Response to Review Comments

Authors: J.C.S. Davie, P.D. Falloon, R. Kahana, R. Dankers, R. Betts, F.T. Portmann, D.B. Clark, A. Itoh, Y. Masaki, K. Nishina, B. Fekete, Z. Tessler, X. Liu, Q. Tang, S. Hagemann, T. Stacke, R. Pavlick, S. Schaphoff, S.N. Gosling, W. Franssen and N. Arnell

We thank all reviewers for their thorough reading of the manuscript and their constructive remarks which are helping us to improve the manuscript. Overall, we agree with the majority of the suggested changes, and propose to make changes based on most of the technical/specific comments. Further details of these are included below in response to each individual comment from the referees.

The main points from the two referees which we propose to address with considerable adjustments to the paper are:

a) Improving clarity of the main objective of the paper, mainly by increasing consistency between the abstract and conclusions so that the key message is clearly the difference in runoff change projections between the ecosystem and hydrological models included.
b) Giving a full explanation within the text of how and why the impact models have been grouped as they have been into the hydrological and ecosystem categories.
c) Explaining, in the Methodology, the reasoning for considering Giorgi regions rather than river basins to compare runoff projections from different model types.
d) Including validation of the modelled runoff.
e) Altering Figures 3 and 5 so that conclusions can be drawn from the graphs more easily. This will be done in an alternative way to that which was suggested (showing the ranges of models within each category), by instead keeping individual models' annual cycles plotted as lines, but with shading to show the ranges of the model types which we hope will make the graphs easier to interpret.
f) Removing water resources from the title of the paper.

Following are detailed responses to the two anonymous referees' comments point by point, which will outline our proposed revisions to the paper regarding each comment. Our response is in bold text.

Anonymous Referee #1

General Comments

In this paper, Davie et al. compare different projections of climate change effects on runoff for different global simulation models. They explicitly distinguish between hydrological models (7) and ecosystem models (4). The main difference between those 2 types of models is that the ecosystems models include vegetation dynamics and CO2 responses, whereas the hydrological models do not (line 25-27). The authors hypothesis is that the two types of models will show different runoff responses, due to the effect of vegetation processes that are considered in the vegetation models, but not in the hydrological models. Specifically, they want to test the effect of vegetation dynamics and of plant CO2 responses on simulated runoff changes In order to focus on the differences between the impact models, and not the whole range of uncertainty in climate change impacts, the authors choose to use only one RCP, and one GCM to force their models with. The authors conclude that ecosystem models generally tend to project larger increases and smaller decreases in runoff than the hydrological

models. This means that the processes that are unique to the ecosystem models tend to decrease evaporation. They plea for a wider range of impact models to be used in impact studies and planning water management than the hydrological models used usually.

The paper shows an important issue, namely that there is even more uncertainty about the effects of climate change on future runoff than previously considered, because the indirect effect of CO2 concentration which is usually not considered in hydrological models. However, I have some difficulties with the general objective of the paper, the way the results are presented and the conclusions drawn from the analysis, which I would like to see improved. Also, in many cases, the model results are presented and described as they are, without a further explanation of the observations. All in all I find the paper not yet readable and therefore not yet suitable for direct publication. While reading the paper, I got very much the impression that the paper was written quite quickly. I found it difficult to read and follow the line of thought, and it looks messy in places. For example, there are author notes still included in the text where references need to be inserted (see technical corrections)

These references were not yet available. They are ISI-MIP papers so very relevant, but the status has not changed for many so cannot currently be updated from Submitted or In preparation. Schewe et al. is now under review so this will be updated. An additional reference, to the ISI-MIP simulation protocol, will be included when information which will be included in Warszawski et al. (In preparation) is also included within it, as this is already available: Simulation Protocol of the Inter-Sectoral Impact Model Intercomparison Project (ISI-MIP) www.pik-potsdam.de/research/climate-impacts-and-vulnerabilities/research/rd2-cross-cutting-activities/isi-mip/simulation-protocol/isi-mipsimulationprotocol-as-of-05-november-2012-website (last accessed: 30th April 2013).

, there are terms introduced without explanation (e.g. the "minimal" settings simulations, the JULES model (r 23 p284), some more information on ISIMIP programme).

