Reply to review of M. Renner:

We thank the reviewer for his valuable comments. A general point of his comment is that he suggests adding much more information from previous work that the present manuscript is based on. We have to choose between providing important information that is necessary for understanding the manuscript and avoiding repeating information that has already been thoroughly discussed in previous publications. In the manuscript we followed the latter strategy, especially with regard to the methods used to produce the multi-model ensemble. It should be noted that the present work is one of the main outcomes from the WATCH project, in which the bias correction method was developed (Piani et al. 2010), the bias-corrected GCM data were created, impacts of the bias correction were discussed (Hagemann et al. 2011), and the GHMs have been compared within the WaterMIP model intercomparison project (Haddeland et al. 2011). Such a large body of work cannot be fully encapsulated in our manauscript, but we provide references so that the reader can get extra details where desired. In Haddeland et al. (2011) differences between the GHMs and their impact on the hydrological simulations have been discussed. Especially with regard to the topic of bias correction, discussing pro and cons is currently a hot topic, and we refer to related literature in sect. 2.1. We think that it does not add much value to the current study when these discussions on pro and cons are merely repeated here. But we agree that we should point the existing literature more specifically in respect to some of the questions raised by the reviewer. We will also include a paragraph in the conclusions and discussion section that deals with this topic.

Major remarks:

- 1.1 Does bias-correction result in higher ensemble mean changes in ET and R? On P1330L11 it is noted that bias-correction reduces the spread of the change signals. Further, the GCM-GHM model chain results in larger average changes, while the noncorrected GCMs show less significant changes. Hence, the GCM-GHM simulation results suggest more confidence in the change signal. This is an important result, which should be taken with care with regard to the inherent assumptions of the general approach. For me the following questions arise:
- To which extent are the changes in ET and R caused by the bias-correction or the use of different GHMs?

A direct comparison of the simulated GHM changes in ET and runoff from uncorrected and bias corrected GCM output cannot be made for the whole GHM ensemble as most of the GHMs did not produce simulations with the uncorrected GCM output. (These costly simulations were not part of the work in WATCH.) By comparing the changes for two of the GHMs in Hagemann et al. (2011), for most of the large catchments considered, the two annual mean GHM climate change signals in ET and runoff differ more than the mean signals obtained with and without bias correction. In addition, the parameterizations of evapotranspiration and runoff vary substantially between the GHMs (Haddeland et al. 2011, we will include a table for the 8 GHMs as mentioned below), and the complicated interactions between the various processes make it infeasible to explain the causes of many simulation differences in detail, as noted in previous model intercomparisons (e.g., Koster and Milly 1997). We will include this information in the conclusions and discussion section.

- Does the bias-correction procedure retains consistent mass and energy balances?

The bias correction does not directly impact the surface water and energy balances of the GHMs applied. The GHM water balances are generally closed. Imbalances in the long–term water balance equation are caused by changes in the water stores between the start and end of the simulation and, for the JULES model, by non-conservation of water for lake surfaces (Haddeland et al. 2011). Most of the GHMs do not compute an energy balance, but as bias-corrected temperatures and GCM radiative fluxes are used as model forcing, it can be assumed that the GHMs calculate energy fluxes consistent to the forcing. Otherwise spurious trends in some surface state variables may occur, which we are not aware of.

- • To which extent does bias-correction changes hydrological behavior? For example changes in precipitation falling as snow or rain.

The fraction of precipitation falling as snow is not changed by the bias correction and, thus, remains at the GCM simulated value (Hagemann et al. 2011). Certainly, the bias correction changes the GHM simulations (but this is a desired effect of bias correction), but not the handling of physical processes by the GHMs. This is part of the BC method described in Piani et al. (2010) and the practical application in Hagemann et al. (2011).

- • Please elaborate the statement on P1330L21 "the consistency between variables is not necessarily the case due to the bias correction".

Precipitation and temperature are corrected independently. Several studies, such as that of Berg et al. (2009), have shown that daily precipitation shows some scaling with temperature so that future improvements of the bias correction method may be achieved with multivariate approaches that take these dependencies into account (Hagemann et al. 2011).

In addition, other GCM variables than precipitation and temperature are not corrected, which certainly introduces some inconsistencies, e.g. for the near surface air humidity used by some of the GHMs as forcing (see also response to major remark 1.2).
But we also mention on p. 1333 – line 3:
"Note that the relative values of projected hydrological change are very similar if also other GCM variables are bias-corrected (Haddeland et al., 2012)."
Thus, it can be assumed that the impact of these inconsistencies is generally rather small.
We will point to the related references more specifically.

