Hydrol. Earth Syst. Sci. Discuss., 11, 4531–4578, 2014
www.hydrol-earth-syst-sci-discuss.net/11/4531/2014/
doi:10.5194/hessd-11-4531-2014

Hydrology and Earth System Sciences
Open Access
Discussions

Uncertainty in runoff based on Global Climate Model precipitation and temperature data – Part 1: Assessment of Global Climate Models

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

[1]Department of Infrastructure Engineering, University of Melbourne, Victoria, 3010, Australia
[2]School of Earth Sciences and ARC Centre of Excellence for Climate System Science, University of Melbourne, Victoria, 3010, Australia

Received: 25 March 2014 – Accepted: 31 March 2014 – Published: 5 May 2014

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

Published by Copernicus Publications on behalf of the European Geosciences Union.

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. Uncertainty between GCM projections of future climate can be assessed through analysis of runs of a given scenario from a wide range of GCMs. 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 objective of this, the first of two complementary papers, is to reduce between-GCM uncertainty by identifying and removing poorly performing GCMs prior to the analysis presented in the second paper. Here we assess how well 46 runs from 22 Coupled Model Intercomparison Project phase 3 (CMIP3) GCMs are able to reproduce observed precipitation and temperature climatological statistics. The performance of each GCM in reproducing these statistics was ranked and better performing GCMs identified for later analyses. Observed global land surface precipitation and temperature data were drawn from the CRU 3.10 gridded dataset and re-sampled to the resolution of each GCM for comparison. Observed and GCM based estimates of mean and standard deviation of annual precipitation, mean annual temperature, mean monthly precipitation and temperature and Köppen climate type were compared. The main metrics for assessing GCM performance were the Nash–Sutcliffe efficiency index and RMSE between modelled and observed long-term statistics. This information combined with a literature review of the performance of the CMIP3 models identified the following five models as the better performing models for the next phase of our analysis in assessing the uncertainty in runoff estimated from GCM projections of precipitation and temperature: HadCM3 (Hadley Centre for Climate Prediction and Research), MIROCM (Center for Climate System Research (The University of Tokyo), National Institute for Environmental Studies, and Frontier Research Center for Global Change), MIUB (Meteorological Institute of the University of Bonn, Meteorological Research Institute of KMA, and Model and

HESSD
11, 4531–4578, 2014

Part 1: Assessment of Global Climate Models

T. A. McMahon et al.

Data group), MPI (Max Planck Institute for Meteorology) and MRI (Japan Meteorological Research Institute).

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. Our primary objective in this paper, the first of two complementary papers, is to assess how well 22 Global Climate Models (GCMs) from the World Climate Research Programme's (WCRP's) Coupled Model Intercomparison Project phase 3 (CMIP3) multi-model dataset (Meehl et al., 2007) are able to reproduce GCM grid-scale observed precipitation and temperature climatological statistics and to identify better performing models for use in our later analyses. We recognise that GCMs model different variables with a range of success and that no single model is best for all variables and/or for all regions (Lambert and Boer, 2001; Gleckler et al., 2008). The approach adopted here is not inconsistent with Dessai et al. (2005) who regarded the first step in evaluating GCM projection skill is to assess how well observed climatology is simulated.

Readers should note that when this project began in 2010, runs from the Coupled Model Intercomparison Project phase 5 (CMIP5) were not available. We are of the view that the methodologies developed in the project (and described here and in Peel et al., 2014) are equally applicable to evaluating CMIP5 simulations for uncertainty in future runoff and reservoir yield estimation. Conclusions about better performing models drawn from this analysis may prove similar to a comparable analysis of CMIP5 runs since most models in CMIP5 are, according to Knutti et al. (2013), "strongly tied to their predecessors". Analysis of the CMIP5 models indicates that the CMIP3 simulations are of comparable quality to the CMIP5 simulations for temperature and precipitation at regional scales (Flato et al., 2013).

HESSD
11, 4531–4578, 2014

A significant limitation to characterising the uncertainty of future runoff is the lack of sufficient GCM runs of historical (20C3M) and future projections (e.g., A1B). Two key sources of uncertainty in projections of future runoff are uncertainty between GCMs and within GCMs (Hawkins and Sutton, 2009, 2011). GCM runs available within the CMIP3 dataset, and being collated within the CMIP5 dataset, can be used to provide insight into the uncertainty between GCMs. However, in the CMIP3 dataset most GCMs have a single run or a very small number of runs for a given scenario, which severely limits any assessment of within GCM uncertainty. The purpose of the second paper is to develop and apply non-stationary stochastic replicates of GCM monthly precipitation and temperature time series for input to an off-line hydrologic model to estimate the variability in the mean and variance of annual runoff. In the absence of a large ensemble of runs (for example, 20 to 100 runs) from each GCM for each projection, this analysis provides an approximation of the uncertainty in projected runoff and also reservoir yield estimated from precipitation and temperature projections from a GCM. The methods developed and tested in the two complementary papers are based on GCM runs from the CMIP3 dataset. These methods will also be applicable to the CMIP5 dataset. In this paper and the complementary paper we use the terms streamflow and runoff synonymously, and adopt the unit of runoff as depth in mm rather than use volume units.

GCM runs for the observed period do not seek to replicate the observed monthly record at any point in time and space. Rather a better performing GCM is expected to produce long-term mean annual statistics that are broadly similar to observed conditions across a wide range of locations. Here, the assessment of CMIP3 GCMs is made by comparing their long-term mean annual precipitation (MAP), standard deviation of annual precipitation (SDP), mean annual temperature (MAT), mean monthly pattern of precipitation and temperature and Köppen climate type with concurrent observed data for 616 to 11 886 terrestrial grid cells world-wide (the number of grid cells depends on the resolution of the GCM under consideration). These variables were chosen to assess GCM performance because they provide insight into the mean annual,

HESSD
11, 4531–4578, 2014

Title Page
Abstract Introduction
Conclusions References
Tables Figures
◀◀ ▶▶
◀ ▶
Back Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

inter-annual variability and seasonality of precipitation and temperature, which are sufficient to estimate the mean and variability of annual runoff from a traditional monthly rainfall–runoff model (Chiew and McMahon, 2002) and from a top-down annual rainfall–runoff model (McMahon et al., 2011).

The GCMs included in this assessment are detailed in Table 1. Model acronyms adopted are also listed in the table. Although no quantitative assessment of the BCCR model is made, this model is included in Table 1 as details of its performance are available in the literature which is discussed in Sect. 2. Other details in the table include the originating group for model development, country of origin, model name given in the CMIP3 documentation (Meehl et al., 2007), the number of 20C3M runs available for analysis, the model resolution, and the number of terrestrial grid cells used in the precipitation and temperature comparisons.

Following this introduction we describe, and summarise in the next section, approaches to estimating runoff and potential evapotranspiration and several previous assessments of CMIP3 GCM performance. We also include some general comments on GCM assessment followed by particular assessments of CMIP3 models at the global scale. In Sect. 3, data (observed and GCM based) used in the analysis are described. Details and results of the subsequent analyses comparing GCM estimates of present climate mean and standard deviation of annual precipitation, mean annual temperature, mean monthly precipitation and temperature patterns and Köppen climate type against observed data are set out in Sect. 4. In Sect. 5 we review the results and compare the literature information with our assessments of the GCMs. The final section of the paper presents several conclusions.

2 Literature

To assess the impact of climate change on surface water resources of a region, it is necessary to determine, as a minimum, how the mean and variability of annual streamflows will be affected. Other factors of less importance are changes in the

HESSD
11, 4531–4578, 2014

auto-correlation of annual streamflow, changes in net evaporation from reservoir water surfaces, and changes in monthly flow patterns, with the latter being more important for relatively small reservoirs. In this paper we deal with the key drivers of streamflow production namely the mean and the standard deviation of annual precipitation and mean annual temperature, the latter is adopted here as a surrogate for potential evapotranspiration, along with a secondary factor, the mean monthly pattern of precipitation and temperature. In addition to understanding the conversion of precipitation and evapotranspiration to streamflow, in this introductory review we examine previous assessments of GCM performance and how they were carried out.

2.1 Estimating streamflow

At least two approaches are available to assess the impact of climate change on the mean and standard deviation of annual streamflow. The first approach uses a rainfall–runoff model calibrated against observed data for the catchment under consideration. Projections of future precipitation and evapotranspiration from a GCM are used as input to the model to produce time sequences of annual runoff from which the mean and variability of annual streamflows can be computed (Xu, 1999). In the second approach, analytical equations are developed that relate the mean and the variance of streamflow to a functional relationship of climate. Future changed climate conditions are then used to estimate a new mean and variance of streamflow. McMahon et al. (2011) developed two simple equations to estimate, respectively, the mean and the standard deviation of annual streamflow using an aridity index (mean annual potential evapotranspiration (PET) divided by mean annual precipitation) and the variances and the covariance of precipitation and potential evapotranspiration.

Computer models of most water resource systems that rely on surface reservoirs to offset streamflow variability adopt a monthly time-step to ensure that seasonal patterns in demand and reservoir inflows are adequately accounted for. However, in a climate change scenario it is more likely that an absolute change in streamflow will have a greater impact on system yield than shifts in the monthly inflow or demand patterns.

HESSD
11, 4531–4578, 2014

T. A. McMahon et al.

This will certainly be the case for reservoirs that operate as carryover systems rather than as within-year systems (for an explanation see McMahon and Adeloye, 2005). Therefore, in this paper we assess the GCMs in terms of annual precipitation and annual temperature, the latter as a surrogate for potential evapotranspiration, and mean monthly precipitation and temperature patterns.

2.2 Potential evapotranspiration

Irrespective of which approach is adopted to estimate future streamflow (a rainfall–runoff model or an analytical equation), an estimate of PET will be required. In this section we discuss briefly how PET can be estimated within the framework of GCM projections of climate variables. A more complete discussion is provided in Appendix A.

The authors appreciate the discussions regarding the observation that the magnitude of evaporation from Class-A pans has decreased over the past several decades during which time annual temperatures have risen (Roderick et al., 2009a). Much of the observed decline in pan evaporation has been attributed to declines in radiation and/or wind speed (Roderick et al., 2009b). The apparent paradox between decreasing pan evaporation and increasing temperature highlights the complex nature of evaporation and the potential danger of adopting an overly simple model for estimating PET. Roderick et al. (2009b) warn against using temperature only PET estimates for climate change studies as they would suggest that rising temperature would lead to rising evaporative demand; the opposite of what has been observed from pan data recently.

A second consideration is how complex does the PET model need to be for satisfactory hydrologic modelling? Andréassian et al. (2004) and Oudin et al. (2005a, b) addressed this issue and concluded a complex estimate of PET is not necessary for successful hydrologic modelling in catchments that are not strongly water or energy limited on an annual basis. Hydrologic models are capable of extracting the PET information they require from simple temperature based PET estimates during calibration. This second consideration is especially pertinent when the inputs to the PET model are from GCM projections. Projections of temperature are considered more reliable

HESSD
11, 4531–4578, 2014

than those of other variables like net radiation at the evaporating surface, wind speed and relative humidity required for Penman based PET estimates. Based on a brief review, see Appendix A, we conclude that adopting a complex PET formulation based on the Penman–Monteith equation (Monteith, 1965) may not improve our hydrologic modelling of climate change projections. A more complex PET formulation requires additional GCM variables other than temperature which are less reliable. Here we use temperature to represent PET, which requires fewer resources and will likely replicate current and future conditions more satisfactorily. However, this simplicity comes at the expense of inadequate representation of important future changes in PET, which may have important negative consequences when modelling streamflow in energy limited catchments. Nevertheless, in the following discussion we concentrate on mean annual temperature as the GCM variable representing PET.

