**Reply to reviewers' comments**

We thank all reviewers for their thorough reading of the manuscript and their helpful remarks that helped us to improve the manuscript. Our reply is structures as follows. First we start with a general reply to the reviewers' comments, which is followed by detailed responses to the remarks of the anonymous reviewers 1 and 2. Then, we provide some additional responses to our reply to the review of Mike Renner that we already published during the discussion phase of the paper. For completeness, we included this reply also in the end of this document.

**General reply to reviewers' comments**

As all reviewers addressed the issue of bias correction, we give a general response to these comments:

Instead of using a bias correction of GCM output, the better way to conduct hydrological projections would be that the GCMs reproduced the current climate (precipitation, temperature and other fields) better than what they currently do. However, it will take a while before the GCM outputs are "good enough" to be used directly in impact studies, and we think hydrological projections should be made even though bias corrections are needed. We agree that this (pragmatic) method includes drawbacks and uncertainties, and we do need to communicate these issues to the user of the results. This is a discussion of a more principal character than can be discussed fully in the paper in question. We have, however, included a short discussion in the current paper, while referring to the other papers (e.g. Haerter et al. 2011; Ehret et al., 2012) for an in-depth discussion of the issue. We still keep the discussions fairly short, though, as we do not think an extensive discussion of the topic "bias correction" is suited for this paper.

We added the following text to Sect. 2.1 after the introduction of the 3 GCMs:
*GCMs exhibit a number of significant systematic biases in their ability to simulate key features of the observed climate system (Randall et al. 2007). Despite the biases, the IPCC concludes that there is still considerable confidence that climate models provide credible quantitative estimates of future climate changes (Randall, 2007). However, until GCMs perfectly reproduce the current climate, GCM outputs cannot be used directly in hydrological impact studies without some form of bias correction. When uncorrected GCM output is used as input to hydrological simulations, the resulting amount and seasonal distribution of runoff may be far from observations, for example see Haddeland et al. (2012), Wood et al. (2004) and Sharma et al. (2007). Consequently, a* statistical …

In addition we added the following text to Sect. 5

Note that the *bias correction also adds uncertainty to the projections. 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 (such as presented by Piani and Haerter 2012) that take these dependencies into account. In addition, GCM variables other than precipitation and temperature are not corrected, which potentially introduces inconsistencies between variables, e.g. between the near surface air humidity and temperature used by some of the GHMs as forcing. However Haddeland et al. (2012) found that the* relative values of projected hydrological change are very similar if other GCM variables are *also* bias-corrected. *Thus, it can be assumed that the impact of these* *inconsistencies is generally rather small. Another uncertainty inherent to the chosen model setup is that the GHM ET does not feed back to the atmosphere, hence, it does not impact GCM precipitation or near surface specific humidity.*

*Despite these uncertainties and inconsistencies, there are currently not many alternatives to this approach for hydrological impact assessments. As mentioned in Sect. 2.1, output from the current generation of GCMs is generally not directly applicable for impact studies, mainly due to the large biases in precipitation and associated biases in surface hydrology (runoff, ET). These biases impact the GCM signals, as do the different GCM parameterizations, and thus* lead to uncertainties in projected changes in terrestrial components of the hydrological cycle that are larger than in the model setup presented here. These *differences can* also lead to different climate change signals, which in the present study, are generally weaker *in the uncorrected GCM output*, but this may be a characteristic of the chosen 3 GCMs.
:
:
*The impact of the bias correction (see above) on the projected changes is probably smaller than that caused by the 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. An indication of the size of the effect can be found by comparing the changes for two of the GHMs in Hagemann et al. (2011) for which, 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.*

**Reply to review of anonymous reviewer 1:**

We thank the reviewer for his valuable comments.

*The difficulties and challenges … …..*
We included more discussions about these difficulties and challenges, especially on the use of bias correction and its implications (some effects of the bias correction as well as the necessity of its usage for impact studies). See general response above.

*Novel but unique are the 8 GHM and the purpose of determination the variability of the different model accesses. That could be better highlighted in the paper.*
In addition to modifications associated with our response to the reviewer's major and minor remarks, we added the following text to the abstract:
… water resources. *This multi-model ensemble allows to investigate how the hydrology models contribute to the uncertainty in projected hydrological changes compared to the climate models.* Due to …
:
for hydrological change, *and that the spread resulting from the choice of the hydrology model is larger than the spread originating from the climate models over many areas.* But there

*The specified and described variability is quite hard to evaluate. For me sometimes important aspects and information in the explanation are missing, skipped or too short cited.*
We improved the manuscript according to the reviewers' remarks (see below and response to review of M. Renner).

