Title: "Improving the quantification of climate change hazards by hydrological models: A simple approach for mimicking the impact of active vegetation on potential evapotranspiration" by Thedini Asali Peiris and Petra Döll, Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2022-230-RC1, 2022

Response to the peer reviewers' comments

Our response is written *in italics*. Changes in the revised text are shown **in bold**.

Comments from Reviewer 1 (Anonymous Referee #1)

In this study, the authors document an approach that aims to represent the influences of vegetation responses to rising CO2 and climate changes on potential evapotranspiration (PET) in hydrological models. The approach is a modified version of Priestley-Taylor (PT), which represents PET as a function of only the change in net energy input and temperature, but removing the long-term temperature trend (PT-MA). The approach is implemented in the WaterGAP model, which, when driven by output from historical and future (e.g., RCP8.5) global climate model simulations, is shown to capture the PET in non-water-stressed regions well compared to PT. Overall, the PT-MA method leads to a smaller increase in PET than PT, so there is a relatively smaller decrease (or larger increase) in water resources in the future.

Overall, the paper is well written, and the method presented has utility. However, there are some flaws in the analysis/interpretation and the framing of the method as "representing the effects of vegetation" is misleading. It would be more accurate to say that PT-MA accounts for processes that might oppose the influences of long-term warming on increasing PET, since it does not actually "mimic" the direct effects of vegetation. I think the paper could be publishable after a significant rewriting and reframing of the results, as well as some additional analysis (e.g., statistical significance testing). Specific suggestions are outlined in the major and minor comments below.

We thank the reviewer for the overall positive evaluation of our manuscript. We address planned improvement regarding the framing and the analysis below.

It is not accurate to say that the approach "mimics the effect of active vegetation in PET estimation" simply because it "removes the long-term temperature trend". Rather than "mimicking the effect of vegetation" (or adding something associated with plants), it is just not including the long-term warming associated with radiative effects - i.e., rather than include temperature trends and stomatal closure, the method proposed here is to include neither. This approach works to some degree for the results from primarily CO2 driven emissions scenarios, where the temperature trends and stomatal closure are both driven primarily by CO2. The method, or at least the description of the method, assumes all processes that do not contribute to PET increasing with temperature are associated with vegetation, but no evidence is provided to justify this assumption. It is likely the method would not work as well for scenarios with larger non-CO2 drivers (e.g., aerosol, non-CO2 GHGs, land-use change). For example, the effects of aerosol and/or non-CO2 greenhouse gases could influence temperature trends without having an opposing influence associated with vegetation, but this method cannot distinguish this difference. To represent the effects of CO2 stomatal closure, it would be better to have the temperature trend adjustment depend on the concentration of CO2 directly. Though even that would only capture one aspect of vegetation effects - many GCMs simulate a large increase in leaf area that can lead to more canopy interception of rainwater and thus more canopy evaporation, which can somewhat offset the stomatal closure effects on actual ET. I suggest rewriting and reframing the paper, so that it does not claim to "mimic active vegetation".

We agree with the reviewer that our approach is certainly not able to take into account all the complex interactions between e.g. non-CO2 drivers. In addition, it is not able to into account any biome specific effects, or effects of nutrient availability, or a number of other factors. These multiple interactions and factors are simulated by complex DGVMs as part of complex GCMs, and each model (GCM or stand-alone land surface model or DGVM) computes very different vegetation responses and effects on future runoff. At the same time, hydrological models from drainage basin to global scales are being applied to estimate the impact of climate change on renewable (ground)water resources, streamflow dynamics including floods and droughts, with models that simply neglect that climate change (including CO2 increase) will have an impact on PET and runoff generation. With our manuscript, we want to address hydrological modelers and provide them with a way to take into account, at least very roughly, the effect of active vegetation in their models and thus avoid (at least in most regions) an overestimation of drying due to climate change.

The verb "to mimic" means "to have the same or similar effect as something else" (Cambridge Dictionary). Our proposed approach aims at leading to a similar effect on PET and runoff as the complex GCMs (with DGVMs) show (at least on average). This is why we used the word "mimicking". However, both reviewers think that it is not correct to use the term "mimic". Therefore, we plan to replace the word "mimic" by the word "consider" and add the term "approximate" in the title, and also adapt the main text. So the revised title reads:

Improving the quantification of climate change hazards by hydrological models: A simple approach for considering the approximate impact of active vegetation on potential evapotranspiration.