2.3 Assessing GCM performance

Ever since the first GCM was developed by Phillips (1956) (see Xu, 1999), attempts have been made to assess the adequacy of GCM modelling. Initially, these evaluations were simple side by side comparisons of individual monthly or seasonal means or multi-year averages (Chervin, 1981). Chervin (1981) compared GCM modelled climate with observed data using the ensemble mean and standard deviation of several climate elements to assess model performance. Legates and Willmott (1992) compared observed with simulated average precipitation rates by 10° latitude bands. On a two-dimensional plot, Taylor (2001) developed a diagram in which each point consisted of the correlation coefficient and the root mean square (RMS) along with the ratio of the variances of the modelled and the observed variables. Recently, many authors have used the Taylor diagram (Covey et al., 2003; Bonsal and Prowse, 2006) or a similar approach (Lambert and Boer, 2001; Boer and Lambert, 2001). Murphy et al. (2004) introduced a Climate Prediction Index (CPI) which is based on a broad range of present-day climates. This index was later used by Johns et al. (2006) for a different set of climate variables than those used by Murphy et al. (2004). Whetton et al. (2005) introduced

HESSD
11, 4531–4578, 2014

a demerit point system in which GCMs were rejected when a specified threshold was exceeded. Min and Hense (2006) introduced a Bayesian approach to evaluate GCMs and argued that a skill-weighted average with Bayes factors is more informative than moments estimated by conventional statistics. Shukla et al. (2006) suggested that differences in observed and GCM simulated variables should be examined in terms of their probability distributions rather than individual moments. They proposed the differences could be examined using relative entropy. Perkins et al. (2007) also claimed that assessing the performance of a GCM through a probability density function (PDF) rather than using the first or a second moment would provide more confidence in the model assessment. To compare the reliability of variables (in time and space) rather than individual models, Johnson and Sharma (2009a, b) developed the Variable Convergence Score which is used to rank the variable based on the ensemble coefficient of variation. They observed the variables with the highest scores were pressure, temperature, and humidity. Reichler and Kim (2008) introduced a Model Performance Index by first estimating a normalised error variance based on the square of the grid-point differences between simulated (interpolated to the observational grid) and the observed annual climate weighted for area and mass averaging and standardised with respect to the variance of the annual observations. The error variance was scaled by the average error found in the reference models and, finally, averaged over all climates. It is clear from this brief review that no one procedure has been universally accepted to assess GCM performance, which is consistent with the observations of Räisänen (2007). We also note the comments of Smith and Chandler (2010, p. 379) who said: "It is fair to say that any measure of performance can be subjective, simply because it will tend to reflect the priorities of the person conducting the assessment. When different studies yield different measures of performance, this can be a problem when deciding on how to interpret a range of results in a different context. On the other hand, there is evidence that some models consistently perform poorly, irrespective of the type of assessment. This would tend to indicate that these model results suffer from fundamental errors which render them inappropriate."

HESSD
11, 4531–4578, 2014

2.4 Assessing CMIP3 GCMs

Räisänen (2007) provided a detailed review of the performance of 21 CMIP3 GCMs in terms of surface air temperature, precipitation and mean sea level pressure by estimating the root mean square error (RMSE) and the spatial correlation. His results, which are reproduced in Table 2, illustrate the wide range of model performances that exist especially for precipitation and mean sea level pressure.

Reichler and Kim (2008) considered 14 variables covering mainly the period 1979–1999 to assess the performance of CMIP3 models using their Model Performance Index. They concluded that there was a continuous improvement in model performance from the CMIP1 models compared to those available in CMIP3 but there are still large differences in the CMIP3 models ability to match observed climates. Gleckler et al. (2008) assessed 22 CMIP3 models using the relative error of each of 22 climate variables and concluded that some models performed substantially better than others. However, they also concluded that it is not yet possible to answer the question: what is the best model? Macadam et al. (2010) assessed the performance of 17 CMIP3 GCMs comparing the observed and modelled temperatures over five 20 year periods and concluded that GCM rankings based on actual temperatures can be consistent over time, although Reifen and Toumi (2009) observed that anomalies relative to climate means can be inconsistent over time.

In summary, Gleckler et al. (2008) stated that the best GCM will depend on the intended application. In the overarching project of which this study is a component, we are interested in the uncertainty in annual streamflow estimated from hydrologic modelling using GCM precipitation and temperature and how that uncertainty will affect estimates of future yield from surface water reservoir systems. Consequently, we are interested in which GCMs reproduce precipitation and temperature well. Based on the references of Reichler and Kim (2008), Gleckler et al. (2008) and Macadam et al. (2010), the performance of 23 CMIP3 GCMs assessed at a global scale are ranked in Table 3. In the table eight models that meet the Reichler and Kim (2008) criterion are also ranked in the upper 50 % based on the Macadam et al. (2010) and Gleckler et al. (2008) references. These models are CCCMA-t47, CCSM, GFDL2.0, GFDL2.1, HadCM3, MIROCM, MPI and MRI.

3 Data

Two data sets are used in the GCM assessment that follows in Sect. 4. One is based on observed data and the other on GCM simulations of present climate (20C3M). It should be noted that of the 22 GCMs examined herein, multiple runs or projections were available for nine models. The resulting 46 runs are identified in the tables summarising the results.

The first data set is based on monthly observed precipitation and temperature gridded at 0.5° × 0.5° resolution over the global land surface from CRU 3.10 (New et al., 2002) for the period January 1950 to December 1999. For grid cells where monthly observations are not available, the CRU 3.10 dataset is based on interpolation of observed values within a "correlation decay distance" of 450 km for precipitation and 1200 km for temperature. The CRU 3.10 data set provides information about the number of observations within the correlation decay distance of each grid cell for each month. In this analysis we defined a grid cell as "observed" if ≥ 90 % of months at that grid cell has at least one observation within the correlation decay distance for the period January 1950 to December 1999. Only "observed" grid cells are used to compute summary statistics in the following analysis.

The second data set is monthly precipitation and temperature data for the present climate (20C3M) from 22 of the 23 GCMs listed in Table 1 and consists of 46 GCM runs. The 20C3M monthly data for precipitation and temperature were extracted from the CMIP3 dataset. As shown in Table 1 the GCMs have a wide range of spatial resolutions, all of which are coarser than the observed CRU data. In order to make comparisons between observed and GCM data either the CRU and/or GCM data must be re-sampled to the same resolution. To avoid re-sampling coarse resolution data to

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

HESSD

11, 4531–4578, 2014

a finer resolution we only re-sampled the CRU data here. Thus in the following analysis the performance of each GCM is assessed at the resolution of the GCM and the CRU data are re-sampled to match the GCM resolution. Therefore, the number of grid cells in each comparison varies with the GCM resolution and ranged from 616 to 11 886 for the temperature comparisons and 425 to 8291 for the precipitation comparisons. The difference in number of grid cells between temperature and precipitation is due to more terrestrial grid cells having observed temperature data than precipitation data over the period 1950–1999.

In the following analysis comparisons are made between observed and GCM values of mean and standard deviation of annual precipitation and mean annual temperature. The GCM values are based on *concurrent raw* (that is, not downscaled or bias corrected) data from the 20C3M simulation. For example, if a grid cell has observed calendar-year data from 1953–1994, then the comparison is made with GCM values from the 20C3M run for the concurrent calendar years 1953–1994. Although the aim of a 20C3M run from a given GCM is not to strictly replicate the observed monthly record, we expect better performing GCMs to reproduce mean annual statistics that are broadly similar to observed conditions. Average monthly precipitation and temperature patterns are also compared to assess how well GCM runs reproduce observed seasonality. Finally, we assess how well the Köppen–Geiger climate classification (Peel et al., 2007) estimated from the CMIP3 data compares with present-day gridded observed climate classification.

4 Comparison of present climate GCM data with observed data

In the analyses that follow, GCM estimates of mean annual precipitation and temperature and the standard deviation of annual precipitation are compared against observed estimates for terrestrial grid cells with ≥ 90 % observed data during the period 1950–1999.

HESSD
11, 4531–4578, 2014

Eight standard statistics – Nash–Sutcliffe efficiency (NSE) (Nash and Sutcliffe, 1970), product moment coefficient of determination (R^2) (MacLean, 2005), standard error of regression (Maidment, 1992), bias (MacLean, 2005), percentage bias (Maidment, 1992), absolute percentage bias (MacLean, 2005), root mean square error (RMSE) (MacLean, 2005) and mean absolute error (MacLean, 2005) – were computed as the basis of comparison but we report only the NSE, R^2 and RMSE in the following discussion. For our analysis, the NSE is the most useful statistic as it shows the proportion of explained variance relative to the 1 : 1 line in a comparison of two estimates of the same variable. R^2 is included because many analysts are familiar with its interpretation. Both NSE and R^2 were computed in arithmetic (untransformed) and natural log space. We have also included RMSE values (computed from the untransformed values) as many GCM analyses include this measure.

In the following sub-sections comparisons between the concurrent raw GCM data and observed values for MAP, SDP, MAT, long-term average monthly precipitation and temperature patterns and Köppen climate classification (Peel et al., 2007) at the grid cell scale are presented and discussed. Our purpose for these comparisons is to select a small number of GCMs for later analysis. Although we rank the models by each selection criteria and combine the ranks by addition, we note the warning of Stainforth et al. (2007) who argue that model response should not be weighted but ruled in or out. We follow this approach in this paper by identifying five GCMs to be used in the analysis reported in the second paper. This approach is consistent with the concept recognised by Whetton et al. (2007, p. 1) that "the closer the current climate simulation of a model is to the observed climate, the closer the enhanced greenhouse response of that model is to the real world response". Furthermore, our comparisons are conducted over the terrestrial land surface, rather than focussing on a single catchment, region or continent. This allows us to assess whether a GCM performs consistently well across a large area and reduces the chance of a GCM being selected due to a random high performance over a small area.

HESSD
11, 4531–4578, 2014

4.1 Mean annual precipitation

Comparisons of mean annual precipitation and the standard deviation of annual precipitation between GCM estimates and observed data for the grid cells across the 46 runs are presented in Table 4. For MAP, the Nash–Sutcliffe efficiency varied from a maximum of 0.68 ($R^2 = 0.69$) with a RMSE value of 335 mm yr^{-1} for model MIUB(3) to −0.54 for GISS-EH3. (GCM run number is enclosed by parenthesis, for example MIUB(3) is run 3 for the GCM MIUB.) The MAP values for MIUB(3) are compared with the observed CRU MAP values in Fig. 1. Each data point in this figure represents a MAP comparison at one of the 632 MIUB(3) terrestrial grid cells where observed CRU 3.10 data were available for the period January 1950 to December 1999. The relationship between GCM and observed MAP shown in this figure is representative of the other GCMs where high MAPs are underestimated and low MAPs are overestimated.

The range of NSE values for the MAP comparisons across the 46 GCM runs is plotted in Fig. 2. The results may be classified into four groups: 5 runs exhibiting NSE > 0.6, 27 runs 0.4 < NSE ≤ 0.6, 6 runs 0 < NSE ≤ 0.4 and 8 runs ≤ 0 where the predictive power of the GCM is less than using the average observed MAP across all grid cells (Gupta et al., 2009).

