

Manuscript: hessd-11-4531-2014: Uncertainty in runoff based on Global Climate Model precipitation and temperature data – Part 1: Assessment of Global Climate Models

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.

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. 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).

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.

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.).

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. 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.

Minor Comments

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

p. 4534 – line 1

... to *characterize* the ...

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.

p. 4545 – line 8

... data set *is* presented ...

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.

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.

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. 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).

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.

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.

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.