**Authors responses (in *red and italics*) to referees comments (in black)**

*We thank the two reviewers for their constructive comments and suggestions and believe our proposed revisions in response to their comments, outlined below, will significantly improve the manuscript.*

**Reviewer #1**

**Major remarks**

The authors consider the performance of 22 GCMs from the CMIP3 exercise. They compare and rank the GCMs using various statistical measures for precipitation (P) and temperature (T), and they compare their results to ranking from literature based on different performance indices. In this way they identified the five best performing GCMs with regard to their plans that they want to use output from these best performing models to simulate runoff in a future study that shall be published in another paper as part 2 of the study.

*Noted.*

Even though I found it interesting to see how ranking based on P and T statistics only compares to more sophisticated multi-variable performance indices, I don't see what is really novel in the present manuscript.

*Our prime objective in Part 1 was to select a sub-set of GCMs that require the least bias correction as input to a hydrologic model for estimating runoff in the second paper (Part 2). Because monthly precipitation (P) depths and temperature (T) values are inputs to the rainfall-runoff model, it is sensible to assess the performance of the GCMs primarily in terms of bias and variance of P and T.*

*The papers are written with surface water hydrologists and water management engineers in mind. Although specifically interested in the runoff results, described in Part 2, hydrologists and water managers are also interested in the performance of GCMs in reproducing hydrologically relevant metrics presented in this paper. It is appropriate, therefore, that the metrics used to assess GCM performance be relevant to hydrologists and engineers (for example, related to bias and absolute RMSE) rather than those generally reported (for example, spatial correlations, normalised RMSE, Taylor plots, PDF and Bayesian approaches, skill scores, etc) and to the hydrologic modelling in Part 2.*

*We believe the GCM performance assessment presented in this paper provides a hydrologic perspective that differs from most GCM performance assessments in the literature. This belief is based on Table A (see end of this document), which presents a review of GCM performance measures from 34 papers published from 1992 to 2014. It is noted that in Table A only five studies report metrics of absolute bias of simulated precipitation and/or temperature as required for Part 2 and of interest to the hydrologic community. (It is not proposed that Table A be incorporated in the revised manuscript.) We are not suggesting the assessment of GCM performance in Part 1 is novel but rather pragmatic and offers a different approach to assessing GCMs in view of the requirement for monthly P and T values in Part 2, and the potential readership of the papers. We note the reviewer "... found it interesting to see how ranking based on P and T statistics only compares to more sophisticated multi-variable performance indices ...".*

*No change to the manuscript.*

I also wonder why the authors do not look additionally at characteristics of annual runoff simulated by the GCMs, and compare these values to observed river discharge. At annual time scales or coarser, the lateral routing does not play a role, and such intercomparison may give insights on how the GCMs generally simulate the terrestrial part of the hydrological cycle (as P is also considered).

*There are two reasons for why we did not look at GCM generated runoff in this study.*

*1) Arora (2001) demonstrated that the quality of GCM generated runoff is dependent mainly on the quality of GCM precipitation and any bias in GCM precipitation is amplified in the resulting GCM runoff. In our paper we assess GCM bias in reproducing observed precipitation conditions and find substantial biases for all GCMs, thus we would expect significant bias in GCM runoff.*

*2) Sperna Weiland et al. (2012) assessed the performance of GCM runoff for two GCMs where the Land Surface Schemes had been tuned to reproduce realistic runoff estimates. They found that runoff estimates from an external hydrologic model generally outperformed the GCM runoff estimates.*

*Therefore in Part 2 of our study we used an external hydrologic model for generating runoff with bias corrected inputs.*

*We will add a comment to a revised manuscript indicating the reasons why we are not using GCM generated runoff.*

While a future paper plan is sufficient to motivate why the authors look at temperature and precipitation measures, it does not really justify larger subsections on “estimating streamflow” and “potential evapotranspiration” as both subsections do not contribute to the main objective of the manuscript. They certainly would contribute to the follow up paper, but this is not part of the present paper. In this respect the complete Appendix is not necessary. The content of the appendix is interesting and probably appropriate for the second paper, but does not contribute to the understanding of the main results of this paper.