Where we need to name our approach in the abstract and main text, we use the term "emulating approach" instead of "mimicking approach", using the definition of "emulators" of GCMs as provided in Chen et al. (2021), p. 219.

Chen, D., M. Rojas, B.H. Samset, K. Cobb, A. Diongue Niang, P. Edwards, S. Emori, S.H. Faria, E. Hawkins, P. Hope, P. Huybrechts, M. Meinshausen, S.K. Mustafa, G.-K. Plattner, and A.-M. Tréguier, 2021: Framing, Context, and Methods. In Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 147–286, doi:10.1017/9781009157896.003

The Introduction is very well written but is somewhat misleading in its discussion of processes that are not represented in GCMs. For example, it discusses biome shifts (changes in plant type distributions), which is not represented interactively in current GCMs. Land use change is included in the RCP scenarios, but these are prescribed changes. Furthermore, it is not clear that this approach would be applicable to biome shifts - e.g., if vegetation died in a region, PET might actually depend more directly on the temperature trends. Also, different plant types will be influenced by water availability at different soil levels depending on their root depth, which is not accounted for here. To show that this approach would be applicable to biome shifts, it would need to include some analysis of models that include this interactively.

*We thank the reviewer for this comment, according to Randall et al., 2017, table 8.1. Most of the GCMs are coupled with a Land surface model, and some GCMs are coupled with DGVMs; for example: the IPSL-CM5A-LR model is coupled with the dynamic vegetation model ORCHIDEE (Sepulchre et al., 2020) . However, we agree with the reviewer that those coupled model GCMs do not fully consider the effect of vegetation dynamics, such as biome shifts. To address this, We added "***but still neglect other relevant vegetation responses***", which now reads as follow:*

DGVMs simulate physiological processes, such as photosynthesis and respiration, and biogeochemical cycles and include the effects of fire, atmospheric CO2, concentration and competition between plant life forms for light, water and nutrients on vegetation dynamics but still neglect other relevant vegetation responses

No statistical significance testing is provided to demonstrate that the differences are meaningful. In particular, the DC metric may appear to have very large differences if the climate change signal is very small. For example, if PT results in a change of 0.001 and PT-MA is -0.001, this will appear to be a much larger difference than changes of 1 and 2, respectively. It would be helpful to first determine if the climate change signal is statistically significant and then assess the impact of the two methods. Furthermore, the color bar choices (yellow vs. light green) make it impossible to determine if a change is very small - I suggest including a color (white) that is centered on zero to indicate very small or no change. Also, where the changes/differences are statistically significant should be indicated with stippling in the Figures.

Line 296: With the color bar used in Figure 5 and no statistical significance testing, it is not possible to determine if the changes in the western US are meaningful. I suggest adding a color (white) centered on zero to indicate regions with no change. And adding some statistical significance testing to these figures.

Line 303: As mentioned in the main comment above, the DC metric may appear to show large differences where the climate change signals are small. I suggest adding some indication in this figure for where the climate change signal is significant.

Figure 6: Again, with the color bar it is not possible to determine if the differences are meaningful for panels a and b. Very small values (e.g., -0.00001 would appear as yellow and 0.00001 would appear as light green) could be from rounding error. I suggest adding a color (white) that is centered on zero (e.g., -0.5 to 0.5), and adding significance testing.

Here we reply to the above four comments concerning Figs. 4, 5, and 6. Testing of the actual significance of the changes of e.g., renewable water resources change (Fig. 5), is conceptually difficult to do at the global scale. The same absolute value of change as in Fig. 5c-f has a quite different societal or ecological significance depending on the value during the reference period (e.g. comparing the humid Amazon basin to dry Southern Africa). This is one reason why we do not consider significance tests to be meaningful for these figures.

We agree with your suggestion to indicate where small changes in renewable water resources (Fig. 5) or PET occur in the DC figures (4 g-h, 5 g-h, 6c-d) by overlaying in grey areas below a certain change in renewable water resources or PET. We also changed the legends of Figs. 4a-f, 5a-f, 6a-b such that small positive and negative changes centered around zero are indicated in one color (light yellow). Thank you for helping us to improve the figures.