4.2 Standard deviation of annual precipitation

For the standard deviation of annual precipitation, HadCM3 was the best performing model with a NSE of 0.57, R^2 of 0.62 and a RMSE of 51 mm yr^{-1}. MIROCH also yielded a NSE of 0.57 and an R^2 of 0.58 but with a RMSE of 63 mm yr^{-1}. These results along with other standard deviation values are listed in Table 4. Figure 3 is a plot for MIUB(3), which is representative (rank 4, 4th best performance of the 46 runs) of the relationship between GCM and observed SDP, and shows the model underestimates the standard deviation of annual precipitation for high values and overestimates at low values of standard deviation compared with observed values.

HESSD
11, 4531–4578, 2014

4.3 Mean annual temperature

The comparison of the GCM mean annual temperatures with concurrent observed data for the grid cells are listed for each model run in Table 4. In contrast to the precipitation modelling, the mean annual temperatures are simulated satisfactorily by most of the GCMs. Except for the IAP and the GFLD2.0 models (NSE =∼ 0.90 and 0.93 respectively), all model runs exhibit NSE values ≥ 0.94 with 17 of the 46 GCM runs having a NSE value ≥ 0.97. A comparison between MIUB(3) estimates of mean annual temperature (NSE = 0.96, rank 33) and observed values from the CRU data set are presented in Fig. 4. Also shown in Fig. 4 is a linear fit between GCM and observed MAT. The average fit for the 46 GCM runs (not shown) exhibited a small negative bias of −1.03 °C and a slope of 1.01.

4.4 Average monthly precipitation and temperature patterns

Since a monthly rainfall–runoff model is applied in the next phase of our analysis (second paper) it was considered appropriate to assess how well the GCMs simulate the observed mean monthly patterns of precipitation and temperature (see also the argument of Charles et al., 2007). The NSE was used for the assessment by comparing the 12 long-term average monthly values. For each GCM run the average precipitation and temperature for each month were calculated for each grid cell. Nash–Sutcliffe efficiencies were computed between the equivalent 12 GCM and 12 CRU based monthly averages. The median NSE values across terrestrial grid cells where observed CRU 3.10 data were available for the period January 1950 to December 1999 for each GCM run are summarised in Table 4. As shown in Table 4 average monthly patterns of precipitation are poorly modelled. In fact, 57 % of the 46 model runs have a median NSE value of < 0. For these GCMs their predictive power for the monthly precipitation pattern is less than using the average of the 12 monthly values at each of the terrestrial grid cells. Only two GCMs have NSE values > 0.25. In contrast, the median NSEs of all monthly temperature patterns are > 0.75, with 41 % > 0.90. The NSE metric reflects

HESSD
11, 4531–4578, 2014

how well the GCM replicates both the monthly pattern and the overall average monthly value (bias). Thus the monthly pattern of temperature is generally well reproduced by the GCMs, whereas the monthly pattern of precipitation is not, which is mainly due to the bias in average monthly precipitation.

4.5 Köppen classification

The Köppen climate classification (see Table 5) provides an alternate way to assess the adequacy of how well a GCM represents climate as the classification is based on a combination of annual and monthly precipitation and temperature data. Two comparisons between the MPI(3) model and CRU observed data are presented in Table 6. The MPI(3) was chosen as an example here as over the three levels of climate classes it estimated the observed climate correctly more often than the other model runs. In Table 6a a comparison at the first letter level of the Köppen climate classification is shown. This comparison reveals how well the GCM reproduces the distribution of broad climate types: tropical, arid, temperate, cold and polar over the terrestrial surface. In Table 6b the comparison shown is for the second letter level of the Köppen climate classification, which assesses how well the GCM reproduces finer detail within the broad climate types; for example, the seasonal distribution of precipitation or whether a region is semi-arid or arid. The diagonal values shown in Table 6a and b represent the number of grid cells correctly classified by the GCM and the off-diagonal values are incorrectly classified by the GCM for the one- and two-letter level respectively. At the first letter level MPI(3) reproduces the correct climate type at 81 % of the terrestrial grid cells. Within this good performance the MPI(3) produces more polar climate and less tropical and cold grids cells than observed. At the second letter level, MPI(3) reproduces the correct climate type at 67 % of the terrestrial grid cells. The model produces less grid cells of tropical rainforest, cold with a dry winter and cold without a dry season than expected and more cold with a dry summer and polar tundra than expected.

Table 7 summarises the overall proportion of GCM grid cells that were classified correctly for each GCM run across the three levels of classification. As we wish to have

HESSD
11, 4531–4578, 2014

a ranking of the comparisons we adopted this simple measure as it is regarded as "...one of the most basic and widely used measures of accuracy..." for comparing thematic maps (Foody, 2004, page 632). From Table 7 we observe that GCM accuracy in reproducing the climate classification decreases as one moves from coarse to fine detail climate classification. The average accuracy (and range) for the three classes are: 0.48 (0.36–0.60) for the three-letter classification, 0.57 (0.47–0.68) for the two-letter classification, and for one-letter 0.77 (0.66–0.82). In other words, at the three-letter scale nearly 50 % of GCM Köppen estimates are correct, increasing to nearly 60 % at the two-letter level and, finally, at the one-letter aggregation more than 75 % are correct across the 46 GCM runs. Across the three classes, the following five models performed satisfactorily in identifying Köppen climate class correctly: HadCM3, HadGEM, MIUB, MPI, and MRI. Of these five models the least successful run was for MIUB(3) with the percentage correct for each class being: 1-letter 78 %, 2-letter 61 % and 3-letter 52 %.

5 Discussion

5.1 Relating GCM resolution to performance

In the analysis presented in the previous section each GCM's performance in reproducing observed climatological statistics was assessed at the resolution of the individual GCM. The question of whether GCMs with a finer resolution outperform GCMs with a coarser resolution is addressed in Fig. 5 where GCM performance in reproducing observed terrestrial MAP and MAT, based on the NSE, is related to GCM resolution, defined as the number of grid cells used in the comparison. The plot suggests there is no significant relationship between GCM resolution and GCM performance beyond 1500 grid cells for either MAP or MAT. Interestingly some lower resolution GCMs, < 1500 grid cells, perform as well as higher resolution GCMs for MAP and MAT, yet for others, they perform poorly. While it is sometimes assumed that higher resolution should normally lead to improved performance, there are many other factors that affect performance.

HESSD
11, 4531–4578, 2014

Part 1: Assessment of Global Climate Models

T. A. McMahon et al.

These include the sophistication of the parameterisation schemes for different sub-grid scale processes, the time spent in developing and testing the individual schemes and their interactions. Hence, as found here, some better developed lower resolution models can have equal or better performance than higher resolution models.

5.2 Joint comparison of precipitation and temperature

In using GCM climate scenarios in a water resources study, it is appropriate to ensure consistency between precipitation and temperature by adopting projections of these variables from the same GCM run. Grid cell based NSEs for mean annual temperature and mean annual precipitation from each GCM are compared in Fig. 6, which illustrates the performance of each GCM for both variables. Models that have relatively high NSEs for precipitation do not necessarily have relatively high values for temperature. It is interesting to note that the rank of the models based on NSE of the MAP is unrelated to the ranking of the models based on MAT. Fortunately, however, most of the NSEs for MAT are relatively high and the acceptance or rejection of a GCM as a better performing model is largely dependent on its precipitation characteristics.

5.3 Identifying better performing GCMs

To identify the better performing GCMs across the different variables assessed, the results in Table 4 are ranked by Nash–Sutcliffe efficiency and summarised in Table 8. The overall rank for each GCM run is based on combining, by addition, the ranks for the individual variables and, finally, identifying the best performing run from each GCM. Selection of the better performing GCMs using these rankings is not inconsistent with Stainforth et al. (2007) who argued that model response should not be weighted but ruled in or out. From Table 8 we identify several GCMs, listed in Table 9, as better performing models for use in our companion paper. These GCMs were selected based on the assumption that performance across the four variables (MAP, SDP, MAT and monthly patterns) is equally weighted. GCMs that achieved MAP NSE > 0.50, SDP

HESSD
11, 4531–4578, 2014

NSE > 0.45, MAT NSE > 0.95 and mean monthly pattern of precipitation NSE > 0.0 were identified as better performing. The following GCMs were selected: HadCM3, INGV, MIROCM, MIUB, MPI, and MRI and are listed in Table 9. INGV was included although it failed the monthly precipitation pattern criterion. These criteria were selected to identify a small number of GCMs that would require less bias correction to produce annual precipitation and temperature consistent with observations.

5.4 Combining literature and our assessments

In Table 9, we summarise our observations from the literature review in Sect. 2 and the results from our analyses in Tables 4 and 8, where we identified six GCMs that satisfied our selection criteria (Table 9, column 1). From the literature review (Table 3), eight GCMs were identified as being satisfactory. We have added MIUB because in the literature review it ranked first overall, although no guidance was available from Reichler and Kim (2008). We also added MIROCH to this list as it performed better according to Gleckler et al. (2008) than several models in the above list and met the performance index of Reichler and Kim (2008). Columns (1) and (2) of Table 9 suggest there is some consistency between our analyses and that reported in the literature. From the table, we identify that, in terms of our objective to assess how well the CMIP3 GCMs are able to reproduce observed annual precipitation and temperature statistics and the mean monthly patterns of precipitation and temperature, the following models are deemed acceptable for the next phase of our project: HadCM3, MIROCM, MIUB, MPI and MRI. Although not used in the selection criteria we note our selected GCMs performed well in the Köppen climate assessment. We note here that INGV also performed satisfactorily but was not included in our adopted GCMs as it was not reviewed in the papers of Gleckler et al. (2008), Reichler and Kim (2008) and Macadam et al. (2010).

HESSD
11, 4531–4578, 2014

5.5 Grid cell based comparison and catchment analysis

In this section we compare results from the GCM vs. observed data comparison at grid cell scale presented in the previous section with results from a similar, but not shown, comparison at the catchment scale. When assessing which GCMs to adopt for a climate change impact assessment usually only grid cell based comparisons are available from the literature. However, the aim of this project is to investigate uncertainty in runoff (streamflow) from catchments. Therefore, it is of interest to know how GCM performance at the grid cell scale aligns with GCM performance at the catchment scale for a range of catchment scales world-wide.

Using the unimpaired catchment data set of Peel et al. (2010) we assessed GCM performance in reproducing observed long-term MAP, SDP and MAT at 699 catchments world-wide. We have not reported the results of this catchment comparison here because many catchments in our data set are smaller than a GCM grid cell and, therefore, the comparison is not strictly appropriate. When we limited the comparison to catchments with larger areas the number and spatial distribution of available catchments reduces considerably. Overall, the catchment based GCM vs. observed results are similar to the grid cell based results. GCM performance in reproducing long-term MAP, SDP and MAT, in terms of NSE, was generally higher for the grid cell based comparison than the catchment comparison. Generally, the better performing GCMs performed well at grid cell scale and catchment scale. This is seen in a comparison of GCM rank (Fig. 7) where the GCM rank from the grid cell analysis (Table 8) is largely consistent with the GCM rank from the catchment analysis (not shown).

