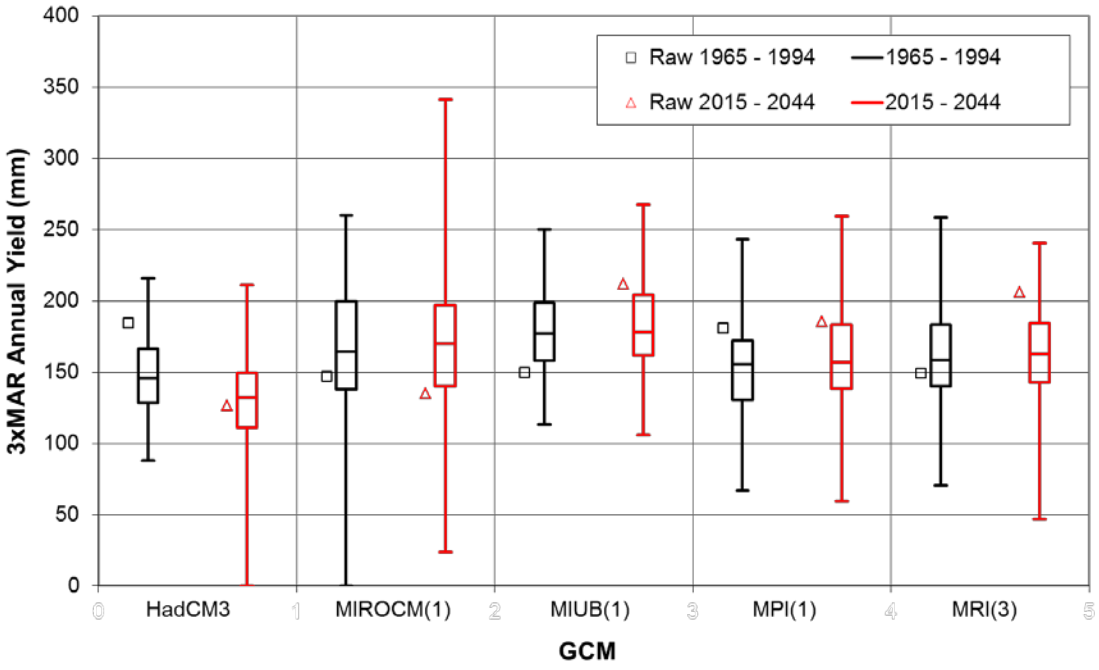


Figure 12. Box-plots of average reservoir yield for the 30-year periods 1965-1994 (20C3M) and 2015-2044 (A1B) for five GCMs. Reservoir yield is estimated using the Gould-Dincer Gamma reservoir storage model for reservoir size τ = 3x MAR and 95% reliability of draft. Runoff metrics for the Gould-Dincer Gamma method are estimated from 100 PERM runs of quantile-quantile bias-corrected stochastic replicates of GCM data for catchment 6058 – Herbert River at Gleneagle (Australia). The box represents the inter-quartile range and the whiskers extend to the maximum and minimum values. The 'Raw' value next to each box-plot represents the Gould-Dincer Gamma output when using the bias-corrected GCM runs that the stochastic replicates are based on.