*Agreed. We will combine and substantially reduce the “estimating streamflow” and “potential evapotranspiration” sections and refocus the context on a brief justification of the variables (precipitation and temperature) to be investigated (see next response). The Appendix provides justification for adopting temperature as a surrogate for potential evapotranspiration and a discussion of the errors entailed in that decision. This material is more appropriate in Part 1 than Part 2 and will be moved to Supplementary Material of Part 1.*

By the way if it should be justified why only precipitation and temperature are necessary to simulate runoff based on literature, then this part clearly lacks to take into account recent research that addresses global multi-model studies with several GCMs and global hydrology models conducted within the EU project WATCH (see, e.g. Haddeland et al. 2011 Multi- Model Estimate of the Global Terrestrial Water Balance: Setup and First Results. J. Hydrometeor. 12, 10.1175/2011JHM1324.1, 869-884.) and the ISIMIP exercise (see http://www.isi-mip.org/). In this respect, the authors stated in the beginning of p. 4535.: “…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).” While I agree that they are sufficient to estimate current climate runoff characteristics, I doubt their validity under global warming conditions. Recent research has shown that temperature-only based estimates of potential evapotranspiration tend to fail under global warming conditions (e.g. Hagemann et al. 2011

Impact of a statistical bias correction on the projected hydrological changes obtained from three GCMs and two hydrology models. J. Hydrometeor. 12, 10.1175/2011JHM1336.1: 556-578.).

*We agree with the reviewer's comment that temperature-only PET models, like Thornthwaite, can be problematic under future projected conditions. In the former Appendix (now proposed to be Supplementary Material to Part 1) we provide a comprehensive discussion of recent research about the pros and cons of PET models in hydrologic modelling of present and future conditions, particularly in energy limited environments. In that discussion we note that conceptual hydrologic models calibrated for a specific catchment are flexible enough to extract the PET information they need from whichever PET data (temperature based or more complex) are used (see Chapman, 2003). The situation of using a calibrated conceptual model on a single catchment is different to the approach described in Haddeland et al. (2011, 2012) and Hagemann et al. (2011) in which the models are not calibrated for a specific catchment.*

*The references to Haddeland et al. (2011; 2012) and Hagemann et al. (2011) provide interesting case studies in the issue of using a simple, or complex, estimate of PET in hydrologic modelling. The resulting mean annual runoff and evapotranspiration estimates for present conditions from these models exhibit large bias. For example, in Haddeland et al. (2012, Table 2) mean annual runoff is estimated from four hydrologic models driven by three climatological datasets (WFD, bias and non-bias corrected ECHAM Cntrl and IPSL Cntrl) for the period 1971-2000. The hydrologic models adopt a range of evaporation models: LPJmL (Priestley-Taylor); VIC (Penman-Monteith); MPI-MH (Thornthwaite); and Water GAP (Priestley-Taylor). The average mean annual runoff estimate for each model, driven by WFD and bias corrected ECHAM and IPSL, is 63,242 $km^3$ $year^{-1}$ (LPJmL), 45,343 $km^3$ $year^{-1}$ (MPI-MH), 54,617 $km^3$ $year^{-1}$ (VIC), and 55,214 $km^3$ $year^{-1}$ (Water GAP) respectively. These values are in stark contrast to the observation-based mean annual runoff value of 37,288 $km^3$ $year^{-1}$ from Dai and Trenberth (2002, page 670; Trenberth et al., 2007) for the period 1979-2000. The large difference in runoff between these models and the observed value of Dai and Trenberth, in the order of 50% for all models except MPI-HM, is concerning. Although Haddleland et al.(2011, page 873) imply this overestimation may be due to an inadequate definition of the terrestrial land area, the methodology of Dai and Trenberth's (2002) and Trenberth et al.'s (2007) was based on runoff volumes rather than runoff depth x area and therefore is largely independent of the global terrestrial land surface. For evapotranspiration, Hagemann et al. (2011, Figure 6) present bias estimates for 12 catchments over the period 1961-1990 and find the LPJmL (Priestley-Taylor PET model) driven with bias corrected inputs results in an average negative bias of approximately -18%, whereas, the MPI-HM (Thornthwaite PET model) has an average positive bias of approximately 1% for the same catchments.*

*In summary, if the references suggested by the reviewer had presented runoff and evaporation results that were consistent with observations, then the argument of the reviewer that "Recent research has shown that temperature-only based estimates of potential evapotranspiration tend to fail under global warming conditions (e.g. Hagemann et al. 2011)" would have provided a stronger justification for adopting a more complex evaporation model. Although we agree that temperature-only PET estimates can be problematic under future conditions, we do not agree that comparison of biased runoff and evapotranspiration estimates presented in these references can be used to support this argument. In fact the results presented reflect our concerns outlined in the former Appendix, that the input data for complex PET models need to be reliable, which supports our methodology of driving our hydrologic model with temperature as a surrogate for PET while being aware of likely PET errors in energy limited environments under future conditions.*