6 Conclusions

This is the first of two complementary papers which together have the overall objective of enhancing our understanding of the uncertainty of future annual streamflows and surface reservoir performance based on GCM projections of future climate. In this

HESSD
11, 4531–4578, 2014

paper we assess which are the better performing GCMs at reproducing observed climatological statistics for use in the second paper where the uncertainty due to within GCM variability is approximated for future streamflows. The GCM selection process was informed by our results presented here and a literature review of better performing GCMs. In terms of the Nash–Sutcliffe efficiency (NSE) there was a large spread in variance explained by the GCMs for observed mean annual precipitation and the standard deviation of annual precipitation over concurrent periods. The highest NSE for mean annual precipitation was 0.68 and 0.52 for the standard deviation of annual precipitation. On the other hand, for mean annual temperatures, the NSEs between modelled and observed data were very high, with median NSE being 0.96. Overall, all GCMs reproduced the Köppen climate satisfactorily at the broad first letter level. From the literature, the following GCMs were identified as being suitable to simulate annual precipitation and temperature statistics: CCCMA-T47, CCSM, GFDL2.0, GFDL2.1, HadCM3, MIROCH, MIROCM, MIUB, MPI and MRI. When combining our results with the literature the following GCMs were considered the better performing models for the next phase of the analyses to be reported in the complementary paper: HadCM3, MIROCM, MIUB, MPI and MRI.

Appendix A

Estimating potential evapotranspiration for climate change impact assessments

Projected changes in water and energy at the catchment scale are the fundamental basis of all hydrologic climate change impact assessments. Hydrologic models require time series of precipitation and, usually, potential evapotranspiration to represent the interaction of water and energy within a catchment. Therefore, for hydrologic climate change impact assessments, an estimate of potential evapotranspiration (PET) is required. For the practitioner the question is which PET method to adopt? Here we briefly review three questions that influence the choice of PET equation: (1) does the equation represent all relevant processes; (2) what PET information does a hydrologic model actually use; and (3) are future projections of variables used to estimate PET reliable?

A1 Does the PET equation represent all relevant processes?

McMahon et al. (2013) discuss a range of PET equations used in rainfall–runoff modelling. Frequently adopted methods to represent PET include Penman (Penman, 1948), Penman–Monteith (Monteith, 1965), FAO reference crop (Allen et al., 1998), Morton (Morton, 1983) and pan evaporation data. Ideally to represent future PET conditions the method adopted should adequately capture all changes in the energy and aerodynamic components of the evaporative process.

The potential danger of using a PET equation that does not adequately represent all relevant processes is highlighted by recent trends in pan evaporation data. Over the past several decades the magnitude of evaporation from Class-A pans has decreased (between -1 to $-4\,\mathrm{mm\,year^{-2}}$) while at the same time annual temperatures have risen (Roderick et al., 2009a). Roderick et al. (2009b) warn against using temperature only PET estimates for climate change studies as they would suggest that rising temperature would lead to rising evaporative demand; the opposite of what has been observed from pan data recently. Roderick et al. (2009b) attribute much of the observed decline in pan evaporation to declines in radiation and/or wind speed. Donohue et al. (2010), using the Penman formulation and gridded Australian data (1981–2006), attributed increasing surface temperature with contributing $+1.5\,\mathrm{mm\,year^{-2}}$ toward evaporative demand. However, the temperature contribution was more than offset by negative contributions from changes in wind speed ($-1.3\,\mathrm{mm\,year^{-2}}$), net radiation ($-0.6\,\mathrm{mm\,year^{-2}}$) and actual vapour pressure ($-0.4\,\mathrm{mm\,year^{-2}}$) to give an overall decrease in evaporative demand of $-0.8\,\mathrm{mm\,year^{-2}}$. Donohue et al. (2010) also compared the performance of five formulations of differing complexity namely Thornthwaite (Thornthwaite, 1948, Priestley–Taylor (Priestley and Taylor, 1972), Morton point and areal (Wang et al., 2001) and Penman, 1948) and preferred Penman, the most complex form, based on its ability to best capture the dynamics of evaporative demand. Overall, Roderick et al. (2009a, b), Chen et al. (2005) and Hobbins et al. (2008) conclude that PET estimates based only on T are problematic, particularly in energy limited environments (cold and polar climates), for climate change studies.

A2 What PET information does a hydrologic model actually use?

Whether hydrologic models require, or make use of, detailed PET data was assessed by Andréassian et al. (2004) and Oudin et al. (2005a, b). They found that hydrologic models perform as well (if not better) with mean monthly estimates of PET, or with temperature based estimates of PET, rather than time varying estimates of PET or more complex Penman based PET (Penman, 1948; Allen et al., 1998). Catchments used in their studies were located in France (Andréassian et al., 2004; Oudin et al., 2005a, b), USA (Oudin et al., 2005a, b) and Australia (Oudin et al., 2005a, b). The vast majority of their catchments have a temperate climate (not strongly water or energy limited on an annual basis). Under these conditions the hydrologic models appear to be largely insensitive to the complexity of the PET data used to drive them. During calibration conceptual hydrologic models are flexible enough to extract the PET information they need from whichever PET data (simple or complex) are used (see Chapman, 2003). Thus, as long as PET estimates are broadly correct in terms of seasonal pattern and annual mean and the hydrologic model was calibrated on that PET data then model performance is likely to be acceptable. For example, Oudin et al. (2005b) tested 27 PET formulations, of varying complexity, over 308 catchments using four daily conceptual models and proposed a simple temperature (mean daily temperature for a given Day-of-Year) and extra-terrestrial radiation (estimated from latitude and Day-of-Year) method that performed as well as the daily Penman method. In summary, a complex estimate of PET is not necessary for successful hydrologic modelling in catchments that are not strongly water or energy limited on an annual basis.

HESSD
11, 4531–4578, 2014

A3 Are future projections of variables used to estimate PET reliable?

In the previous two sections we have seen that a simple PET formulation may be good enough for hydrologic modelling, but not good enough to represent projected changes in PET. The final question relates to whether GCMs are able to provide reliable outputs on which to base a complex estimate of PET? Kay and Davies (2008) used IPCC third assessment report runs for 5 GCMs and 8 regional climate models driven by the HadGEM (Hadley Centre GCM) to calculate PET using Penman–Monteith and the temperature/radiation (T/R) method of Oudin et al. (2005b). They compared their two PET estimates derived from GCM data against observation based gridded values of Penman–Monteith PET for Britain. Overall, the GCM estimate of PET using T/R performed better than GCM Penman–Monteith at reproducing observed Penman–Monteith for all climate models. Future values of PET based on Penman–Monteith were also more variable than those based on T/R, which they suggest may reflect reliability issues with GCM variables, other than temperature, used to estimate Penman–Monteith. Kingston et al. (2009) also highlight reliability issues with GCM inputs to the Penman–Monteith equation. Although confidence in GCM-simulated temperature is generally high, Kingston et al. (2009, p. 4) note "less confidence can be placed in cloud cover and vapour pressure", which influence GCM-simulation estimates of net radiation at the evaporating surface and relative humidity. Overall, Kay and Davies (2008) suggest hydrologic modellers should be pragmatic and use as many GCMs as possible and estimate PET in a consistent way for any impact analysis.

A4 Summary

Ideally, estimates of PET should be based on methodologies that include all key evaporative processes to ensure future changes in PET are accurately represented. A Penman based equation is thus an ideal methodology to adopt. However, the reliability of future PET estimates is dependent on the reliability of GCM projections of input variables. For example, the Penman equation requires inputs of air temperature, net

HESSD
11, 4531–4578, 2014

radiation at the evaporating surface, wind speed and relative humidity. In this paper we have found that mean monthly and mean annual temperature are well reproduced by CMIP3 GCMs. However, reported confidence in GCM estimates of net radiation at the evaporating surface, wind speed and relative humidity is much lower. Therefore, although Penman based methodologies have the capacity to represent future trends due to changes in all key evaporative processes, GCM projections of those process variables, other than temperature, may be unrealistic. Thus at this time PET based on Penman may actually increase uncertainty in future PET, as seen in Kay and Davies (2008). PET based on Penman will be preferable once GCM projections of net radiation at the evaporating surface, wind speed and relative humidity become more reliable.

As GCM projections of temperature are considered reliable, here we adopt a simple PET formulation based only on temperature. A temperate based PET is likely to provide sufficient PET information for successful hydrologic modelling if the model is calibrated on that data, which is our intention in the companion paper. However, by adopting a temperature based PET estimate we acknowledge that the projected trend in PET will be an increase, when in reality the trend may increase or decrease due to changes in temperature, net radiation at the evaporating surface, wind speed and/or relative humidity. This error in PET trend is unlikely to be important for hydrologic modelling of water limited catchments, where changes in precipitation are the main driver of changes in runoff. However, in energy limited catchments, PET is a key driver of runoff and errors in PET trend will result in errors in runoff trend. For the 699 global catchments, described in Peel et al. (2010), used in the companion paper, ~ 30 % of total catchment area experience cold or polar climates that are likely to be energy-limited. The future trend in PET will be important to runoff for these catchments and using a temperature based PET estimate may cause errors in modelled runoff. In the remaining ~ 70 % of our total catchment area that experience tropical, arid and temperate climates, future runoff will be dominated more by changes in precipitation than PET as these areas are more water limited and errors in modelled runoff due to incorrect PET trends are unlikely.

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 Ian Smith assisted with extraction and analysis of CMIP3 GCM data. Lionel Siriwardena also assisted with extraction and analysis of the CRU 3.10 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, US Department of Energy.

References

Allen, R. G., Pereira, L. S., Raes, D., and Smith, M.: Crop evapotranspiration Guidelines for computing crop water requirements, FAO Irrigation and Drainage Paper 56, Food and Agriculture Organization of the United Nations, Rome, Italy, 1998.

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.

Boer, G. J. and Lambert, S. J.: Second order space–time climate difference statistics, Clim. Dynam., 17, 213–218, 2001.

Bonsal, B. T. and Prowse, T. D.: Regional assessment of GCM-simulated current climate over Northern Canada, Arctic, 59, 115–128, 2006.

Chapman, T. G.: Estimation of evaporation in rainfall–runoff models, MODSIM 2003 International Congress on Modelling and Simulation, vol. 1, 14–17 July 2003, Townsville, Australia, 148–153, 2003.

Charles, S. P., Bari, M. A., Kitsios, A., and Bates, B. C.: Effect of GCM bias on downscaled precipitation and runoff projections for the Serpentine catchment, Western Australia, Int. J. Climatol., 27, 1673–1690, 2007.

Chen, D., Gao, G., Xu C- Y., Guo, J., and Ren, G.: Comparison of the Thornthwaite method and pan data with the standard Penman–Monteith estimates of reference evapotranspiration in China, Clim. Res., 28, 123–132, 2005.

Chervin, R. M.: On the comparison of observed and GCM simulated climate ensembles, J. Atmos. Sci., 38, 885–901, 1981.