Line 10: It is not clear in the abstract how the approach attempts to capture the effects of "active vegetation". I suggest adding a sentence or two that explains/justifies this connection more directly.

*We revised the sentence (adding "***which simulates the active vegetation***"), which now reads as follows:*

Our approach is based on the work of Milly and Dunne (2016) (MD), which compared the change of non-water-stressed actual evapotranspiration (NWSAET) as computed by an ensemble of global climate models (GCM), which simulate the active vegetation, with various methods for computing PET change.

Line 110: Why is the approach only validated against 3 (or 4) GCMs? Why not use all 16 models that MD used? What criteria were used to choose the GCMs that were used in this study?

To conduct future runs, GHM needs the bias-corrected GCM-generated climate data (temperature, rainfall, radiation and etc.). Therefore, the analysis is limited to 4 GCMs where the bias-corrected climate data is produced and published by ISIMIP. The explanation is provided in section 2.3, lines 180 – 186. We refer to Frieler et al. (2017), where the explanation for the GCM selection is provided.

Line 124: This paragraph describes how the different components of evapotranspiration are calculated, as dependent on PET. Is the PT-MA method for calculating PET applied to all components or just some? Further down (line 175) it is stated that modified T is not used for open water, but what about regions where there is little vegetation (i.e., leaf area index is low), why use a version of PET adjusted for vegetated conditions in those regions? Also, vegetation effects would mostly have an influence during the growing season, so would it be more appropriate to use the original approach during winter?

The PT-MA method is applied to canopy evaporation, snow sublimation, and evapotranspiration because in Milly and Dunne (2016), total non-water-stressed evapotranspiration of GCMs was analyzed. We did not apply it to evaporation from surface bodies because WaterGAP takes into account other and mostly more surface water bodies than the GCMs. Regarding vegetation cover and seasonality, we believe that our global hydrological model cannot take this into account in a meaningful way.

Line 126: Since canopy evaporation is calculated as a function of PET and leaf area index, is leaf area index from the GCM simulations also used? Leaf area increases due to rising CO2 and will influence both canopy evaporation and transpiration.

No. In addition, leaf area does not change with climate change in WaterGAP (because the model does not simulate vegetation development).

Line 170: Given that vegetation effects will be most important during the growing season, why not remove the long-term monthly (or seasonal) trend? For example, if the summer is warming more quickly than the winter, would it be more appropriate to remove the summer trend rather than the annual trend?

The Milly and Dunne approach was developed using only a few grid cells and months with NWSAET. So they could not evaluate any seasonal differences; consequently, we do not have any basis to do so and would lack any validation data.

Line 235: How is it possible that actual ET is higher than net radiation (i.e., HadGEM2-ES)? What is the energy balance of net radiation, latent heating, sensible heating, and ground heat flux?

In Table 1, non-water-stressed actual ET is higher than net radiation, which can be explained: It is the net radiation of WGHM, which WGHM computes from bias-corrected short-wave down and long-wave down GCM radiation plus WGHM specific estimates of short-wave up and long-wave up radiation (lines 139-144). And it is the NWSAET directly from the GCM, which is based on net radiation that is different from WGHM net radiation. In the caption of Table 1, we indicate the source of the listed non-water-stressed actual ET and net radiation.

Line 244: A fixed value of 0.8 would likely not apply in all regions (as the authors discussion in section 4.2), so it makes sense that it would not match the PT-MA results. It could be that it is less than or more than 80% of net radiation in some regions. If you calculated the net radiation directly from the GCMs to compare to WGHM, you could determine if it is the radiation that is different or if it is the "scaling factor", which would help the discussion here.

This is beyond the scope of the work. In addition, GCMs do not provide net radiation as an output, but the components for the radiation include surface solar radiation downward, Surface upwelling short-wave radiation, Surface upwelling long-wave radiation, etc.

Line 263: Why is it intended that there is no difference before 2001? Was there no long-term temperature trend in those locations prior to 2001? Why do you not remove the temperature trend for the entire time series?

Figure 2/3: Why is there no difference between PT vs PT-MA (and T vs modified T) before 2000? With a reference period of 1981-2000, it would be assumed that actual temperature T would be lower than the modified temperature for the early 1900s, since a long-term warming trend would be present from 1900 to 2000.