The "minimal setting" simulations will be explained in the Methodology with the following text: "The main simulations analysed in this study were ISI-MIP "minimal settings" simulations (Warszawski et al., In Preparation). For hydrological models, these were model runs with the "nosoc" SSP scenario, naturalised runs with no human impact, and for ecosystem models, these were runs with varying CO_2, with concentration specified by the RCP scenario. These were the runs provided to ISI-MIP by the largest set of impact models. The ecosystems models were also further analysed in order to investigate the importance of including individual processes, depending on the availability of sensitivity experiments. These included models with both a varying CO_2 run and a constant CO_2 run (concentration kept constant from 2000), and four JULES runs, which include the ISI-MIP minimal setting (Warszawski et al., In Preparation) and sensitivity experiment runs, with and without varying CO_2 and dynamic or non-dynamic vegetation. Table 2 gives an overview of the experiments analysed in this study. The aim of this study is to show that impact models including carbon dioxide impacts and/or vegetation dynamics may give differing projections to models not considering these, and therefore should be included in hydrological impact assessments." The JULES model is referenced;

however, due to the large number of models in the ISI-MIP ensemble, it would be infeasible to describe them individually within the text in much detail.

The main objective of the paper remains somewhat unclear. If the main objective is to draw general conclusions on the runoff response of ecosystem models vs hydrological models? Or to discuss each model individually?

The main objective is to draw general conclusions on the runoff response of ecosystem models vs. hydrological models. This will be explained better and made clearer, mainly by altering the abstract so that it focuses more on ecosystems vs. hydrological models rather than CO_2 effects.

The figures might be changed so that the conclusions that the authors draw from them are more convincing.

This is addressed below in response to the specific comments.

We propose to revise the text keeping in mind the general points made by the reviewers.

Specific Comments

Some illustrations that illustrate the general comments:

- Title. I suggest to remove the water resources from the title, since it is not part of the paper. The paper only focuses on runoff.

Water resources will be removed from the title.

- Abstract. The main conclusion of the abstract seems to be that there is more uncertainty in runoff response to elevated CO2 than previously considered. However from the conclusion part of the paper I understand that the main message is to include ecosystem models in water resource management and impact studies, because they give different results. This is somewhat unclear and should be consistent.

We propose to alter the abstract to focus more on hydrological vs. ecosystem models' runoff projections, rather than the response to elevated CO_2, to make it more consistent with the conclusions. More generally, we find the effect of CO_2 on the runoff response is more uncertain than in previous studies. This uncertainty does not mean that this process can be ignored altogether, as is currently the case in most hydrological models.

- Definition of ecosystem models is that they inlcude vegetation dynamics (line 26, p 282). In table 1 it is shown that VISIT model does not include vegetation dynamics...

We propose to include a better classification of the impact models in each category: "Ecosystem models have been taken to be those which contribute to the ecosystem sector of ISI-MIP and the hydrological models to the hydrological sector only. As can be seen in Table 1, hydrological models in this study have been classified as those which do not include carbon dioxide impacts or vegetation dynamics, whereas ecosystems models include carbon dioxide impacts and three of the four also include vegetation dynamics. However, the distinction between hydrological and ecosystem models is unclear, with two of the impacts models (JULES and LPJmL) considered to be ecosystem models in this study, considered as hydrological models elsewhere."

- It would be very good to have a short description of the used models in the text. Now the models are presented without any description on main differences

Due to the number of impact models used, it had been decided that the use of a table giving more details of the different models would be most appropriate. If the referees feel that it is necessary to include more detailed description of model