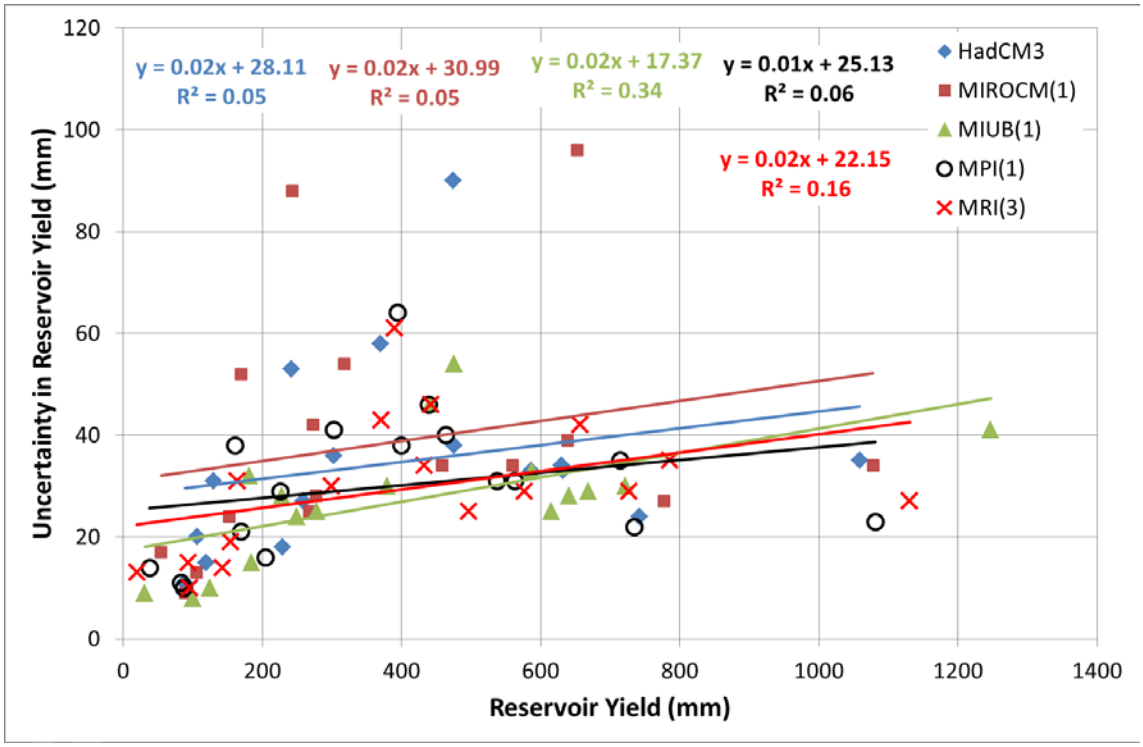


Figure 13. Within-GCM uncertainty (mm year^{-1}) in reservoir yield versus reservoir yield based on 100 Gould-Dincer Gamma estimates (hypothetical capacity: τ = 3, reliability of draft: 95%) from PERM runs of monthly precipitation and temperature replicates (2015-2044, A1B) from five GCMs.

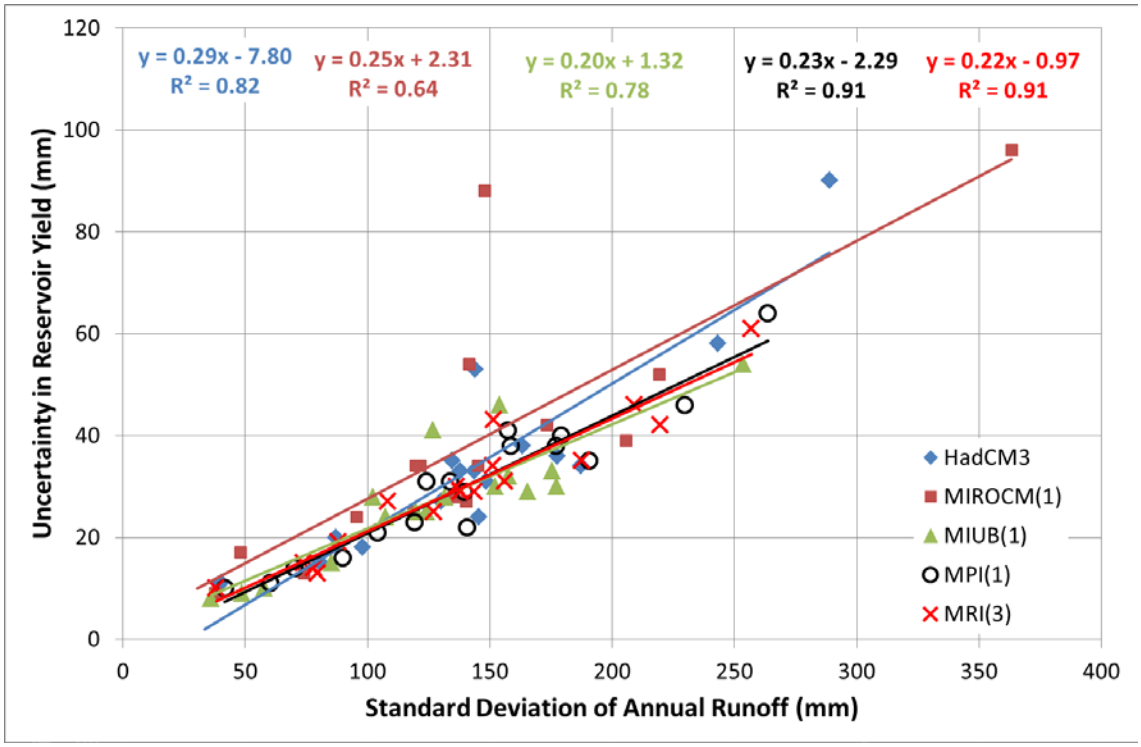


Figure 14. Within-GCM uncertainty (mm year^{-1}) in reservoir yield versus standard deviation of annual runoff based on 100 Gould-Dincer Gamma estimates (hypothetical capacity: τ = 3, reliability of draft: 95%) from PERM runs of monthly precipitation and temperature replicates (2015-2044, A1B) from five GCMs.