We did not explain the objective of the approach well enough. We thank the reviewer for this constructive comment. We have now added to the second but last paragraph of the Introduction the following sentence:

The approach is applicable for estimating the change of hydrological variables between a reference period and a period in the future.

We selected as reference period the reference period used in Milly and Dunne (2016) (1981-2000) so that we could validate the approach. In line 417 in the Conclusion, we have written that "the reference period for another climate change study can be easily adjusted."

Line 260: Why not include all the models in Figure 3? Since there are only 4 GCMs, it seems somewhat random to choose 2 to put in the main manuscript and 2 others for the appendix.

Given that the figures include many panels, it is not possible to include all four GCMs in one figure. One alternative would be to move Figures B1, B2 and B3 to the main but comparison would not be made easier by this, and the flow of the text would be disturbed.

Line 272: Again, why are these 2 GCMs chosen over the other 2? Or why not show the multi-model mean?

The objective of Figure 3 (and Fig. B1) is to show how the variables T, PET and PET-to-Rn ratios relate to each other. This is only possible when showing individual GCMs, not the ensemble mean.

Line 291: I suggest removing "and open water bodies", since there wouldn't be any limit on water availability in open water bodies.

The words should not be removed as the relative increase of renewable water resources due to the proposed approach depends on the fraction of the grid cell that is covered by open water bodies, as there is no change of PET for the area covered by open water bodies. To clarify this, we added the following:

If a large part of a grid cell is covered by open water bodies, the RWR increase will be small as the PET of open water is not affected by the PT-MA method.

Comments from Reviewer 2 (Anonymous Referee #2)

The authors addressed the lack of adaption of vegetation to increased atmospheric CO2 in global hydrological models. Based on ideas from Milly & Dunne, 2016 they introduced a method to modify the potential evaporation method Priest-Taylor to address the adaption of vegetation to CO2 change by removing the long-term temperature trend. The used their new method for the WaterGap model with different climate forcing from different RCPs and GCMs. As a results there will be less transpiration from vegetation and therefore more water resources in future compared to the PT approach.

The paper is well written and the PT-MA method is presented in a good way. But some details of the model have to be revamped and the idea of reproducing result of GCM dynamic vegetation maybe has to be change into an idea of improving vegetation process in GHM.

We thank the reviewer for the overall positive evaluation of our manuscript. We explain below how we plan to improve the manuscript following the suggestions of the reviewer.

The starting point is fine,global hydrological models (GHM) do not address adaption to increasing CO2 and there is a need to include this adaptation. The proposed PT-MA approach is described as mimic approach to the way GCM model compute AEP. It might be that the GCM have a more comprehensive way to include the CO2 process but they are lacking the water availability part. The PET-AET approach in GHMs is still useful, because it can address water stress, water shortage, irrigation need water demand, water withdrawal from different sources etc. While the PT approach does not have any CO2 adaptation included, the PT-MA using a trend removal seems to me as the maximum possible adaptation vegetation can have. According to fig A1 the temperature reduction factor can go up to 10°C. I am not a vegetation expert, but I doubt that boreal forest can adapt to climate change this fast (see also Kropp et al. 2017). I think the PT-MA is useful to show the uncertainty of vegetation adaption between no adaptation and max. adaption, but it should not be described as mimicking GCM of dynamic vegetation.

We agree with the reviewer that our approach is certainly not able to take into account all the complex interactions between e.g. non-CO2 drivers. In addition, it is not able to take into account any biome-specific effects, effects of nutrient availability, or a number of other factors. These multiple interactions and factors are simulated by complex DGVMs as part of complex GCMs, and each model (GCM or stand-alone land surface model or DGVM) computes very different vegetation responses and effects on future runoff. At the same time, hydrological models from drainage basin to global scales are being applied to estimate the impact of climate change on renewable (ground)water resources, streamflow dynamics, including floods and droughts, with models that simply neglect that climate change (including CO2 increase) will have an impact on PET and runoff generation. With our manuscript, we want to address hydrological modelers and provide them with a way to take into account, at least very roughly, the effect of active vegetation in their models, and thus avoid (at least in most regions) an overestimation of drying due to climate change.