differences, this can be added as an appendix. Where appropriate, we propose to discuss model differences which are relevant to the message within the text.
- Line 21, p283. Whether the model shows sensitivity for CO2 depends very much on the model. Before it was stated that most hydrological models do not consider CO2 concentration at all, therefore they are not sensitive to CO2. I agree that the difference between ecosystem models and hydrological models is not always very clear (in other studies LPJmL and JULES are regarded as hydrological models), but that should be explained. Now it is rather confusing what are the differerences between the two.
The distinction between ecosystem and hydrological models will be made clearer, with models classified accordingly – see response for comment which is 2 comments above.
- Line 23, p 284 introduces JULES without any references or description
References will be added here for the JULES model.
- Line 21, p 284. It would be good to know what the 'minimal settings' are.
The ISI-MIP description paper is referenced, which explains what the minimal settings are. A brief description of the runs included in the minimal setting is proposed to be added to the text – see response for comment which is 4 comments above.
- Line 2, p 286. Why not use basins?
It is proposed to explain our reasoning in the Methodology. There is a detailed response to this point in the response to Referee #2 comments below.
-Line 4, p 286. Which regional differences? Unclear sentence.
This is referring to Giorgi region averages, so will be worded differently and be included within the updated Giorgi region paragraph in the Methodology.
- p 287 first paragraph is already described in the intro.
This paragraph will be adjusted so that information is not duplicated from the introduction.
- Line 10-19 p 287 start with a discussion of this study's results, before the results are even presented. I think it should start with fig 1 and after that the discussion on what is presented in fig 1 etc. This would make the
The ordering of this section will be adjusted so that the results are presented before they are discussed.
-Line 21 p 287. What is the definition of 'consensus' here? Does it mean that 50% of the models show the same signal? OR also in the same magnitude class of change? This needs some explanation.
The method of calculating consensus for Figure 1 is proposed to be explained in the text, and follows Kaye et al. (2012). Consensus was calculated as the proportion of models agreeing on the same magnitude class of change, with different hues representing different classes of change, and the saturation of the hue representing model agreement on that class of change. The proposed text in the results referring to the consensus plots will read: "Across all models (Fig. 1), there was consensus for annual runoff increases over high latitudes and India, parts of China, the Great Lakes and Eastern Canada, with more than 70% of models agreeing on increases in runoff greater than 10% over large parts of these regions, and for decreases in runoff over Southern and Central Europe, the Amazon and Indonesia, with many places having more than 70% model agreement on a runoff reduction of 50% or more. The ecosystems models showed less agreement in runoff changes over Europe, Central Africa and the Amazon, compared to the hydrological models, which showed that more than

70% of hydrological models agreed that runoff would increase by more than 10%. Supplementary Figure 2 shows the runoff changes for individual models."

- Line 25 and further p 287. It is not very strange that other patterns of change are found than in other studies, because they might have forced with different GCMS which show very different patterns of precipitation change regionally. Also Hagemann et al (2012) used three different GCMs, therefore, runoff changes between this study and Hagemann et al. cannot be directly compared. Tang and Lettenmaier foudn that spatial patterns are stable across emission scenarios, but NOT across GCMs!!

The reasons for the differences in patterns of change between this and other studies will be explained, stating that the findings cannot be directly compared due to different GCMs and scenarios, with the present study considering only HadGEM2-ES RCP8.5. For example, Portmann et al. (Accepted) found differences in precipitation pattern of the projections of the 5 GCMs used within ISI-MIP and Nohara et al. (2006) used an ensemble of 19 AOGCM simulations. The sentence about Tang and Lettenmaier was not intended to imply that spatial patterns of runoff sensitivity should be stable across GCMs, but to suggest that we might expect similar spatial patterns under different future scenarios than RCP8.5. This will be made clearer in the text.

- Line 8, p 288. What is actually said here is that the ecosystem system models are generally the upper points in the graph. I find the figure too messy to draw this conclusion, because all the experiments (including sensitivity) are included in the graph. I think the figure would be stronger when just the 'standard' runs are presented, with changing CO2 and dynamic vegetation. It would become much clearer and convincing in this way.

We propose to replace the graph, showing runoff only from the standard ISI-MIP runs.

- Line 18 p 288. Why does the fact that the points are not on a 1:1 line indicate that there are regional variations? I don't understand this conclusion. I would say that it means that not all additional precipitation is added to runoff, but also partly enhancing evapotranspiration.

This should have read: "The differences in distance from the 1:1 line for each region (regionally-averaged precipitation change) differ, suggesting regional variation of how the changes in input moisture will be affected by evaporation."

- Line 21 p 288. For readibility, I suggest to add the abbreviations of the regions as used in the figure here.

These are to be added in brackets when region names are used within the text to link better with the labels in the figures.

- Figures (3 and 5) have too many lines to be readable. Since the individual models can not be distinguished, it is also possible to show the ranges for each type of model.

It is agreed that these plots are difficult to interpret as there are a lot of lines. Although plotting only the ranges of the different types of models would make it simpler to draw conclusions from the graphs, these plots would not give a true representation of the annual cycle for any particular model. We will therefore use shading to show the ranges of the two model types with lines for the individual models over-plotted using thinner lines to make them clearer. This will hopefully result in no loss of detail, while gaining clarity in what is being shown by the graphs.