*We note that in Part 2 that although the PERM hydrologic model uses T as an input, it does so through a transfer function (PERM parameter a) that allows the model to estimate actual evaporation. This transfer function is calibrated at each catchment and effectively represents the evaporative environment of each catchment. The evaporation model in PERM is not a temperature-only model like Thornthwaite, which estimates only the 12 monthly average values, but is a time-series broadly representing the radiation and aerodynamic evaporation processes experienced by the catchment during the calibration period. A necessary assumption of any hydrologic modelling under climate change is that the calibrated parameters apply into the future.*

*In view of the comments above, we see no need to change the manuscript.*

I suggest a) either to clearly focus on the ranking and the comparison to other skill scores, i.e. removing all parts of the present manuscript that do not contribute to this issue. But here question arise whether such a study would qualify to be a full HESS paper? A paper has to qualify by itself, and not because there shall be a second part that is more hydrology oriented. Thus, it might also be appropriate to b) merge this manuscript with the second paper part to have a complete study. But then the limitations addressing the use of temperature-only based PET estimates need to be thoroughly taken into account.

*We note the reviewer presents two options. We have given serious consideration to option (b), but have decided to not merge Parts 1 & 2 for the following two reasons. (1) We believe Part 1 does qualify as a full paper (see second comment above) following the proposed revisions outlined in response to both reviewers. (2) A merged paper would significantly dilute the novel material presented in Part 2 and be far too long.*

*We believe the proposed revisions to Part 1 in response to the review comments, option (a), will improve the manuscript.*

Please note that I also wrote comments to paper parts that may be removed if option a) is chosen. In summary, I suggest major revisions to be conducted before the paper may be accepted for publication.

*Noted.*

**Minor Comments**

In the following suggestions for editorial corrections are marked in Italic.

p. 4534 – line 1

… to characterize the …

*Agreed, will change.*

p. 4543 – line 21-24

Even though the citation is from a published paper, I disagree with this. While I agree that it is very important that a GCM captures the current climate reasonably well, this does not mean that it will be closer to the future greenhouse response of the real world. Here, it is important that the GCM has the right climate sensitivity and that it adequately captures certain feedbacks that play an important role in this response. The latter, e.g., is currently being (and will be) investigated in research focusing on emergent constraints.

*Agreed. A sentence to reflect this comment will be added.*

p. 4545 – line 8

… data set is presented …

*Agreed, will change.*

p. 4547 – Sect. 5.1

The findings in this section are really interesting. They should be more highlighted and are worth to be explored further.

*We agree these results are interesting, but further exploration is beyond the scope of the paper. No change.*

p. 4550 – Sect. 5.5

You shouldn't talk about results that are not considered in this study. E.g. it is mentioned that "…for a range of catchment scales world-wide."

On one hand one would ask, Which scales? On the other hand, it is written: "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."

Thus, it is sufficient to talk about the results of large catchments, as it doesn't really make sense (as also realized by the authors) to look at small catchments if GCM results are considered.

*Agreed, we will remove this section and Figure 7 from a revised manuscript.*

p. 4553 – line 23-25

This conclusion does not apply for global climate change studies as relevant climate changes are ongoing exactly in those areas that are strongly energy and water limited, especially a sit is projected that many dry areas will become even drier, i.e. more areas become stronger water limited.

*We note that our conclusion was for areas that 'are not strongly water or energy limited', which the reviewer has misread as 'strongly energy and water limited'.*

On the other hand the largest warming signal is observed in high latitudes that are strongly energy limited. Thus, I cannot follow that argument that complex PET formulations are not necessary. Especially if there are indications that the very simple ones based on temperature only fail under climate change conditions (see above).

*In the former Appendix A (now proposed to be incorporated as Supplementary Material), we reviewed the literature dealing with this issue. Based on the studies of Chapman (2003), Andréassian et al. (2004) and Oudin et al. (2005a, b) we concluded that for estimating PET in calibrated hydrologic models a method based on temperature alone is satisfactory for the purposes of our study. In the Appendix we acknowledged that using temperature only information to drive a calibrated hydrologic model could result in errors in runoff estimation in energy limited environments, as noted by the reviewer. However, we also argued that errors in runoff estimation would be small in water limited regions and regions that are not strongly water or energy limited as the key hydrologic driver is precipitation and soil moisture, not energy input. For example, in a hot dry region a change in PET* *(either increase or decrease) will have negligible impact on runoff as the limiting variable is precipitation.*

*We do not believe incorporating the additional climate variables required to estimate potential evaporation by the Penman-based formulation of wind speed (at 2 m height), net radiation (requiring both short and long wave, incoming and outgoing, to be estimated) and relative humidity estimated at 2 m is justified given the difficulty in obtaining GCM data that are credible. Johnson and Sharma (2009) have shown that in terms of their Variable Convergence Score (VCS, scaled between 0 and 100, where 100 is perfect convergence between GCMs) the predictions of the surface wind and specific humidity have VCS scores of approximately 40, net long wave radiation about 20 compared with surface temperature and net short wave radiation of about 70 and precipitation at 10.*

*A revised manuscript and Supplementary Material will be amended to reflect this argument.*

p. 4554 – line 17-21

But less confidence should not outweigh wrong physical behaviour in the projections. I.e., just because there is less confidence in some variables, their effect on water and energy cycles at the land surface shouldn't be neglected.