The verb "to mimic" means "to have the same or similar effect as something else" (Cambridge Dictionary). Our proposed approach aims at leading to a similar effect on PET and runoff as the complex GCMs (with DGVMs) show (at least on average). This is why we used the word "mimicking". However, both reviewers think that it is not correct to use the term "mimic". Therefore, we plan to replace the word "mimic" by the word "consider" and add the term "approximate" in the title, and also adapt the main text. So the revised title reads:

Improving the quantification of climate change hazards by hydrological models: A simple approach for considering the approximate impact of active vegetation on potential evapotranspiration.

Where we need to name our approach in the abstract and main text, we use the term "emulating approach" instead of "mimicking approach", using the definition of "emulators" of GCMs as provided in Chen et al. (2021), p. 219.

Chen, D., M. Rojas, B.H. Samset, K. Cobb, A. Diongue Niang, P. Edwards, S. Emori, S.H. Faria, E. Hawkins, P. Hope, P. Huybrechts, M. Meinshausen, S.K. Mustafa, G.-K. Plattner, and A.-M. Tréguier, 2021: Framing, Context, and Methods. In Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 147–286, doi:10.1017/9781009157896.003

PT-MA is using a reference period 1981-2000 and a trend removal from 2001 on. It is not clear why from 2001 onwards plants adapting to CO2 increase and not before this date?

Why 1981-2000 as reference, maybe including another reference period for comparison because it will change the results

As mentioned before the reference date 1981-2000 seems to be random. You get different results if you change the reference period or if you calculate backward (i.e. plant adaptation to cooler climate)

Line 263: that is exactly the shortcoming of your method.

We answer the last four comments jointly. We did not explain the objective of the approach well enough. We have now added to the second but last paragraph of the Introduction the following sentence:

The approach is applicable for estimating the change of hydrological variables between a reference period and a period in the future.

We selected as reference period the reference period used in Milly and Dunne (2016) (1981-2000) so that we could validate the approach. In line 417 in the Conclusion, we have written that the reference period for another climate change study can be easily adjusted. Changing the reference period will affect the resulting changes from the reference period.

A description why PT-MA is a mimic approach, that includes a description how GCM e.g. Sepulchre 2020, Krinner 2005 , model dynamic vegetation. Or even better describe PT-MA as a maximum vegetation adaption approach.

In the Introduction, we describe, in lines 55-58, the processes modeled by DGVMs are shortly described and refer to two papers that review a number of DGVMs, and in line 73 we explain that GCMs integrate DGVMs. To go into further in this text is out of the scope of this paper. PT-MA is not a maximum vegetation approach but represents what GCMs, on average, computed as the best estimate of vegetation adaptation. Results derived with PT-MA can certainly be used as one ensemble member of a multi-model study or in a sensitivity analysis.

A description why the PT-MA can be used in hydrological models with a PET-AET approach and has therefore an advantage over the PT-EO approach which is a substitute to the GCM – NWSAET approach.

The PT-EO approach can only be directly applied to estimating long-term changes in (annual) PET. However, a hydrological model must be able to compute daily values of PET with spatial variations between the many thousands of grid cells. This is achieved by the PT-MA approach, which is the core innovation of our study.

A revamp of fig 4-6 is necessary, to filter out areas with low changes Table 1 and figure 2 have to be inline. Maybe showing PET-EO in both.

We agree with your suggestion. Small changes in renewable water resources (Fig. 5) or PET occur in the DC figures (4 g-h, 5 g-h, 6c-d) by overlaying in grey areas below a certain change in renewables water resources or PET. We also changed the legends of Figs. 4a-f, 5a-f, 6a-b such that small positive and negative changes centered around zero are indicated in one color (light yellow). Thank you for helping us to improve the figures.

The PET-EO value in Table 1 are mean values over the reference period and changes to mean of 2080-2099 as extracted from Milly and Dunne (2016). We cannot show PET-EO as annual time series in Fig. 2 as they have not been made available. Instead, we show the WGHM-derived net radiation times 0.8, which is similar to the change of PET-EO.

Line 55: Here a general description of dynamic vegetation models is given, but not how dynamic vegetation model calculate NWSAET. It does not need all details but more than one sentence.

The computations relevant for deriving the PT-MA approach are not computations by stand-alone DGVMs but those integrated in GCMs. Therefore, it is not meaningful to describe how DGVMs compute NWSAET within our context.

Line 110: For WaterGap you used the meteo forcing of ISIMIP2 which has only 3 GCM in common with Miley & Dunne . You have to explain it here.