Chiew, F. H. S. and McMahon, T. A.: Modelling the impacts of climate change on Australian streamflow, Hydrol. Process., 16, 1235–1245, 2002.
Covey, C., Achutarao, K. M., Cubasch, U., Jones, P., Lambert, S. J., Mann, M. E., Phillips, T. J., and Taylor, K. E.: An overview of results from the Coupled Model Intercomparison Project, Global Planet. Change, 37, 103–133, 2003.
Dessai, S., Lu, X., and Hulme, M.: Limited sensitivity analysis of regional climate change probabilities for the 21st century, J. Geophys. Res., 110, D19108, doi:10.1029/2005JD005919, 2005.
Donohue, R. J., McVicar, T. R., and Roderick, M. L.: Assessing the ability of potential evaporation formulations to capture the dynamics in evaporative demand within a changing climate, J. Hydrol., 386, 186–197, 2010.
Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S. C., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., Forest, C., Gleckler, P., Guilyardi, E., Jakob, C., Kattsov, V., Reason, C., and Rummukainen, M.: Evaluation of climate models, in: Climate Change 2013: The Physical Science Basis, Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.
Foody, G. M.: Thematic map comparison: evaluating the statistical significance of differences in classification accuracy, Photogramm. Eng. Rem. S., 70, 627–633, 2004.
Gleckler, P. J., Taylor, K. E., and Doutriaux, C.: Performance metrics for climate models, J. Geophys. Res.-Atmos., 113, D06104, doi:10.1029/2007JD008972, 2008.
Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling, J. Hydrol., 377, 80–91, 2009.
Hawkins, E. and Sutton, R.: The potential to narrow uncertainty in regional climate predictions, B. Am. Meteorol. Soc., 90, 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.
Hobbins, M. T., Dai, A., Roderick, M. L., and Farquhar, G. D.: Revisiting the parameterization of potential evaporation as a driver of long-term water balance trends, Geophys. Res. Lett., 35, L12403, doi:10.1029/2008GL033840, 2008.

Johns, T. C., Durman, C. F., Banks, H. T., Roberts, M. J., Mclaren, A. J., Ridley, J. K., Senior, C. A., Williams, K. D., Jones, A., Rickard, G. J., Cusack, S., Ingram, W. J., Crucifix, M., Sexton, D. M. H., Joshi, M. M., Dong, B.-W., Spencer, H., Hill, R. S. R., Gregory, J. M., Keen, A. B., Pardaens, A. K., Lowe, J. A., Bodas-Salcedo, A., Stark, S., and Searl, Y.: The new Hadley Centre climate model (HadGEM1): evaluation of coupled simulations, J. Climate, 19, 1327–1353, 2006.

Johnson, F. M. and Sharma, A.: GCM Simulations of a Future Climate: How Does the Skill of GCM Precipitation Simulations Compare to Temperature Simulations, 18th World IMACS/MODSIM Congress, 13–17 July 2009, Cairns, Australia, 2618–2624, 2009a.

Johnson, F. M. and Sharma, A.: Measurement of GCM skill in predicting variables relevant for hydroclimatological assessments, J. Climate, 22, 4373–4382, 2009b.

Kay, A. L. and Davies, H. N.: Calculating potential evaporation from climate model data: a source of uncertainty for hydrological climate change impacts, J. Hydrol., 358, 221–239, 2008.

Kingston, D. G., Todd, M. C., Taylor, R. G., Thompson, J. R., and Arnell, N. W.: Uncertainty in the estimation of potential evapotranspiration under climate change, Geophys. Res. Lett., 36, L20403, doi:10.1029/2009GL040267, 2009.

Knutti, R., Masson, D., and Gettelman, A.: Climate model genealogy: generation CMIP5 and how we got there, Geophys. Res. Lett., 40, 1194–1199, 2013.

Lambert, S. J. and Boer, G. J.: CMIP1 evaluation and intercomparison of coupled climate models, Clim. Dynam., 17, 83–106, 2001.

Legates, D. R. and Willmott, C. J.: A comparison of GCM-simulated and observed mean January and July precipitation, Global Planet. Change, 5, 345–363, 1992.

Macadam, I., Pitman, A. J., Whetton, P. H., and Abramowitz, G.: Ranking climate models by performance using actual values and anomalies: implications for climate change impact assessments, Geophys. Res. Lett., 37, L16704, doi:10.1029/2010GL043877, 2010.

MacLean, A.: Statistical evaluation of WATFLOOD (Ms), Dept. of Civil & Environmental Engineering, University of Waterloo, Ontario, Canada, 2005.

Maidment, D. R.: Handbook of Hydrology, McGraw-Hill Inc., New York, USA, 1992.

McMahon, T. A. and Adeloye, A. J.: Water Resources Yield, Water Resources Publications, Denver, Colorado, USA, 220 pp., 2005.

HESSD
11, 4531–4578, 2014

McMahon, T. A., Peel, M. C., Pegram, G. G. S., and Smith, I. N.: A simple methodology for estimating mean and variability of annual runoff and reservoir yield under present and future climates, J. Hydrometeorol., 12, 135–146, 2011.
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.
Min, S.-K. and Hense, A.: A Bayesian approach to climate model evaluation and multi-model averaging with an application to global mean surface temperatures from IPCC AR4 coupled climate models, Geophys. Res. Lett., 33, L08708, doi:10.1029/2006GL025779, 2006.
Monteith, J. L.: Evaporation and environment, in: The State and Movement of Water in Living Organisms, edited by: Fogg, G. E., Symposium Society Experimental Biology, Cambridge University Press, London, 19, 205–234, 1965.
Morton, F. I.: Operational estimates of areal evapotranspiration and their significance to the science and practice of hydrology, J. Hydrol., 66, 1–76, 1983.
Murphy, J. M., Sexton, D. M. H., Barnett, D. N., Jones, G. S., Webb, M. J., Collins, M. J., and Stainforth, D. A.: Quantification of modelling uncertainties in a large ensemble of climate change simulations, Nature, 430, 768–772, 2004.
Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models Part 1 – A discussion of principles, J. Hydrol., 10, 282–290, 1970.
New, M., Lister, D., Hulme, M., and Makin, I.: A high-resolution data set of surface climate over global land areas, Clim. Res., 21, 1–25, 2002.
Oudin, L., Michel, C., and Anctil, F.: Which potential evapotranspiration input for a lumped rainfall–runoff model?, Part 1 – Can rainfall–runoff models effectively handle detailed potential evapotranspiration inputs?, J. Hydrol., 303, 275–289, 2005a.
Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., and Loumagne, C.: Which potential evapotranspiration input for a lumped rainfall–runoff model? Part 2 – Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling, J. Hydrol., 303, 290–306, 2005b.
Peel, M. C., Finlayson, B. L., and McMahon, T. A.: Updated world map of the Köppen-Geiger climate classification, Hydrol. Earth Syst. Sci., 11, 1633–1644, doi:10.5194/hess-11-1633-2007, 2007.

HESSD
11, 4531–4578, 2014

Part 1: Assessment of Global Climate Models

T. A. McMahon et al.

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:10.1029/2009WR008233, 2010.

Peel, M. C., Srikanthan, R., McMahon, T. A., and Karoly, D. J.: Uncertainty in runoff based on Global Climate Model precipitation and temperature data – Part 2: Estimation and uncertainty of annual runoff and reservoir yield, Hydrol. Earth Syst. Sci. Discuss., 11, 4579–4638, doi:10.5194/hessd-11-4579-2014, 2014.

Penman, H. L.: Natural evaporation from open water, bare soil and grass, Proc. R. Soc. Lon. Ser.-A, 193, 120–145, 1948.

Perkins, S. E., Pitman, A. J., Holbrook, N. J., and McAneney, J.: Evaluation of the AR4 climate models simulated daily maximum temperature, minimum temperature and precipitation over Australia using probability density functions, J. Climate, 20, 4356–4376, 2007.

Phillips, N. A.: The general circulation of atmosphere: a numerical experiment, Q. J. Roy. Meteor. Soc., 82, 123–164, 1956.

Priestley, C. H. B. and Taylor, R. J.: On the assessment of surface heat flux and evaporation using large scale parameters, Mon. Weather Rev., 100, 81–92, 1972.

Räisänen, J.: How reliable are climate models?, Tellus A, 59, 2–29, 2007.

Reichler, T. and Kim, J.: How well do coupled models simulate today's climate? B. Am. Meteorol. Soc., 89, 303–311, 2008.

Reifen, C. and Toumi, R.: Climate projections: past performance no guarantee of future skill?, Geophys. Res. Lett., 36, L13704, doi:10.1029/2009GL038082, 2009.

Roderick, M. L., Hobbins, M. T., and Farquhar, G. D.: Pan evaporation trends and the terrestrial water balance. I. Principles and observations, Geography Compass, 3, 746–760, 2009a.

Roderick, M. L., Hobbins, M. T., and Farquhar, G. D.: Pan evaporation trends and the terrestrial water balance. II. Energy balance and interpretation, Geography Compass, 3, 761–780, 2009b.

Shukla, J., DelSole, T., Fennessy, M., Kinter, J., and Paolino, D.: Climate model fidelity and projections of climate change, Geophys. Res. Lett., 33, L07702, doi:10.1029/2005GL025579, 2006.

Smith, I. and Chandler, E.: Refining rainfall projections for the Murray Darling Basin of south-east Australia – the effect of sampling model results based on performance, Clim. Change, 102, 377–393, 2010.

Stainforth, D. A., Allen, M. R., Tredger, E. R., and Smith, L. A.: Confidence, uncertainty and decision-support relevance in climate predictions, Philos. T. R. Soc. A, 365, 2145–2161, 2007.
Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, J. Geophys. Res., 106, 7183–7192, 2001.
Thornthwaite, C. W.: An approach toward a rational classification of climate, Geogr. Rev., 38, 55–94, 1948.
Wang, Q. J., Chiew, F. H. S., McConachy, F. L. N., James, R., de Hoedt, G. C., and Wright, W. J.: Climatic Atlas of Australia Evapotranspiration, Bureau of Meteorology, Commonwealth of Australia, 2001.
Whetton, P., McInnes, K. L., Jones, R. J., Hennessy, K. J., Suppiah, R., Page, C. M., and Durack, P. J.: Australian Climate Change Projections for Impact Assessment and Policy Application: a Review. CSIRO Marine and Atmospheric Research Paper 001, available at: www.cmar.csiro.au/e-print/open/whettonph_2005a.pdf (last access: 14 April 2014), 2005.
Whetton, P., Macadam, I., Bathols, J., and O'Grady, J.: Assessment of the use of current climate patterns to evaluate regional enhanced greenhouse response patterns of climate models, Geophys. Res. Lett., 34, L14701, doi:10.1029/2007GL030025, 2007.
Xu, C. Y.: Climate change and hydrologic models: a review of existing gaps and recent research developments, Water Resour. Manag., 13, 369–382, 1999.

HESSD
11, 4531–4578, 2014

Table 1. Details of 23 GCMs considered in this paper.