- • Why and how do you arrive at that statement P1330L24: "...these results show another advantage of the chosen model setup compared to the direct use of GCM data for impact assessment, ..."?

GCMs cannot be used for projection impact studies without some form of bias correction. When only the climate change signal is taken from simulations, instead of the raw GCM output, this is tantamount to applying a bias correction only to the mean. That said, it is a matter of scientific debate whether the bias correction is adding or uncovering another level of uncertainty that is related to the uncertainty induced by the choice of the GCM (Hagemann et al. 2011). In this respect, we cite several references where this has been discussed.

As mentioned on p.1330, the spreads and associated uncertainties that are caused by GCM biases in the original GCM output can be reduced with the chosen model setup. We consider this as a positive point. To make this clearer, we will rephrase the sentence as:
"   these results show a beneficial characteristic of the chosen model setup …"

- 1.2 How useful are decoupled ET estimates in a climate simulation setting?

As mentioned above, the GHM surface water balances are generally closed. We agree with the reviewer that the GHM ET does not feed back to the atmosphere, thus it doesn't impact GCM precipitation or near surface specific humidity. While the first is corrected, the latter is still taken as uncorrected GCM values, which certainly introduces an inconsistency in the atmospheric water balance. On the other hand, as also mentioned above, the current GCM output is generally not directly applicable for impact studies, mainly due to the large biases in precipitation and associated biases in surface hydrology, i.e. runoff, ET. These biases are impacting the GCM signals in addition to the use of different GCM parameterizations. We would add this discussion in a revised version of the paper.

Please also note that when running a GCM, bias correction must be avoided in order to keep water and energy balance closed. However, generally, future runoff change is widely concerned by public. In order to make simulated runoff comparable to observation (or real world), removing the bias of precipitation and temperature of GCMs is virtually indispensable, which eventually leads to collapse the energy and water balance of GCMs. That's why impact models are usually driven by GCM output after some kind of bias correction and/or downscaling have been performed.

- 1.3 Hydro-climatological assessment

We felt that adding regional figures/analyses for different hydroclimates would in fact increase the number of panels/figures, so that we refrained from doing this and would like to keep the global maps. As in our study we are dealing with the impact of climate change on hydrological fluxes and water resources, we don't need to separate these effects from changes in catchment properties, such as is done in e.g. Roderick and Farquhar (2011). Anyhow, we believe that an application of the framework of Renner and Bernhofer (2012) may be a valuable future extension of our study by linking some of these changes to the aridity index. Thus, we will point to this in our conclusions section.
While we will acknowledge in the conclusion that individual catchment properties can affect the magnitude of hydrological response (with reference to Arora, 2002; Roderick and Farquhar, 2011; Renner and Bernhofer, 2012), the present maps facilitate comparisons with other global-scale assessments (which, unfortunately, have not all applied the approach suggested by the reviewer; e.g. http://www.sciencedirect.com/science/article/pii/S0959378011000161). The novelty of our paper is in the consideration of hydrological model uncertainty, through the application of multiple hydrological models. To this end, readers may wish to compare our ensemble mean maps with those presented elsewhere for individual hydrological models (e.g. aforementioned references)".

- 1.4 Temporal scale of changes daily vs. annual

We have no objections against this point, so we will change the unit if this is consensus by the reviewers.

Minor remarks:

- abstract, I miss the mentioning that bias-correction is employed

We will add this.

- P1323L5: please be more detailed on the differences between hydrological models

See 3rd response below

- section 2.1 models: please give an brief overview with respect to the main differences of the GCM models; this task should not be left for the reader

The GCM data are described in Hagemann et al. (2011).., and we note on p. 1329: "However, the chosen GCMs belong to different model families and cover some range in projected precipitation change among the CMIP3 (Meehl et al., 2007; see also Sect. 5) ensemble (Mason and Knutti, 2011). The selection of GCMs for this study was imposed by the availability of climate model data necessary to force the GHMs. A respective analysis of the original GCM results over Europe was provided by Hagemann et al. (2008)."

We think the included information is sufficient for this paper. Paying attention to the reviewer's remark, we will move this statement to Sect. 2.1.

- bias-correction: please be more detailed on the pitfalls of bias-correction; this is a major modeling step and should be clearly reflected for the reader

See main response to reviewer's comment above.

- please be more detailed on the GHMs; maybe a table for input data, ET formulation and other important processes would be beneficial; As the use of 8 different models is the stated novelty of this manuscript, their differences should be discussed.

Even though we wanted to minimise the repetition of previous work, we agree with the reviewer that some more information on the GHMs may be helpful. Thus, we will include parts of Table 1 from Haddeland et al. (2011).