- Line 23, line 290 Over Wester Africa.... hydrological models. This conclusion cannot be drawn based on fig 5, because hydrological models are not shown in there.

This sentence should have referred to Figure 3, as well as Figure 5, which includes hydrological models. This will now be amended in the text: "Over Western Africa (WAF), the impact of CO_2 was largest in summer (Fig. 5), and most ecosystems models also projected larger increases in runoff than hydrological models at this time of year (Fig. 3)."

- Line 16-20 p 291. Isn't this very logical? In some locations there won't be a vegetation shift, in some locations from high to low vegetation and in other lcoations from low to high vegetation.

Yes, it is logical. It will be mentioned that this is what would be expected.

- Line 22, p292. Changes in vegetation had more effect on what?

The changes in vegetation had more effect on runoff change projections. This has been added to the text.

- Line 11, p 295. What is the definition of runoff deficit?

This was referring to Prudhomme et al. in which runoff deficit is defined as when runoff falls below a defined threshold of Q90 (daily) for the reference period. We propose to add a brief description to the text, however the calculations involved will not be explained. Small runoff deficits imply that typically there is more runoff than large deficits, which is how the findings were being compared with the current study.

- Line 15, p 296. It would be helpfull to start with a summary of the main differences.

We propose to adjust the conclusion so that it begins by outlining the main differences between the ecosystem and hydrological models' runoff projections.

Technical Comments

- Reference Falloon et al, 2012a and 2012b are the same.

Falloon et al., 2012b is the discussion paper version, and there were changes between this and Falloon et al., 2012a.

- Pag 285, line 14, text between brackets should be removed.

This text between brackets will be removed.

- Page 286, line 10, include a reference here

A reference to the relevant paper will be included.

- Fig 3 caption. Add reference here.

A reference to the ISI-MIP description paper will be included.

- P 295, line 11. Insert constant before CO2.

The word constant will be inserted before CO_2 in the relevant place.

Anonymous Referee #2

This article deals with a very interesting topic on comparing hydrological and ecosystem models for projections of future changes in runoff. The authors revealed that ecosystem models tended to project larger increases and smaller decreases in runoff than the hydrological models, likely due to the hydrological models ignore the effects of CO2 and vegetation dynamics. The results are important to global hydrological study, and inspiring for selecting useful models for simulations. The paper should be published after major revision considering the following comments:

General Comments

(1)In the introduction, put the description of WaterMIP (Page 282, line 3-27) to the end of Introduction. It is not suggested to put this part at the beginning.

The description of WaterMIP is background information rather than what is included in this study, so it makes sense to leave it where it is to show what has been found already related to what is being investigated with this study, which is

described later in the introduction. We propose to say, "for example, the WaterMIP intercomparison…" to make it clearer that it is being used to put the current study in context of other related studies.

(2)There is a need to present calibration and validation of all models for runoff simulations.

We think that a detailed section on these issues is not necessary within the present study. Most of the impact models were not calibrated specifically to the ISI-MIP forcing data which was applied. Some of the models were calibrated independently by each model author(s) against observed global datasets, and for others, specific processes parameterisations were tuned on local or regional scales for specific regions, which are then applied globally. Therefore, impact model calibration issues are not really relevant in the present study. More information on the individual models can be found in the relevant model description papers, given as references in Table 1. We propose to address validation by comparing the Fekete dataset with the different models by comparing globally and regionally averaged runoff values, and by plotting annual cycles. We would like to emphasise that the aim of this paper is to look at differences in projections into the future. Most of the models which participated in ISI-MIP are models that have been validated and used previously at the global scale, for example in WaterMIP.

(3)River basins are recommended for runoff study. The authors need to give reasons why they did not study at a river basin level.

Although we are aware that there are deficiencies in using Giorgi region averages for analysis, as on a sub-region scale there may be robust and plausible hydrological responses which would not be captured from spatial averaging, we have decided that the Giorgi regions are best suited for our analysis, for the following reasons, which we propose to explain in the Methodology.