*Many of the figures are too small sized …*
We enlarged the size of the too small sized figures (see also specific response to reviewers' remarks (M. Renner and reviewer 1).

**Major Remarks**

*2.1 First a table …*

We agree with the reviewer that some more information on the GHMs would be helpful. Thus, we included parts of Table 1 from Haddeland et al. (2011), will add the following text to section 2.1:
*The major model characteristics are listed in Table 1. The GHMs differ in their evapotranspiration and runoff schemes, and the differences in model parameterizations are to some extent reflected in the forcing variables that are used by each (Table 1). For associated model references and validation of GHM model results using quasi-observational forcing data*, see Haddeland …
.

*2.1 Can we guarantee that within the bias correction consistency and conservation aspects are sufficiently regarded? How are …… This paper would win with a more detailed and critical discussion to where the uncertainties come from.*

We added a discussion on bias correction and consistencies in Sect. 5 (see general response above)

*2.2. I recommend a more transparent analysis and discussion to the obvious BC impact.*

See general response stated above.

*2,2 – P1329 L25ff.*
Modified :text
… the projected A2 changes from the original GCM output are often *significantly lower, up to 50-70% less, than the respective changes projected by the GCM-GHM ensemble.*

Added text:
*This may partly be due to the small sample of 3 GCMs that with regard to the calculation of ET and their projected changes may not cover the full space of possible model solutions.*

In addition, see general response above.
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.

2.2 P1330
Modified text:

Note that absolute standard deviations *are shown in Fig. 8 (previously 4) while the spreads represented by the CV in the Figs. 4 (previously 2), 7, 11 and S2 (previously 5) are relative values. Thus, Fig. 8 shows catchment averaged absolute standard deviations of the original GCM output that* are partially smaller than the corresponding standard deviations in the GCM-GHM ensemble due to the choice of the GCM, *especially for runoff*. But as the projected mean changes in the direct GCM output are mostly weaker than for the GCM-GHM ensemble (see above), the associated *relative values become much larger for the direct GCM output. Thus, the* spreads represented by the CV are often larger for the direct GCM output than their GCM-GHM counterparts (Fig. *11*).

*2.3. Natural climate variability*

As has been shown in previous studies nature climate variability is usually the smallest source of uncertainty if you consider climatological time scales and average over larger areas such as catchments (e.g. Deque et al.  ). In addition, to bias correct the other two ECHAM5 ensemble members and force the GHMs with this bias corrected GCM output is simply beyond the scope of the present study.

*2.4. Unit of changes*

We recalculated all projected annual changes into the unit mm/year [mm/a] and updated all figures accordingly.

*2.4. seasonal perspective*

We added a new Figure 13 with seasonal changes in runoff obtained from the GCM-GHM ensemble and the following text to Sect. 4:

*The analysis presented above has been conducted on the annual scale, but in some regions available water resources are also affected by seasonal changes. Figure 13 shows the projected changes in runoff per season that can be compared to the annual mean changes shown in Fig. 3b. Regions that might experience a seasonal reduction in runoff that is more severe than in the respective annual mean are potentially affected by a seasonal reduction in available water resources. These regions comprise parts of southern Africa in DJF (Fig. 13a), central eastern South America, eastern US and eastern Europe in MAM (Fig. 13b), almost the whole of Europe and western Siberia as well as western US and southern and western Canada in JJA (Fig. 13c), and north-western South America in SON (Fig. 13d).*

*2.4. interannual variability*

Even though we believe that it is interesting to consider how changes in inter-annual variability may affect available water resources, we think that this is a much broader topic that is beyond the scope of the present study. Such changes should be considered together with the change in drought characteristics (especially frequency and duration). To notify their potential importance we added the following text to Sect. 4:

*It has to be noted that even if the long-term mean annual change in annual water resources may be quite small for some regions, they might be strongly affected by changes in inter-annual variability and the occurrence of droughts. An analysis of these effects is beyond the scope of the present study, but it is an important topic for future studies. In this respect, Prudhomme et al. (2013) investigated the impact of climate change on hydrological droughts.*

*2.5. …I missed a section with essential information to GCM and GHM calibration and validation.*

We think that a detailed section on these issues is not necessary within the present study. On one hand, most of the GHMs were not calibrated specifically to the forcing data we applied. Instead, some of the GHMs were calibrated against observed global datasets in independent calibration exercises carried out by each model author(s). For other GHMs, specific processes parameterizations were tuned on the local or regional scale for specific regions, which then are applied globally. Thus, GHM calibration issues are not really relevant in the present study. None of the GCMs is calibrated either. On the other hand, validation information has been published in previous studies. For the GCMs, a respective analysis of the original GCM results over Europe was provided by Hagemann et al. (2008). This information is now moved to Sect. 2.1.