- section 2.2: so ET is derived by the GHMs; Is there a check for consistency of the (surface) energy balance and water balance within the GCM-GHM model chain?

See response to 2nd major remark above.

- P1326L6ff: for the precipitation change results, is the GCM output shown, or the bias-corrected output?

Bias corrected GCM output. We will add the information in the text and Figure caption.

- P1326L20 To which extent are ET and R driven by precipitation changes? How important are changes in other forcing variables such as net radiation and temperature?

The results of Haddeland et al. (2011) indicate that, globally averaged, the majority of the interannual variation in precipitation feeds directly through to the runoff and that the evapotranspiration is constrained by other atmospheric factors such as temperature, radiation, and humidity. The same is valid for the future changes in runoff and ET, whereas ET will also be affected by precipitation changes in transitional wet regions where the availability of soil moisture directly affects the evaporative fraction (Seneviratne et al. 2010). We will add this information in the discussion of results.

Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., Orlowsky, B., and Teuling, A. J.: Investigating soil moisture–climate interactions in a changing climate: A review, Earth-Science Reviews, 99, 125-161, 10.1016/j.earscirev.2010.02.004, 2010.

- What are the spatial patterns, hydro-climatological patterns of the changes?

As mentioned above, we felt that adding regional figures/analyses for different hydroclimates would increase the number of panels/figures, so that we refrained from doing this and would like to keep the global maps. Fig. 2a and b show the maps of mean changes in ET and runoff, respectively. We will increase the map sizes to allow a better identification of the spatial patterns of change. In addition, we are already describing the most noticeable changes in Sect. 3. p. 1326 and 1327.

- P1327L8-23: In general I like the idea of separating the differences induced by the climate or the hydrological models. Further, please indicate how exactly you defined the coloring.

In this case the maximum spread is defined as the largest spread (SD) coming from one of the three sources (GCM, GHM, scenario). In order to avoid misunderstanding with the maximum absolute spread (not the SD) we will rephrase the text by replacing “maximum” with “largest”.

- P1328L8-12 “Over the high latitudes ...” unclear sentence

We will rephrase the sentence as:
“Over Siberia, the scenario spread for runoff seems to be related to the combined effect of scenario spreads in precipitation and ET (Fig. S1 in the Supplement).”

- P1328L16-18: First, it is stated that maximum spread is important to judge the robustness for average changes. However, in Fig. 4 the standard deviation is used. Hence, I would like to see the maximal spread of projected changes and not the sd.

It is rather common to show the SD for quantifying the spread in ensemble approaches. By showing the maximum spread, results may become strongly blurred by single outliers. In order to avoid misunderstanding with the maximum absolute spread (not the SD) we will rephrase the text by replacing “maximum” with “largest”.

- P1329L19-23: the text only repeats the colors of the corresponding figure

We will remove the color information from the text and rephrase the text as:
“Here, areas indicating larger changes in the GCM-GHM ensemble exceed areas with larger changes in the original GCM output for both ET and runoff. Areas where the sign of change differs are relatively scarce.”

- P1330 Original GCM output means that there is no bias correction? If so, this should be noted explicitly.

Yes, we will add “original (uncorrected) GCM output” in line 2, p. 1330.

- Fig. 1 Is that bias corrected output? Panel b) by using the CV, dry regions are overly emphasized, while for more wet regions the map suggests high confidence.

Yes, bias corrected output. We will add the information in the figure caption. Despite these characteristics of the CV, we think it is easier (and also common) to look at than the spread represented by the absolute standard deviation (SD). Note that for showing absolute SDs, wet regions would stick out more due to the larger means and commonly larger associated SDs, and thus these would be overly emphasized.

- Fig.: 2 The 8 global maps are hardly readable. Especially the legend is too small. The authors should remove some of these maps or make separate figures, if important.

We will enlarge the maps, remove the southern ocean part of the maps and increase the size of the legend. We will also separate the panels e-h from Fig. 2 into a new figure.

- Fig. 4 try to increase the readability of the figures (colors, too much overlay). Please display the full spread and not only the standard deviation (requires nonskewed samples).

We will try to improve the readability of Fig. 4. But for reasons mention above (outlier blurring results in full spreads) we want to keep the standard deviations.

- Fig. 5 I would rather like to see a scatterplot comparing the changes / or absolute ensemble means than the maps. It is enough to state that spatial patterns are similar.

We will make some scatterplots, and then we will judge which kind of presentation (map/ scatterplot) gives more useful information to the reader.

- Fig.6 Similar as Fig.5 and unclear units used in the legend

No units are used. We will remove the numbers in the legend.