- **The Giorgi regions tend to cover larger areas than river basins, meaning that the entire land area of the world could be covered by a manageable number of similarly-sized boxes to give a broad global picture.**
- **The regions are of a large enough horizontal scale for GCMs to be considered useful for climate simulations.**
- **Particular to our method of analysis (of using scatter plots), the use of similar-sized regions is preferable as different sized river basin regions would give a visual bias as small and large regions are given the same weight.**
- **Other papers have also used relatively large regions rather than river basins. For example, Betts et al. (2007) and Gedney et al. (2006) both consider runoff at the continental scale rather than at a river basin scale.**
- **Comparing results from other ISI-MIP papers with our general findings, we would expect that using smaller regions would be unlikely to change the overall conclusions and message of the paper. For example, Schewe et al. considered runoff at a country-scale (calculated using basins) and global scale, and found that JULES and LPJmL had a lower proportion of the global population under water stress than the other hydrological models in the future, which is in agreement with our findings. Prudhomme et al. considered runoff at a global scale and GEO sub-region scale, drawing similar conclusions when considering JULES in relation to the other hydrological models.**

(4)There is a need to present how the models consider vegetation change.

This will be addressed with more detailed description of model differences.
(5)Page 297, Line 15-19: How the decision makers plan water resources management with a large uncertainty from different models?
The paper does not attempt to address this question. It would depend on the range of possible changes in runoff from the different models, and would also need to be put in the context of climate variability. More generally, if the uncertainty is large, we feel that it is better to state this rather than ignoring it altogether.
(6)It is not suggested using many unpublished references.
These references are ISI-MIP papers so very relevant to be included. If the status of any of these are updated while this paper is in revision, the references will be updated accordingly. For example, Schewe et al. (2013) is now under review and the reference will be updated to reflect this.
(7)The English language and manuscript format should be further improved.
Improving the language and manuscript format will be kept in mind while revising the paper.

Specific comments
(1) Page 281, line 25: the first sentence should be improved.
This sentence will be improved.
(2) Page 283, Line 1-4: References should be added to the statement "Firstly, plant structural changes,... alter evapotranspiration rates and albedo". Possible references include:
- Liu J. and Yang H., 2010. Spatially explicit assessment of global consumptive water uses in cropland: green and blue water. Journal of Hydrology 384: 187-197. Doi: 10.1016/j.jhydrol.2009.11.024
- Liu, J., A. J. B. Zehnder, and H. Yang, 2009. Global consumptive water use for crop production: The importance of green water and virtual water, Water Resources Research 45, W05428, doi:10.1029/2007WR006051.
(3) Page 283, Line 7: the following references should be added:
Liu J., Folberth C., Yang H., Röckström J., Abbaspour K., et al. (2013) A Global and Spatially Explicit Assessment of Climate Change Impacts on Crop Production and Consumptive Water Use. PLoS ONE 8(2): e57750. doi:10.1371/journal.pone.0057750
We thank the reviewer for these suggestions; however we feel that they may not be directly relevant and so do not propose to include them.
(4)Page 284: Line 9: correct the format of (Fallon et al., 2012b), and delete ";"
The format seems to be consistent with the rest of the paper.
(5)Page 286, Line 5: format Meehl et al. (2007) is not correct
This will be corrected to (Meehl et al., 2007).
(6)Page 286, Line 8: "information about which is lost ..." this sentence is not clear for me.
This is about the heterogeneity of the Giorgi regions. This point will be addressed in the updated text for the Methodology, explaining why we have used Giorgi regions rather than river basins.
(7) In Fig. 2, regional names are sometimes too close to each other. This also applies to Fig. 4, Fig. 6
Region names will be spread out more vertically to improve legibility in these 3 figures.
(8) C02 in figure legend should be capital (not co2). [this applies to many graphs]

This will be corrected in all graphs where capital letters have not been used for CO_2.

(9) Supplementary Fig. 2, the legend and title are put in a crowd way. Pls adjust the formats.

The formatting of the title and legend will be changed so that they are more spread out.

Additional Proposed Revisions

In addition to these changes suggested by the reviewers, we propose to include an additional impact model, PCR-GLOBWB, to the analysis, which we have been made aware was also available but was not originally included in the current study. This means that we would also have two more co-authors of the paper, Yoshihide Wada and Dominik Wisser. The hydrological model, PCR-GLOBWB would be added to Table 1 and Table 2 and runoff projections from the model included in the figures. The message of the paper remains the same with the inclusion of this additional model.

The figure captions will be clear to state that HadGEM2-ES RCP8.5 climate forcing data was used.

The reference Hagemann et al. (2012) in Earth Systems Dynamic Discussions has now been updated to Hagemann et al. (2013) in Earth System Dynamics. Additional references for PCR-GLOBWB would also be included.