Comment on hess-2022-230
Anonymous Referee #1

Referee comment on "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

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.

Major Comments:
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".

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.

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.

Minor Comments:

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.
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?

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?

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.

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?

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?

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.

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.

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.
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?

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

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

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.