*We understand, but do not accept this argument. There are two issues here: (1) the physical representation of PET; and (2) the quality of the input data to the PET equation. For hydrologic climate change impact assessments the practitioner is faced with a trade-off between these two issues. Using a Penman-based formulation will include all the PET relevant physical processes, but the results may not be useful if the input data are not of adequate quality (see Kay and Davies, 2008; Kingston et al., 2009). Furthermore, it should be noted that if the Penman equation were used for example, a decision will be required as to which wind function to adopt which can affect the resultant evaporation estimate by up to ~19% (McMahon et al., 2013, Supplementary Material page 28), so there is some uncertainty even if perfect data were available. Whereas, a simpler method based only on reliable input data may not include all PET physical processes, but there will be more confidence in the processes modelled. We adopted the latter approach in our study and acknowledged where this approach was likely to lead to error under future conditions (in energy limited environments).*

p. 4555 – line 7

It is written: "GCM projections of those process variables, other than temperature, may be unrealistic."

But if Temperature-only based parameterizations behave unrealistic in certain regions, too, then all climate change impact maybe unrealistic. That's actually some uncertainty one has to live with and pay regard to it.

*We agree with this comment. See previous response, no change in manuscript.*

p. 4555 – line 18

It is written:

"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."

As mentioned above: While this may be true for the present climate it has been shown that errors in PET behaviour can become important under future climate conditions.

*In water limited environments now, and in the future, changes in PET will have negligible impact on runoff as precipitation (hence soil moisture) is the dominant driver of runoff. In this environment, errors in PET have little to no impact on estimates of runoff from a calibrated rainfall-runoff model. As discussed in the Appendix, the case for energy limited environments is different.*

*We see no need to amend the manuscript.*

**Reviewer #2**

This manuscript and its companion paper attempts to quantify the of impact internal climate variability (as simulated by GCMs) on runoff risk assessments. This (the first paper) focus on selecting a subset of GCMs to focus on for further analysis in the companion paper.

*Noted.*

I have three comments on Part 1 that the authors may want to address.

1) Typically one would see uncertainty in climate change simulation to arise from three types of uncertainty: a) emission scenario uncertainty, b) model structural uncertainty (between GCMs) and c) internal climate variability (as represented by running an ensemble using one GCM changing only the initial conditions). In this paper the authors only discuss the latter two sources, though it could well be the case that emission scenario is the one with the largest uncertainty (particularly in the far future and for variables with strong direct relationship to global warming, i.e. temperature). It would be useful for the authors to provide some details on all sources of uncertainty and maybe comment on what sources are more dominant at what time horisons (see e.g. papers cited at http://climate.ncas.ac.uk/research/uncertainty/ as a starting point).

*We agree that scenario uncertainty is important, particularly in the far future, and will modify the text accordingly. In this study we focused on two periods, 1965-1994 and 2015-2044, where within-GCM and between-GCM uncertainties dominate over scenario uncertainty. This view is supported by Wilby and Harris (2006) who assessed the uncertainty in future low flows for the River Thames and ordered the uncertainties from largest to smallest as follows: GCM > (empirical) downscaling method > structure of hydrologic model > hydrological model parameters > emission scenario. In a paper on understanding uncertainty in regional temperature prediction, Hawkins and Sutton (2009) estimated the contributions to the total prediction uncertainty from internal variability, model uncertainty and scenario uncertainty where scenario uncertainty is least important. Hawkins and Sutton carried out a similar analysis for regional precipitation again concluding that internal variability contributes most of the uncertainty and scenario modelling is generally small (Hawkins and Sutton, 2011).*

*Based on the above, we will include a comment in a revised manuscript regarding the levels of uncertainty of internal variability, model uncertainty and emission scenario uncertainty.*

2) In Part 1, the focus lies on selecting a sub-set of GCMs from the CMIP3 archive. However, given that the methods that the researchers are using are statistical, would it not be possible to conduct experiments on all possible GCMs bar those that are outright bad (e.g. use metrics to remove poor models rather than selecting a few good ones?). The reason for keeping more models in would be that

a) you could well imagine that some models perform better in some regions than others but overall isn't one of the high scorers b) others have shown that models that do well in current climate do not necessarily agree on a future pathway, thus would you not want to be inclusive rather than exclusive? At the very least it would be good to see where the models you selected fall within the spectra of CMIP3 models, do they simulate an overall wet future/dry future/ relative to the CMIP3 ensemble?