Acronym	Originating group	Country	Model name in CMIP3	Number of 20C3M runs available	Resolution		Number of precipitation grid cells[b]	Number of temperature grid cells[c]
					Lat (°)	Long (°)		
BCCR	Bjerkness Centre for Climate Research	Norway	bccr-bcm2.0	na[a]	1.9	1.9	na	na
CCCMA-t47	Canadian Centre for Climate Modeling and Analysis	Canada	cccma_cgm3_1_t47	1	~ 3.75	3.75	631	916
CCCMA-t63	Canadian Centre for Climate Modeling and Analysis	Canada	cccma_cgm3_1_t63	1	~ 2.8	2.8125	1169	1706
CCSM	National Centre for Atmospheric Research	USA	ccsm	8	~ 1.4	1.40625	5184	7453
CNRM	Météo-France/Centre National de Recherches Météorologiques	France	cnrm	1	~ 2.8	2.8125	1169	1706
CSIRO	Australia CSIRO	Australia	csiro_mk3_0	1	~ 1.87	1.875	2820	4068
GFDL2.0	NOAA Geophysical Fluid Dynamics Laboratory	USA	gfdl2_cm2_0	1	2	2.5	1937	2828
GFDL2.1	NOAA Geophysical Fluid Dynamics Laboratory	USA	gfdl2_cm2_1	1	~ 2	2.5	1911	2758
GISS-AOM	NASA Goddard Institute of Space Studies	USA	giss_aom_r1, 2	2	3	4	754	1076
GISS-EH	NASA Goddard Institute of Space Studies	USA	giss_eh1, 2,3	3	3 and 4	5	425	616
GISS-ER	NASA Goddard Institute of Space Studies	USA	giss_model_e_r	3	3 and 4	5	425	616
HadCM3	Hadley Centre for Climate Prediction and Research	UK	hadcm3	1	2.5	3.75	982	1421
HadGEM	Hadley Centre for Climate Prediction and Research	UK	HadGem	1	1.25	1.875	4316	6239
IAP	Institute of Atmospheric Physics, Chinese Acad. Sciences	China	iap_fgoals1.0_g	3	6.1 ~ 2.8	2.8125	1159	1664
INGV	National Institute of Geophysics and Vulcanology, Italy	Italy	ingv20c ECHAM4.6	1	~ 1.1	1.125	8291	11 886
INM	Institute for Numerical Mathematics, Russia	Russia	inmcm3.0	1	4	5	420	620
IPSL	Institute Pierre Simon Laplace	France	ipsl_cm4	1	~ 2.5	3.75	980	1403
MIROCH	Center for Climate System Research (The University of Tokyo), National Institute for Environmental Studies, and Frontier Research Center for Global Change	Japan	miroc3_2_hires (mirochi)	1	~ 1.1	1.125	8291	11 886
MIROCM	Center for Climate System Research (The University of Tokyo), National Institute for Environmental Studies, and Frontier Research Center for Global Change	Japan	miroc3_2_medres (mirocmedr)	3	~ 2.8	2.8125	1169	1706
MIUB	Meteorological Institute of the University of Bonn, Meteorological Research Institute of KMA, and Model and Data group	Germany South Korea	miub_echo_g	3	~ 3.7	3.75	631	916
MPI	Max Planck Institute for Meteorology	Germany	mpi_echam5 (mpi)	3	~ 1.8	1.875	2820	4068
MRI	Japan Meteorological Research Institute	Japan	mri_cgcm2_3_2a (mri)	5	~ 2.8	2.8125	1169	1706
PCM	National Center for Atmospheric Research	USA	pcm	1	~ 2.8	2.8125	1169	1706

[a] na: not available.
[b] Based on mean annual precipitation comparison between GCM and CRU.
[c] Based on mean annual temperature comparison between GCM and CRU.

Table 2. Comparison between CMIP3 GCM modelled and observed variables (reproduced from Räisänen, 2007, Table 1).

Variable	RMS error				Spatial correlation			
	Mean	Min	Max	21-model mean[a]	Mean	Min	Max	21-model mean
Surface temperature (°C)	2.32	1.58	4.56	1.43	0.989	0.968	0.994	0.996
Precipitation (mm day^{-1})	1.35	0.97	1.86	0.95	0.775	0.597	0.867	0.872
MSLP (hPa)	3.96	2.06	8.33	2.65	0.880	0.497	0.984	0.945

[a] 21-model mean is mean value for the 21 CMIP3 GCMs.

HESSD
11, 4531–4578, 2014

Part 1: Assessment of Global Climate Models

T. A. McMahon et al.

Table 3. Summary of performance of 23 CMIP3 GCMs in simulating present climate.

Source:	Macadam et al. (2010)	Gleckler et al. (2008)		Reichler and Kim (2008)
Variables:	MAT	P		Many
Method:	Ranking[c]	Rel. error ranking	Overall rank	Perf. Index[a]
GCM				
BCCR	15	= 13	14	No
CCCMA-t47	9	= 1	= 3	Yes
CCCMA-t63	na[b]	= 3	na	Yes
CCSM	6	= 13	= 10	Yes
CNRM	8	= 19	13	No
CSIRO	12	= 3	7	No
GFDL2.0	16	= 3	= 10	Yes
GFDL2.1	5	= 3	2	Yes
GISS-AOM	na	= 13	na	No
GISS-EH	na	= 19	na	No
GISS-ER	1	= 17	9	No
HadCM3	2	= 9	5	Yes
HadGEM	14	= 17	15	Yes
IAP	na	= 9	na	No
INGV	na	na	na	Yes
INM	13	= 19	16	No
IPSL	10	= 9	= 10	No
MIROCH	na	= 9	na	Yes
MIROCM	7	= 3	= 3	Yes
MIUB	4	= 1	1	na
MPI	3	= 13	8	Yes
MRI	11	= 3	6	Yes
PCM	17	22	17	No

[a] As summarised in Smith and Chandler (2010) (The performance index is based on the error variance between modelled and observed climate for 14 climate and ocean variables. "Yes" indicates the variance error is less than the median across the GCMs.)
[b] na: not available or not applicable.
[c] Rank 1 is best rank.

HESSD
11, 4531–4578, 2014

Part 1: Assessment of Global Climate Models

T. A. McMahon et al.

Table 4. Performance statistics comparing GCM mean and standard deviation of annual precipitation, mean annual temperature, and mean monthly patterns of precipitation and temperature with concurrent observed data. (Analysis based on untransformed data.).

GCM Name	MAP			SDP			MAT			Monthly pattern	
	R^2	NSE	RMSE	R^2	NSE	RMSE	R^2	NSE	RMSE	NSE Prec	NSE Temp
CCCMA-t47	0.498	0.457	435	0.342	0.252	63	0.984	0.953	3.14	0.409	0.838
CCCMA-t63	0.519	0.458	447	0.397	0.328	65	0.984	0.940	3.59	0.364	0.797
CCSM(1)[a]	0.496	0.483	460	0.426	0.413	71	0.982	0.981	2.06	−0.178	0.910
CCSM(2)	0.488	0.473	464	0.423	0.411	71	0.982	0.981	2.03	−0.210	0.912
CCSM(3)	0.493	0.479	462	0.418	0.403	71	0.981	0.980	2.08	−0.195	0.908
CCSM(4)	0.500	0.488	457	0.426	0.410	71	0.982	0.980	2.08	−0.174	0.911
CCSM(5)	0.493	0.480	461	0.423	0.410	71	0.983	0.981	2.02	−0.210	0.909
CCSM(6)	0.494	0.480	461	0.437	0.426	70	0.982	0.981	2.04	−0.181	0.909
CCSM(7)	0.496	0.483	460	0.429	0.420	71	0.982	0.981	2.06	−0.173	0.907
CCSM(9)	0.500	0.488	457	0.400	0.393	72	0.982	0.980	2.08	−0.157	0.910
CNRM	0.445	0.246	527	0.479	0.321	65	0.979	0.967	2.67	−0.631	0.879
CSIRO	0.387	0.363	503	0.462	0.452	65	0.971	0.959	2.99	0.034	0.825
GFDL2.0	0.544	0.528	434	0.588	0.460	63	0.980	0.934	3.79	−0.092	0.760
GFDL2.1	0.534	0.518	436	0.570	0.196	77	0.979	0.970	2.54	0.071	0.884
GISS-AOM(1)	0.330	−0.093	624	0.142	0.039	73	0.972	0.969	2.55	−0.325	0.873
GISS-AOM(2)	0.330	−0.087	623	0.132	0.027	74	0.972	0.970	2.54	−0.306	0.876
GISS-EH(1)	0.373	−0.510	692	0.210	−0.397	78	0.963	0.956	3.03	−0.856	0.858
GISS-EH(2)	0.375	−0.502	690	0.176	−0.589	83	0.962	0.955	3.07	−0.920	0.852
GISS-EH(3)	0.368	−0.535	697	0.181	−0.521	81	0.962	0.955	3.06	−0.858	0.856
GISS-ER(1)	0.386	−0.347	653	0.254	−0.115	70	0.970	0.960	2.87	−0.819	0.854
GISS-ER(2)	0.381	−0.357	656	0.203	−0.372	77	0.970	0.959	2.90	−0.739	0.850
GISS-ER(4)	0.386	−0.340	652	0.223	−0.214	72	0.970	0.960	2.88	−0.742	0.854
HadCM3	0.662	0.630	363	0.618	0.572	51	0.988	0.973	2.43	0.227	0.893
HadGEM	0.571	0.302	531	0.457	0.178	82	0.977	0.953	3.22	0.046	0.824
IAP(1)	0.496	0.438	456	0.191	0.096	75	0.963	0.894	4.64	−0.910	0.777
IAP(2)	0.493	0.433	458	0.188	0.041	77	0.962	0.895	4.61	−0.989	0.779
IAP(3)	0.499	0.440	455	0.186	0.048	77	0.963	0.896	4.60	−0.922	0.781
INGV	0.681	0.672	371	0.492	0.468	70	0.983	0.973	2.45	−0.263	0.882
INM	0.450	0.439	431	0.287	0.099	65	0.969	0.952	3.21	−0.247	0.833
IPSL	0.394	0.116	563	0.421	0.223	68	0.967	0.957	3.05	−0.147	0.846
MIROCH	0.588	0.370	514	0.583	0.570	63	0.974	0.971	2.54	0.107	0.906
MIROCM(1)	0.555	0.512	424	0.477	0.454	58	0.970	0.969	2.58	0.061	0.899
MIROCM(2)	0.552	0.508	425	0.525	0.501	56	0.970	0.969	2.58	0.054	0.900
MIROCM(3)	0.549	0.505	427	0.459	0.428	60	0.971	0.970	2.52	0.041	0.902
MIUB(1)	0.689	0.676	336	0.527	0.510	51	0.979	0.960	2.92	0.166	0.870
MIUB(2)	0.684	0.671	338	0.529	0.513	51	0.979	0.962	2.85	0.155	0.867
MIUB(3)	0.691	0.678	335	0.524	0.515	51	0.979	0.958	2.99	0.167	0.860
MPI(1)	0.543	0.538	429	0.464	0.437	66	0.985	0.984	1.88	0.014	0.939
MPI(2)	0.541	0.536	430	0.462	0.415	67	0.985	0.983	1.90	−0.002	0.939
MPI(3)	0.542	0.536	430	0.507	0.479	63	0.986	0.984	1.87	0.007	0.940
MRI(1)	0.617	0.535	414	0.507	0.499	56	0.977	0.969	2.57	0.217	0.912
MRI(2)	0.615	0.537	413	0.513	0.491	56	0.976	0.968	2.64	0.216	0.907
MRI(3)	0.617	0.541	411	0.523	0.505	55	0.977	0.969	2.57	0.222	0.911
MRI(4)	0.619	0.539	412	0.532	0.523	54	0.977	0.969	2.60	0.195	0.911
MRI(5)	0.615	0.538	412	0.503	0.487	56	0.977	0.968	2.62	0.211	0.907
PCM	0.360	0.190	546	0.336	0.135	73	0.975	0.943	3.49	−0.415	0.798

[a] In the paper a parenthesis after a GCM name indicates run number.

HESSD
11, 4531–4578, 2014

Part 1: Assessment of Global Climate Models

T. A. McMahon et al.

Table 5. Köppen climate classification (adapted from Peel et al., 2007).