With regard to the GHMs, we point to this now by using the following text added to Sect. 2.1:
*For associated model references and validation of GHM model results using quasi-observational forcing data, see Haddeland et al. (2011). The variability among the GHM results forced with bias-corrected GCM output and associated runoff biases for the control period 1971-2000 are in accordance with the validation shown in Haddeland et al. (2011).*

*2.5. Do the GHM simulations with GCM input match the observed hydrological data?*
We will add the following text to Sect. 2.1:
*The variability among the GHM results forced with bias-corrected GCM output and associated runoff biases for the control period 1971-2000 are in accordance with the validation shown in Haddeland et al. (2011).*

*2.5.ET is somehow part of GCM and GHM: congruent differences?*
ET both varies within the GHM ensemble (see also previous remark) as well as within the uncorrected output of the 3 GCMs for the control period. Consequently ET also varies in the associated projected changes. Comparing multi-model mean ET from the two different ensembles shows positive and negative differences over various regions for the control period. It seems that GCM-GHM ET tends to be smaller than the ET from original GCM output except over the Tropics where the first tends to be larger.

**Minor Remarks**

P1324 L17
Sentence will be rephrased to:
… observed (Weedon et al., 2011) and *the original GCM* data and then applied

P1324 L17-25
Please see general response (see above) with regard to the issue of bias correction.

P1325 L7
A schematic figure is added (new Fig. 1). We also added the following text in the beginning of Sect. 2.2:
*Figure 1 presents an overview on the global modelling chain employed within the WATCH project (cf. sect. 2.1).*

P1325 L12
Modified text:
P1325 L12: For the *high emission* A2 scenario, simulations …
P1325 L14: For the *low emission* B1 scenario, 18 …

P1325 L15
Modified text:
… hydrological variable *(evapotranspiration and runoff)* was…

P1326 L10
We recalculated all projected annual changes into the unit mm/year [mm/a] and updated all figures accordingly. Also, we modified the text to:
These future changes in precipitation *show* similar *patterns to as* to…

P1326 L17
We added the following sentence in the beginning of Sect. 3.
*In the following, projected changes are associated with the A2 scenario if not mentioned otherwise.* According to the results ….

P1327 L12 and P1327 L13
Modified text (Fig. 2e-h changed now to Fig. 4a-d):
… uncertainty *(Fig. 2g)* where the spread originating from the GCMs is rather low *(Fig. 2e)*. For runoff, the CV values representing the GCM spread *(Fig. 2f)* are often comparable to *those for* the GHM spread *(Fig. 2h)* even though…

P1327 L25
Figure S1 is included in the supplementary material that can be downloaded at the ESDD web page where this paper is located. We modified the text:
… scenario (*supplementary Fig.* S1).

P1327 L24+25 & P1328 L1+2
We kindly ask the reviewer to take a look into the supplementary material where the necessary information is provided in Fig. S1.

P1328 L16ff/Fig. 1342
Modified caption text:
… *about* the A2 mean *(i.e. ±Std.)* over …

*Improve and better explain Fig.4*
We separated Fig. 4 (now Fig. 8) into two figures 4a and 4b to increase the size (the ESD page format has disadvantages compared to A4 in which we prepared the manuscript.), and in the figure caption, we replaced spread by standard deviation to avoid confusion with the spread defined as the CV. We also added text to Sect. 3:
*Here, the spreads are expressed by the absolute standard deviations about (±) the respective mean change so as to allow direct comparisons between them.*

P1329 L14-16
Modified text:
… with the *uncorrected* climate model …

P1332 L20ff
Please see general response (see above) with regard to the issue of bias correction.

P1333 L1
We removed ‘on an individual basis’

P1322 L10
In the abstract, this is a summarizing statement on model agreement. From Fig. 2, also the ensemble mean changes over different areas can be obtained. To make this clearer, we modified the text:
… indicative of higher confidence *in this ensemble mean signal*.