*We remind the reviewer that we do analyse, assess and report on the performance of P and T for all GCMs and associated runs in the paper. It is not until Section 5.3 that we discuss which models we plan to carry through to Paper 2.*

*For this project our resources were limited so that the adoption of five GCMs was a trade-off between more analyses and interpretation as against less resources being devoted to runoff uncertainty and reservoir yield analysis. We believe we have achieved a reasonable compromise.*

*As requested by the reviewer, we include in this response Figure A in which we compare the ratio of mean annual precipitation from 2015-2034 to 1965-1994 against the mean annual temperature difference between 2015-2034 and 1965-1994 for the global land surface. Figure A shows the relative positions of the selected five GCM runs compared with the 23 CMIP3 GCMs including all 44 ensemble members. The five selected GCM runs are well distributed amongst the 44 GCM ensemble members, which indicate that they are reasonably representative of the range of future GCM projections. We observe that most GCM runs are clustered around the median response, except for the seven CCSM runs in the top right quadrant with a precipitation ratio > ~1.04.*

*In a revised manuscript we will include a sub-section that addresses this issue and include Figure A.*

3) The metrics is based on absolute data for rainfall, could this be an issue? All models have a bias but are we note more interested in the relative change? Since you bias correct before you use the data for the impact work, should bias really be equally important in selecting models as the other metrics used?

*No, we do don't believe this is an issue. To assess GCM performance from a hydrologic perspective requires absolute, rather than relative, values and our aim was to identify a sub-set of GCMs that required the least bias correction in precipitation and temperature that would be used in Paper 2 to estimate runoff.*

*No change to manuscript.*

Finally just on a structural comment, the content in Appendix A doesn't really sit comfortably within this paper. It would probably sit better in Part 2.

*Partially agree. The Appendix will be moved to Supplementary Material (see earlier comment in response to Reviewer 1).*

## ***References***

*(The references that relate to Table A would not be included in a revised manuscript.)*

*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.*

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

*Baker, N.C and, Huei-Ping Huang, H.-P.: A Comparative Study of Precipitation and Evaporation between CMIP3 and CMIP5 Climate Model Ensembles in Semiarid Regions. Journal of Climate 27:10, 3731-3749, 2014.*

*Biemans, H., Hutjes, R.W.A., Kabat, P., Strengers, B. J., Gertens, D. and Rost, S.: Effects of precipitation uncertainty on discharge calculations for main river basins. Journal of Hydrometeorology 10, 1011-1025, 2009.*

*Boer, G J. and Lambert, S. J.: Second-order space-time climate difference statistics, Climate Dynamics, 17, 213-218, 2001.*

*Bonsal, B. R. and Prowse, T. D.: Regional assessment of GCM-simulated current climate over Northern Canada, Artic, 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.*

*Chiew, F. H. S., Teng, J., Vaze, J., Post, D. A., Perraud, J. M., Kirono, D. G. C. and Viney, N. R.: Estimating climate change impact on runoff across southeast Australia: Method, results, and implications, Water Resources Research, 45, W10414, doi:10.1029/2008WR007338, 2009*

*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 and Planetary Change, 37, 103-133, 2003.*

*Dai, A. and Trenberth, K. E.: Estimates of freshwater discharge from continents: Latitudinal and seasonal variations, Journal of Hydrometeorology 3, 660-687, 2002.*

*Falloon, P., Betts, R., Wiltshire, A., Dankers, R., Mathison, C. and McNeall, D.: Validation of River Flows in HadGEM1 and HadCM3 with the TRIP River Flow Model. Journal of Hydrometeorology 12, 1157-1180, 2011.*

*Gleckler, P. J., Taylor, K. E. and Doutriaux, C.: Performance metrics for climate models. Journal of Geophysical Research 113, D06104, doi:10.1029/2007JD008972, 2008.*

*Haddeland, I., Clark, D. B., Franssen, W., Ludwig, F., Voβ, F., Arnell, N.W., Bertrand, N., Best, M., Folwell, S., Gerten, D., Gomes, S., Gosling, S. N., Hagemann, S., Hanasaki, N., Harding, R., Heinke, J., Kabat, P., Koirala, S, Oki, T., Polcher, J., Stacke, T., Viterbo, P., Weedon, G. and Yeh, P.: Multimodel estimate of the global terrestrial water balance: Setup and first results, J. Hydrometeor, 12, 869-884, 2011.*