Köppen class	Description of climate
Af	Tropical, rainforest
Am	Tropical, monsoon
Aw	Tropical. savannah
BWh	Arid, desert hot
BWk	Arid, desert cold
BSh	Arid, steppe hot
BSk	Arid, steppe cold
Csa	Temperate, dry and hot summer
Csb	Temperate, dry and warm summer
Csc	Temperate, dry and cold summer
Cwa	Temperate, dry winter and hot summer
Cwb	Temperate, dry winter and warm summer
Cwc	Temperate, dry winter and cold summer
Cfa	Temperate, without dry season and hot summer
Cfb	Temperate, without dry season and warm summer
Cfc	Temperate, without dry season and cold summer
Dsa	Cold, dry and hot summer
Dsb	Cold, dry and warm summer
Dsc	Cold, dry and cool summer
Dsd	Cold, dry summer and very cold winter
Dwa	Cold, dry winter and hot summer
Dwb	Cold, dry winter and warm summer
Dwc	Cold, dry winter and cool summer
Dwd	Cold, dry winter and very cold winter
Dfa	Cold, without dry season and hot summer
Dfb	Cold, without dry season and warm summer
Dfc	Cold, without dry season and cool summer
Dfd	Cold, without dry season and very cold winter
ET	Polar, tundra
EF	Polar, frost

HESSD
11, 4531–4578, 2014

Table 6a. Köppen climate estimated by MPI(3) compared with the observed Köppen climate for **(a)** the one-letter and **(b)** the two-letter climate classification.

		CRU					
	Land Surface	A	B	C	D	E	Sum
	A	**414**	19	8	0	0	441
	B	68	**339**	52	17	0	476
GCM	C	24	62	**319**	27	0	432
	D	0	76	16	**1085**	17	1194
	E	0	6	7	143	**121**	277
	Sum	506	502	402	1272	138	2820

HESSD
11, 4531–4578, 2014

Table 6b. Continued.

	Land Surface	CRU													
		Af	Am	Aw	BW	BS	Cs	Cw	Cf	Ds	Dw	Df	ET	EF	Sum
	Af	**57**	0	2	0	0	0	0	0	0	0	0	0	0	59
	Am	24	**19**	13	0	0	0	0	0	0	0	0	0	0	56
	Aw	25	49	**225**	0	19	0	4	4	0	0	0	0	0	326
	BW	2	1	2	**134**	50	3	4	0	0	0	2	0	0	198
	BS	4	11	48	50	**105**	13	19	13	4	0	11	0	0	278
	Cs	0	0	0	10	18	**35**	9	20	1	0	6	0	0	99
GCM	Cw	0	1	17	0	5	0	**62**	1	0	1	0	0	0	87
	Cf	2	2	2	3	26	1	35	**156**	0	0	19	0	0	246
	Ds	0	0	0	0	33	2	1	1	**38**	1	40	0	0	116
	Dw	0	0	0	0	5	0	1	0	0	**102**	2	0	0	110
	Df	0	0	0	3	35	0	4	7	2	57	**843**	17	0	968
	ET	0	0	0	0	6	2	2	3	8	22	113	**93**	0	249
	EF	0	0	0	0	0	0	0	0	0	0	0	11	**17**	28
	Sum	114	83	309	200	302	56	141	205	53	183	1036	121	17	2820

HESSD
11, 4531–4578, 2014

Part 1: Assessment of Global Climate Models

T. A. McMahon et al.

Table 7. Proportion of GCM grid cells (20C3M) that reproduce observed CRU Köppen climate classification over the period January 1950–December 1999.

GCM Name	Köppen climate class[a]		
	Three-letter	Two-letter	One-letter
CCCMA-t47	0.498	0.620	0.753
CCCMA-t63	0.429	0.558	0.709
CCSM(1)	0.488	0.558	0.749
CCSM(2)	0.489	0.563	0.748
CCSM(3)	0.424	0.545	0.744
CCSM(4)	0.466	0.549	0.749
CCSM(5)	0.444	0.519	0.727
CCSM(6)	0.490	0.563	0.757
CCSM(7)	0.488	0.556	0.749
CCSM(9)	0.489	0.560	0.755
CNRM	0.539	0.602	0.775
CSIRO	0.506	0.601	0.775
GFDL2.0	0.430	0.563	0.726
GFDL2.1	0.508	0.590	0.781
GISS-AOM(1)	0.460	0.559	0.773
GISS-AOM(2)	0.456	0.561	0.773
GISS-EH(1)	0.407	0.487	0.751
GISS-EH(2)	0.402	0.482	0.741
GISS-EH(3)	0.400	0.473	0.744
GISS-ER(1)	0.426	0.478	0.732
GISS-ER(2)	0.424	0.468	0.722
GISS-ER(4)	0.426	0.478	0.732
HadCM3	0.549	0.624	0.797
HadGEM	0.563	0.676	0.818
IAP(1)	0.362	0.484	0.790
IAP(2)	0.368	0.480	0.784
IAP(3)	0.369	0.490	0.784
INGV	0.495	0.616	0.815
INM	0.452	0.526	0.731
IPSL	0.459	0.544	0.749
MIROCH	0.496	0.631	0.806
MIROCM(1)	0.477	0.597	0.749
MIROCM(2)	0.477	0.594	0.759
MIROCM(3)	0.469	0.583	0.748
MIUB(1)	0.528	0.604	0.783
MIUB(2)	0.528	0.604	0.783
MIUB(3)	0.520	0.610	0.778
MPI(1)	0.599	0.666	0.801
MPI(2)	0.593	0.657	0.805
MPI(3)	0.602	0.669	0.808
MRI(1)	0.534	0.644	0.808
MRI(2)	0.521	0.625	0.798
MRI(3)	0.527	0.632	0.798
MRI(4)	0.528	0.634	0.799
MRI(5)	0.532	0.641	0.803
PCM	0.397	0.481	0.660

[a] The three-, two- and one-letter climate classes are listed in Table 5.

HESSD
11, 4531–4578, 2014

Part 1: Assessment of Global Climate Models

T. A. McMahon et al.

Table 8. GCM run rank (rank 1 = best) based on Nash–Sutcliffe efficiency values from comparison of 20C3M and concurrent observed grid cell data.

GCM Name	MAP rank	SDP rank	MAT rank	Monthly pattern rank	Rank Sum	Overall GCM rank
CCCMA-t47	28	30	38	19	115	12
CCCMA-t63	27	28	42	22	119	13
CCSM(1)	21	22	7	18	68	8
CCSM(2)	26	23	5	17	71	
CCSM(3)	25	26	10	21	82	
CCSM(4)	20	25	11	16	72	
CCSM(5)	24	24	4	21	73	
CCSM(6)	23	19	6	20	68	
CCSM(7)	22	20	8	19.5	69.5	
CCSM(9)	19	27	9	17	72	
CNRM	36	29	26	30.5	121.5	14
CSIRO	34	16	32	28.5	110.5	11
GFDL2.0	14	14	43	34	105	10
GFDL2.1	15	32	15	17.5	79.5	9
GISS-AOM(1)	40	39	20	30.5	129.5	
GISS-AOM(2)	39	40	17	29.5	125.5	15
GISS-EH(1)	45	44	35	35.5	159.5	22
GISS-EH(2)	44	46	37	39	166	
GISS-EH(3)	46	45	36	36.5	163.5	
GISS-ER(1)	42	41	28	36	147	19
GISS-ER(2)	43	43	31	36.5	153.5	
GISS-ER(4)	41	42	29	36	148	
HadCM3	5	1	13	12	31	1
HadGEM	35	33	39	28	135	17
IAP(1)	31	36	46	44	157	
IAP(2)	32	38	45	45	160	
IAP(3)	29	37	44	44	154	21
INGV	3	13	12	28	56	5
INM	30	35	40	35	140	18
IPSL	38	31	34	29.5	132.5	16
MIROCH	33	2	14	14.5	63.5	7
MIROCM(1)	16	15	22	17	70	
MIROCM(2)	17	8	21	17	63	6
MIROCM(3)	18	18	16	17.5	69.5	
MIUB(1)	2	6	30	18.5	56.5	
MIUB(2)	4	5	27	19.5	55.5	4
MIUB(3)	1	4	33	19	57	
MPI(1)	9	17	2	10.5	38.5	
MPI(2)	12	21	3	12	48	
MPI(3)	11	12	1	10.5	34.5	2
MRI(1)	13	9	18	5	45	
MRI(2)	10	10	25	10.5	55.5	
MRI(3)	6	7	19	5.5	37.5	3
MRI(4)	7	3	23	8	41	
MRI(5)	8	11	24	11.5	54.5	
PCM	37	34	41	38.5	150.5	20

HESSD
11, 4531–4578, 2014

Part 1: Assessment of Global Climate Models

T. A. McMahon et al.

HESSD
11, 4531–4578, 2014

Table 9. Better performing GCMs identified from the literature and our analyses.

Grid cells (Tables 4 and 8) (Col. 1)[a]	Literature (Table 3) (Col. 2)	Adopted GCMs (Col. 3)
	CCCMA-t47	
	CCSM	
	GFDL2.0	
	GFDL2.1	
HadCM3	HadCM3	HadCM3
INGV		
	MIROCH	
MIROCM	MIROCM	MIROCM
MIUB	MIUB	MIUB
MPI	MPI	MPI
MRI	MRI	MRI

[a] Included only the six highest ranked GCMs from our analyses.

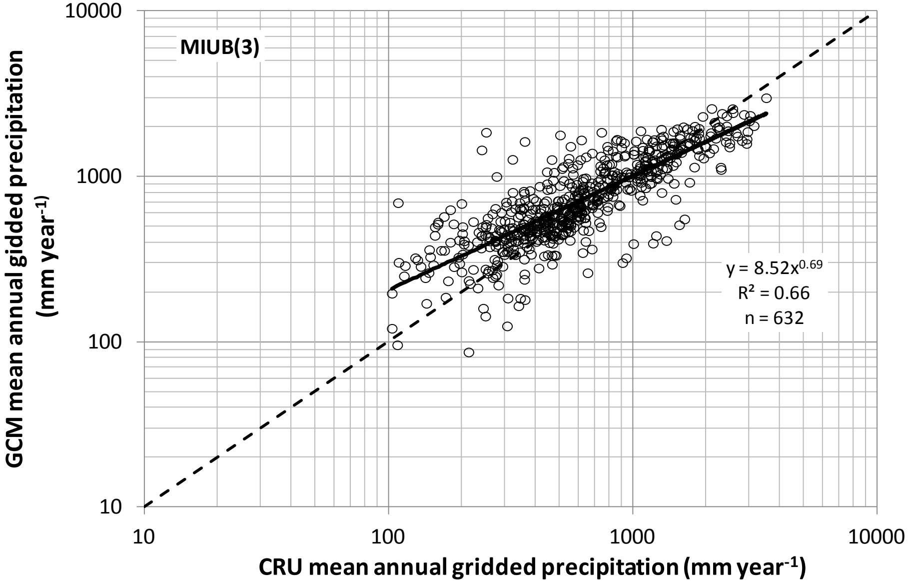

Fig. 1. Comparison of MIUB(3) model estimates of observed mean annual precipitation with CRU estimates. (Based on untransformed precipitation NSE = 0.678, rank 1 of 46 runs, and $R^2 = 0.691$.)