P1340 Fig. 2
We enlarged the maps of Fig. 2 (now Fig. 3), removed the southern ocean part of the maps and increased the size of the legend. We also separate the panels e-h from Fig. 2 into a new figure (Fig. 4).

P1344 Fig. 6
We will remove the numbers in the legend as no units are used. We also will separate the figure into two figures to shorten the caption.
Caption of new Fig. 10 comprising the upper two panels of Fig. 6:
*Comparison of mean A2 evapotranspiration (left panel) and runoff (right panel) changes (2071–2100 compared to 1971–2000) projected by the GCM-GHM ensemble and the original uncorrected GCMs. Areas are indicated where the projected decreases and increases are larger in the GCM-GHM ensemble (red and blue, respectively) than in the original GCM output and vice versa (orange and turquoise, respectively), as well as areas where the sign of projected change differs between them (green).*

Caption of new Fig. 11 comprising the lower two panels of Fig. 6:
*Comparison of the spreads associated with the mean A2 evapotranspiration (left panel) and runoff (right panel) changes due to the choice of the GCM for the GCM-GHM ensemble and the original uncorrected GCMs. Areas are indicated where the CV is larger for the GCM-GHM ensemble (blue) or for the original GCM output (red).*

**Reply to review of anonymous reviewer 2:**

We thank the reviewer for his valuable comments.

*Which scheme is used to model ET?*

As mentioned in our reply to the other reviewers’ comments, we will include parts of Table 1 from Haddeland et al. (2011). We will add the following text, and the Table, to section 2.1:

*The major model characteristics are listed in Table 1. The GHMs differ in their evapotranspiration and runoff schemes, and the differences in model parameterizations are to some extent reflected in the forcing variables that are used by each (Table 1). For associated model references and validation of GHM model results using quasi-observational forcing data*, see Haddeland …

*Which datasets are used (such as for the soil characteristics)?*

The models use their default soil datasets. We have opted not to include information about the datasets, for two reasons: 1) for many models this is difficult to summarize in a Table, or, if summarized the values provide limited useful information (e.g. the soil depth in WaterGAP varies between 0.1 and 4 meters across the globe), and 2) there are other equally important characteristics, such as vegetation characteristics, which are even harder to summarize in a Table. We concluded that Table 1 (now inserted as mentioned above) shows the most important information and that further details should be found in the paper(s) cited for each model. This difference will likely affect the final spread somewhat, but is an important part of the analyses of future projections/spread.

*How are the GHMs calibrated?*
Calibration issues are not really relevant in the present study since most GHMs did not use calibration.

*More critical analysis of observed patterns – distinguish between water- and energy-limited basins.*
We added new figures 6 and 7 and included the following text in Sect. 3:
*We now consider whether the projected changes and associated spreads behave differently in dry (humid) areas where ET tends to be limited primarily by the availability of moisture (energy). These areas are represented by low (high) values of the ensemble mean runoff coefficient (runoff R divided by precipitation P) for the present day climate (Fig. 6a). Figures 6b and c show that humid areas expect increases in ET and runoff, while decreases in both variables occur only over some medium wet (e.g. Danube) to dry (e.g. Murray) areas (R/P < 0.6). Considering the spreads, it can be noted that for both ET and runoff the GCM spread tends to be larger for dry areas than for humid areas (Fig. 7 a, b). For the GHM spread, there is a less clear tendency, even though some larger CVs (CV > 1.7 for ET, CV > 1.2 for runoff) only occur over medium wet to dry areas (R/P < 0.5).*

**Additional response to the review of M. Renner:**

- 1.1 precipitation falling as snow or rain.

Even though the fraction of precipitation falling as snow is not changed by the bias correction, in practice it does change for most GHMs as many GHMs use total precipitation as input and use temperature to partition between rain and snow, so when temperature is bias corrected, fraction snow/rain will also be affected. But this is also a desired effect as wrong temperatures negatively impact the accumulation and melt of snow.

- 1.3

Even though we didn't add a full hydro-climatological assessment for different river basins, we now consider the different behaviour over dry and humid areas (see response to reviewer 2, last topic).

- Fig. 5

We moved the old figure 5 to the supplementary material and included some scatter plots (new Fig. 9). We modified the text accordingly.

Just for completeness:

**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 manuscript, 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 retain 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 modelling 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.