*Haddeland, I., Heinke, J., Voβ, F., Eisner, S., Chen, C., Hagemann, S. and Ludwig, F.: Effects of climate model radiation, humidity and wind estimates on hydrological simulations. Hydrol. Earth Syst. Sci., 16, 305-318, 2012.*

*Hagemann, S., Chen, C., Haerter, J.O., Heinke, J., Gerten, D. and Piani, C.: Impact of a statistical bias correction on the projected hydrological changes obtained from three GCMs and two hydrology models. J. Hydrometeor. 12, 556-578, 2011.*

*Hawkins, E. and Sutton, R.: The potential to narrow uncertainty in regional climate predictions. Bulletin of the American Meteorological Society DOI:10.1175/2009BAMS2607.1, 1095-1107, 2009.*

*Hawkins, E. and Sutton, R.: The potential to narrow uncertainty in projections of regional precipitation change, Climate Dynamics, 37, 407-418, 2011.*

*Heo, K.-Y., Ha, K.-J., Yun, K.-S., Lee, S.-S., Kim, H.-J. and Wang, B.: Methods for uncertainty assessment of climate models and model predictions over East Asia. International Journal of Climatology 34:10.1002/joc.2014.34.issue-2, 377-390, 2014.*

*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., Richard, G. J., Cusack, S., Ingram, W. R., 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, Journal of 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, Cairns, Australia, 2618-2624, 2009a.*

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

*Kay, A.L., Davies, H.N.: Calculating potential evaporation from climate model data: A source of uncertainty for hydrological climate change impacts. Journal of Hydrology 358: 221-239, 2008.*

*Kingston, D.G, Todd, M.C., Taylor, R.G., Thompson, J.R., Arnell, N.W.: Uncertainty in the estimation of potential evapotranspiration under climate change. Geophysical Research Letters 36, L20403, 2009.*

*Knutti, R., Furrer, R., Tebaldi, C., Cermak, J and Meehl, G.A.: Challenges in combining projections from multiple climate models Journal of Climate, 23, 2739-2758, 2010.*

*Kumar, D., Kodra, E. and Ganguly, A.R.: Regional and seasonal intercomparison of CMIP3 and CMIP5 climate model ensembles for temperature and precipitation. Climate Dynamics, 2014.*

*Lambert, S. J. and Boer, G. J.: CMIP1 evaluation and Intercomparison of coupled climate models, Climate Dynamics, 17, 83-106, 2001.*

*Legates, D. R. and Willmott, C. J.: A comparison of GCM-simulated and observed mean January and July precipitation, Palaeogeography, Palaeoclimatology, Palaeoecology (Global and Planetary Change Section), 97,345-363, 1992.*

*McMahon, T. A., Peel, M. C., Lowe, L., Srikanthan, R., and McVicar, T. R.: Estimating actual, potential, reference crop and pan evaporation using standard meteorological data: a pragmatic synthesis, Hydrol. Earth Syst. Sci., 17, 1331-1363, 2013.*

*Maxino, C. C., McAvaney, B. J., Pitman, A. J. and Perkins, S. E.: Ranking the AR4 climate models over the Murray-Darling Basin using simulated maximum temperature, Minimum temperature and precipitation, International Journal of Climatology, 28, 1097-1112, 2008.*

*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, Geophysical Research Letters, 33, L08708, doi:10.1029/2006GL025779, 2006.*

*Murphy, J. M., Sexton, D. M. H., Barnett, D. N., Jones, G. S., Webb, M. J., Collins, M. and Stainforth, D. A.: Quantification of modelling uncertainties in a large ensemble of climate change simulations, Nature, 430, 768-772, 2004.*

*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.*

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

*Pierce, D.W., Barnett, T.P., Santer, B.D. and Gleckler, P.J.: Selecting global climatic models for regional climate change studies. Proceedings of the national Academy of Sciences of the United States of America, 106, 21, 8441-8446, 2009.*

*Raju, K.S. and Kumar, D.N.: Ranking of global climate models for India using multicriterion analysis. Climate Research, 60, 103-117, 2014.Raisanen, J.: How reliable are climate models? Tellus, 59A, 2-29, 2007.*

*Reifen, C. and Toumi, R.: Climate projections: Past performance no guarantee of future skill? Geophysical Research Letters, 36, L13704, doi:10.1029/2009GL038082, 2009.*

*Reichler, T. and Kim, J.: How well do coupled models simulate today's climate? Bulletin of the American Meteorological Society DOI:10.1175/BAMS-89-3-303, 303-311, 2008.*