To explain it already in the Introduction, we extended the last sentence of the second but last paragraph of the Introduction as follows:

It is validated by implementing PT-MA in the global hydrological model WaterGAP 2.2d using the bias-adjusted output of four GCM available on the ISIMIP data portal (Frieler et al. 2017), and comparing PET changes simulated by WaterGAP to NWSAET changes of three GCMs included in MD.

Line 126: evapotranspiration from the soil. Here you mean evaporation from soil and transpiration from vegetation and exactly this transpiration you are interested in.

Line 128: AET from soil is a function of soil PET (calculated as the difference between total PET, snow sublimation and canopy evaporation) and soil water saturation. This correct but in detail you are interested in the effect on transpiration of plants.

WGHM does not distinguish evaporation from soil and transpiration from vegetation.

Line 127: AET from the snow (i.e., sublimation) is determined as the fraction of PET that remains after canopy evaporation – sounds a bit strange as description (but you do not have to describe all the processes you are not interested in, anyway)

This is further described in Müller Schmied et al. (2021) as referenced in the paragraph. In WGHM, snow on the canopy is not distinguished from snow on the ground, and canopy evaporation is assumed to occur first.

Table 1: HadGem has higher PET than Rn why?

In Table 1, non-water-stressed actual ET is higher than net radiation, which can be explained: It is the net radiation of WGHM, which WGHM computes from bias-corrected short-wave down and long-wave down GCM radiation plus WGHM specific estimates of short-wave up and long-wave up radiation (lines 139-144). And it is the NWSAET directly from the GCM, which is based on net radiation that is different from WGHM net radiation. In the caption of Table 1, we indicate the source of the listed non-water-stressed actual ET and net radiation.

In Table 1 you show PET-EO but in figure 2 you show 0.8*Rn

The PET-EO value in Table 1 are mean values over the reference period and changes to mean of 2080-2099 as extracted from Milly and Dunne (2016). We cannot show PET-EO as annual time series in Fig. 2 as they have not been made available. Instead, we show the WGHM-derived net radiation times 0.8, which is similar to the change of PET-EO.

PET-EO is 0.8*(Rn-G) but it seems very different to 0.8*Rn.

We neglected G because 1) WGHM and other standard hydrological models cannot consider G and 2) at the daily time steps of hydrological models G can be more easily neglected than at the the sub-daily time steps of GCMs.

Part 3.2 Temporal development of PET at two locations does not really bring a different view than part 3.1.

The objective of section 3.2 is to show the relations between the temporal developments of T, PET and the PET-to-Rn ratio (Fig. 3), which is not shown in section 3.1 (Fig. 2). In addition, Fig. 2 shows the slope of the trend line between PET and Rn and allows a comparison to the ideal value of 0.8. Therefore, the information content is different and important for the understanding of the PT-MA method.

Figure 3: Instead c and d and g and H you can put the temperature change in a and b and show all 4 climate models (or explain why you choose these two climate models)

Figs. 3c and 3d show both temperature and PET-to-Rn ratio, already requiring the two available y-axes. Figs. 3a and 3 b require two different y-axes, PET and PET difference. Therefore they cannot be combined from four to two figures.

Figure 4-6: For all world maps I suggest putting in a neutral color (e.g. gray) to show values around no change i.e. from -0.5 to + 0.5.

Thank you for the suggestion. We will do so

Also for the DC value, please check if the absolute change is very small and therefore a small difference in the PT approaches lead to a high DC. For example fig 4 the colors blue and light red in fig4 c and e might lead to high % in Fig4 g even if the values in c and e are close to 0.

We agree with your suggestion to indicate where small changes in renewable water resources (Fig. 5) or PET occur, in the DC figures (4 g-h, 5 g-h, 6c-d) by overlaying in grey areas below a certain change in renewables water resources or PET. Thank you for helping us to improve the figures.

Instead of Figure 6 it would be niche to have a map of model agreement (some sort of the figure of Schewe 2014: Multimodel assessment of water scarcity under climate change).

This paper describes a new method and not aims to provide best estimates and uncertainties of future climate change impacts. A distinction of the individual GCMs helps to understand the impact of the proposed PT-MA approach much better than the representation of the ensemble mean (with model agreement). Therefore, we want to keep Figure 6.