HESSD
11, 4531–4578, 2014

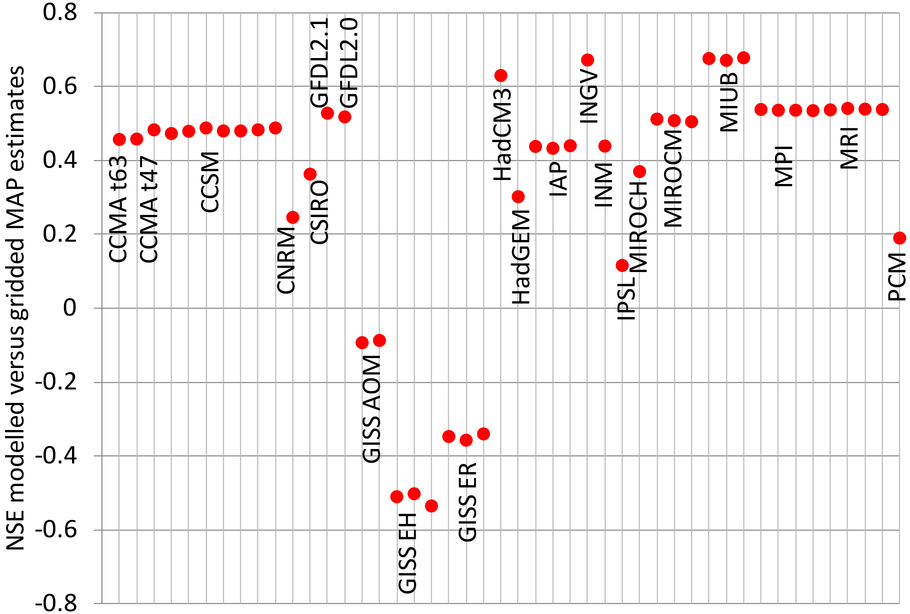

Fig. 2. Nash–Sutcliffe efficiency (NSE) values for modelled vs. observed MAP untransformed estimates.

HESSD
11, 4531–4578, 2014

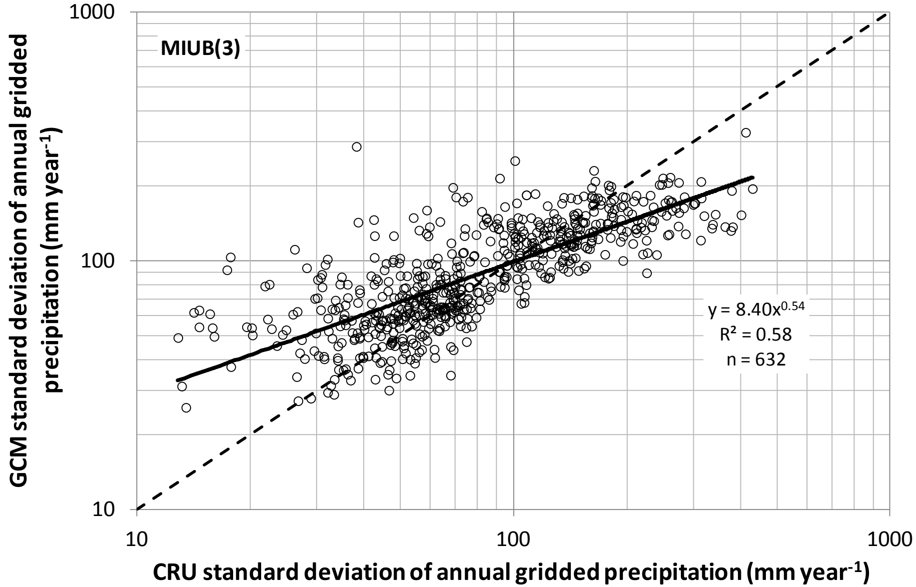

Fig. 3. Comparison of MIUB(3) model estimates of the standard deviation of annual precipitation with CRU observed estimates. (Based on untransformed precipitation NSE = 0.515, rank 4 of 46 runs, and R^2 = 0.524.)

HESSD
11, 4531–4578, 2014

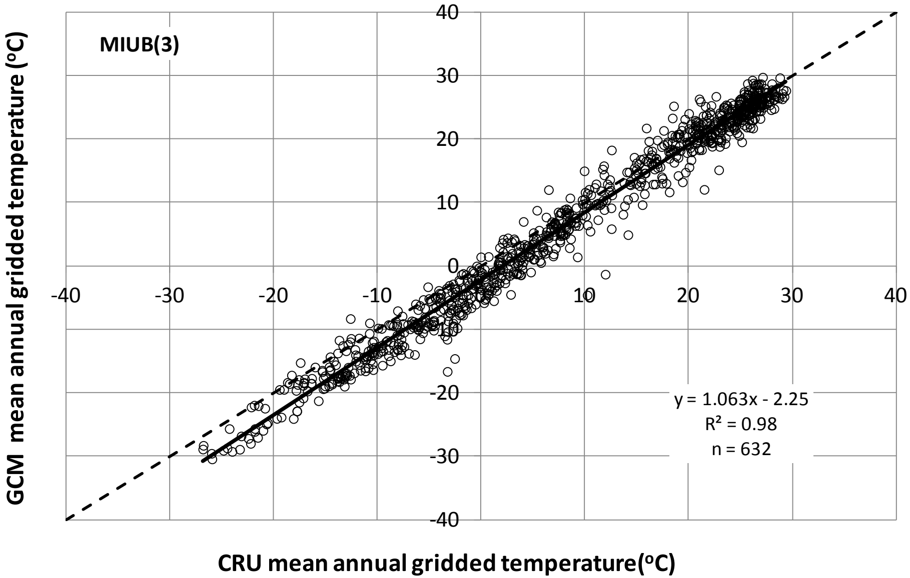

Fig. 4. Comparison of MIUB(3) model estimates of mean annual temperature with CRU estimates. (Based on untransformed temperature NSE = 0.958, rank 33 of 46 runs, and R^2 = 0.979).

HESSD
11, 4531–4578, 2014

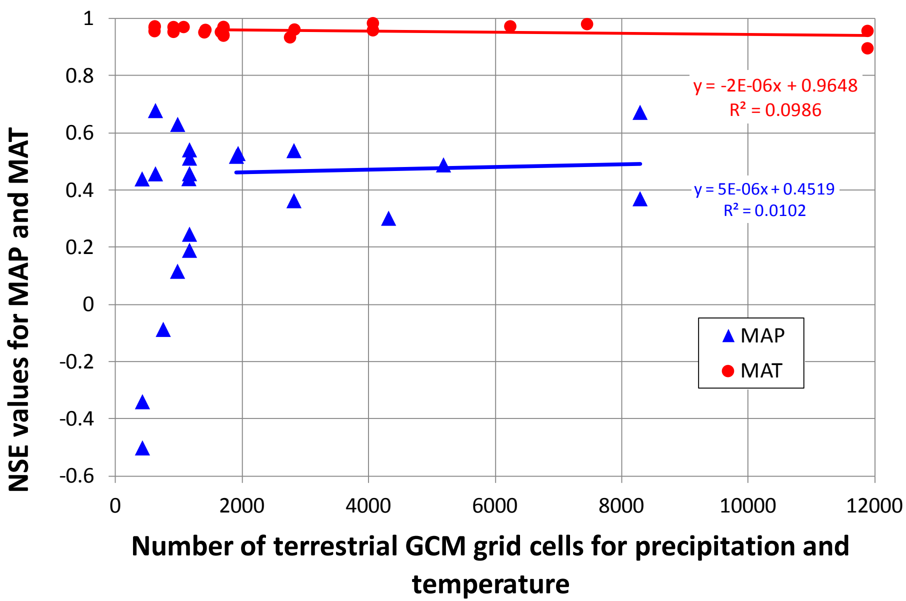

Fig. 5. Relating 22 GCM resolutions (as the number of terrestrial grid cells for MAP) to model performance based on Nash–Sutcliffe efficiency for mean annual precipitation and mean annual temperature. (The trend lines are fitted to data with > 1500 grid cells.)

HESSD
11, 4531–4578, 2014

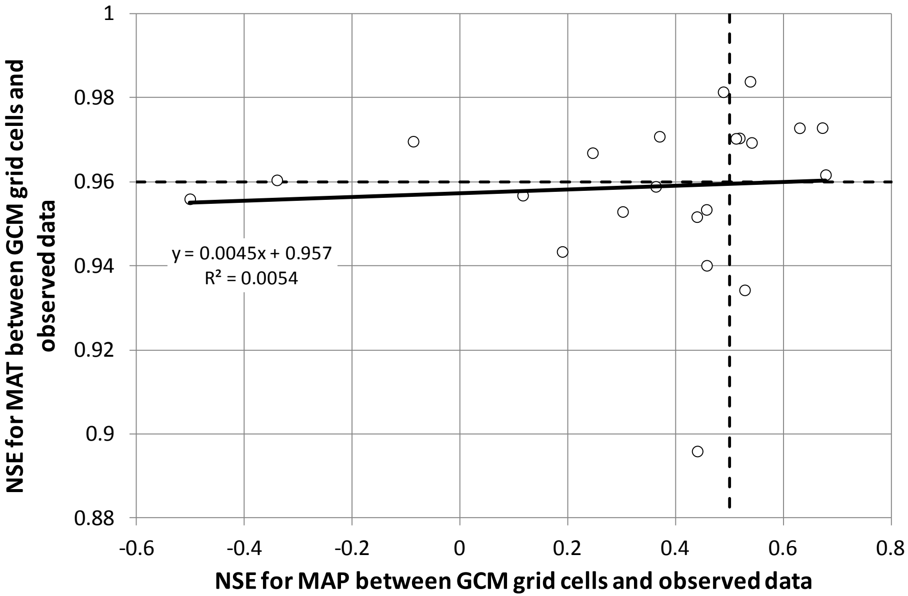

Fig. 6. Comparison of Nash–Sutcliffe efficiency values between GCM and observed MATs with Nash–Sutcliffe efficiency values between GCM and observed MAPs.

HESSD
11, 4531–4578, 2014

Part 1: Assessment of Global Climate Models

T. A. McMahon et al.

Title Page
Abstract | Introduction
Conclusions | References
Tables | Figures
◀| |▶
◀ ▶
Back | Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

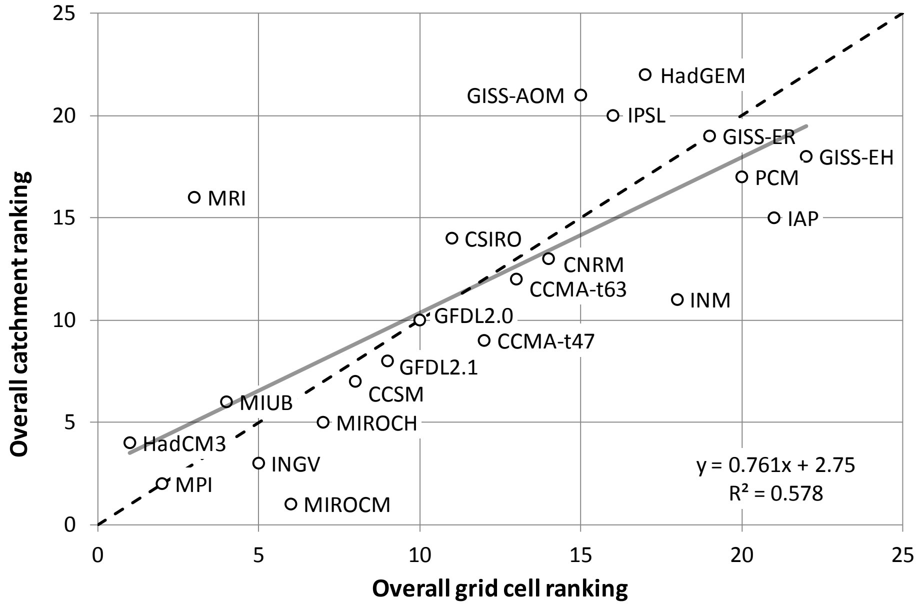

Fig. 7. Comparison of overall GCM rank from grid cell (Table 8) and catchment analyses (not shown).

HESSD
11, 4531–4578, 2014

Part 1: Assessment of Global Climate Models

T. A. McMahon et al.