*Sheffield, J., Barrett, A.P., Colle, B., D. Fernando, D.N., Fu, R., Geil, K.L., Hu, Q., Kinter, J., Kumar, S., Langenbrunner, B., Lombardo, K., Lindsey N. Long, L.N., Maloney, E., Mariotti, A., Meyerson, J.E., Mo, K.C., J. Neelin, D., Nigam, S.,Pan, Z., Ren, T., Ruiz-Barradas, A., Serra, Y.L., Seth, A., Thibeault, J.M., Stroeve, J.C., Yang, Z. , and Yin, L.: North American Climate in CMIP5 Experiments. Part I: Evaluation of Historical Simulations of Continental and Regional Climatology. Journal of Climate, 26, 9209-9245, 2013.*

*Shukla, J., DelSole, T., Fennessy, M., Kinter, J. and Paolino, D.: Climate model fidelity and projections of climate change. Geophysical Research Letters, 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, Climate Change, 102, 377-393, 2010.*

*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, Journal; of Hydrometeorology, 13, 140-154, 2012.*

*Suppiah, R., Hennessy, K. L., Whetton, P. H., McInnes, K., Macadam, I., Bathols, J., Ricketts, J. and Page, C. M.: Australian climate change projections derived from simulations performed for IPCC 4th Assessment Report. Aust. Met. Mag, 56, 131-152, 2007.*

*Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, Journal of Geophysical Research, 106, 7183-7192, 2001.*

*Trenberth K. E., Smith, L., Qian, T., Dai, A. and Fasullo, J.: Estimates of the global water budget and its annual cycle using observational and model data, Journal of Hydrometeorology, 8, 758-769, 2007.*

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

*Watterson, I. G.: Calculation of probability density functions for temperature and precipitation change under global warming. Journal of Geophysical Research 113, D12106, doi:10.1029/2007JD009254, 2008.*

*Whetton, P. H., McInnes, K. L., Jones, R. N., Hennessy, K. J., Suppiah, R., Page, C. M., Bathols, J. and Durack, P. J.: Australian climate change projections for impact assessment and policy application: A review., CSIRO Marine and Atmospheric Research Paper 1, 2005).*

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

.

*Table A GCM assessment climate measures based on literature review*

| Reference | Mean of P | Variance of P | Rel. bias or anomaly of P | Mean of T | RMS or similar metric | Spat. Correl. | Taylor plots | Space-time cmsd* | Climate prediction index | Bayesian approach | PDF, entropy & related measures | Skill score including demerit points | Var. converg. Score+ | Signal noise ratio | ENSO |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Legates & Willmott (1992) | yes | | | | | | | | | | | | | | |
| Taylor (2001) | | | | | | | yes (abs)^ | | | | | | | | |
| Lambert & Boer (2001) | | | | | | | | yes | | | | | | | |
| Boer & Lambert (2001) | | | | | | | | yes | | | | | | | |
| Covey et al. (2003) | | | | | | | yes (norm)# | | | | | | | | |
| Murphy et al. (2004) | | | | | | | | | yes | | | | | | |
| Whetton et al. (2005) | | | | | | | | | | | | yes | | | |
| van Oldenborgh et al (2005) | | | | | | | | | | | | | | | yes |
| Bonsal & Prowse (2006) | | | | | | | yes (abs) | | | | | | | | |
| Johns et al. (2006) | | | | | | | | | yes | | | | | | |
| Min & Hense (2006) | | | | | | | | | | yes | | | | | |
| Shukla et al. (2006) | | | | | | | | | | | yes | | | | |
| Raisanen (2007) | | | | | yes (abs) | yes | | | | | | | | | |
| Perkins et al. (2007) | | | | | | | | | | | yes | yes | | | |
| Suppiah et al. | | | | | yes | yes | | | | | | | | | |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (2007) | | | | | (abs) | | | | | | | | | | |
| Reichler & Kim (2008) | | | | | yes (rel)* | | | | | | | | | | |
| Gleckler et al. (2008) | | | | | yes (rel) | | yes (norm) | | | | | yes | | | |
| Watterson (2008) | | | | | | | | | | | yes | | | | |
| Maxino et al. (2008) | yes | | | yes | | | | | | | yes | yes | | | |
| Johnson & Sharma (2009a) | | | | | | | | | | | | | yes | | |
| Johnson & Sharma (2009b) | | | | | | | | | | | | | yes | | |
| Reifen & Toumi (2009) | | | | | yes (rel) | | | | | | | | | | |
| Chiew et al. (2009) | | | | | yes (abs) | yes | | | | | | | | | |
| Pierce et al. (2009) | | | | | | | | yes | | | | yes | | | |
| Smith & Chandler (2010) | | | yes | | yes (abs) | yes | | | | | yes | yes | | | yes |
| Knutti et al. (2010) | | | | yes | Yes (abs) | yes | | | | | yes | | | | |
| Falloon et al. (2011) | yes | | yes | | yes (rel) | | | | | | | | | | |
| Hagemann et al. (2011) | yes | Yes as figure | yes | | | | | | | | | | | | |
| Sperna Weiland et al. (2012) | | | yes | | | | yes (abs) | | | | | | | | |
| Sheffield et al. (2013) | | | yes | | yes (abs) | yes | yes (norm) | | | | | | | | |
| Baker and Huang (2014) | | | yes | | | | | | | | | | | | |
| Heo et al. (2014) | | | | | | | yes (norm) | | | | | | | yes | |
| Kumar et al. (2014) | | | | | | | yes (norm) | | | | | yes | | | |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Raju and Kumar (2014) | | | yes | | yes (norm) | | | | | | | | | | |
| Number of references | | | | | | | | | | | | | | | |
| 34 | 4 | 1 | 7 | 2 | 11 | 6 | 8 | 3 | 2 | 1 | 6 | 7 | 2 | 1 | 2 |

** cmsd: climate mean square difference; ^ (abs): based on absolute data ; # (norm): based on normalized data; * (rel): relative; @var. converg.: variable convergence*

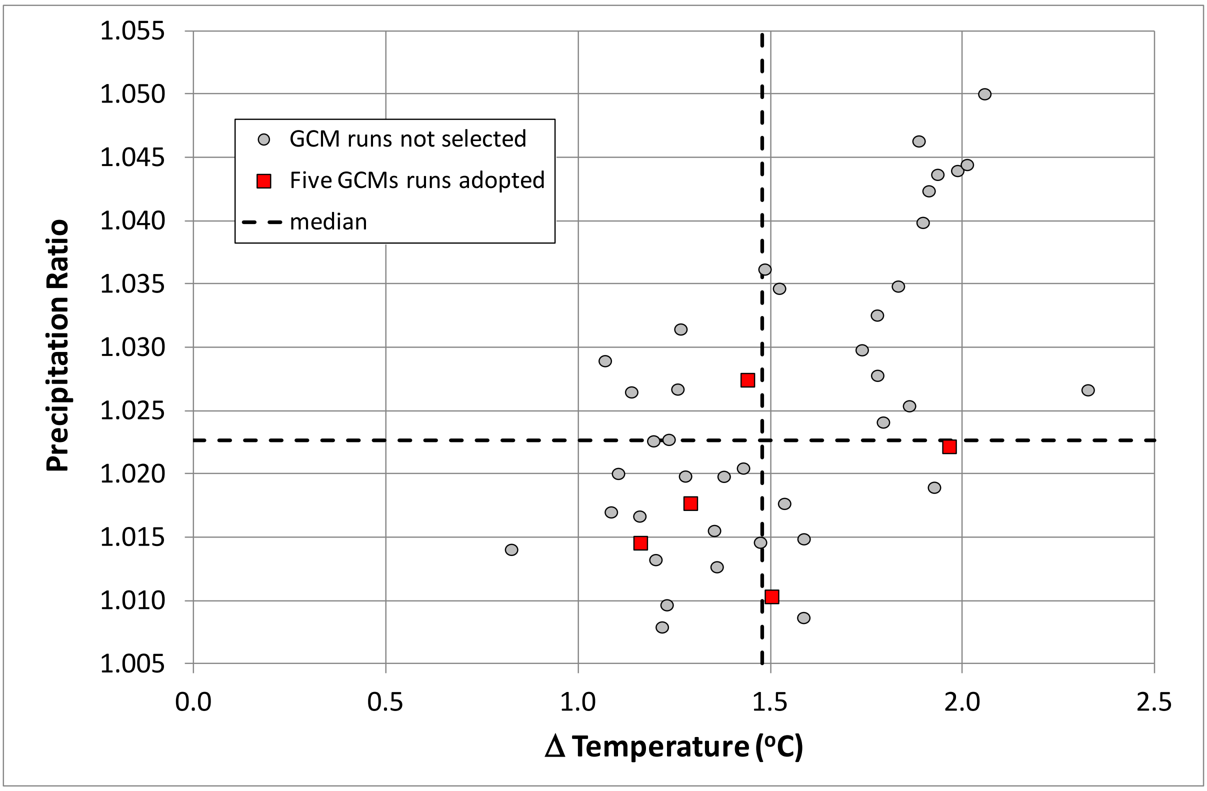


*Figure A Ratio of 2015-2034 to 1965-1994 mean annual precipitation compared with the increase in mean annual temperature ([2015-2034] – [1965-1994]) for the selected five GCMs runs compared with the 23 CMIP3 GCMs including all ensemble members for the global land surface*