Review of the manuscript "Unveiling the impact of potential evapotranspiration method selection on trends in hydrological cycle components across Europe", submitted to HESS by Vishal Thakur et al.

General comments

Thank you for giving me the opportunity to review the manuscript "*Unveiling the impact of potential evapotranspiration method selection on trends in hydrological cycle components across Europe*" by Vishal Thakur and co-authors. I enjoyed reading this manuscript and consider the findings to be valuable and highly relevant: Even though potential evapotranspiration is a crucial component of the hydrological cycle, factors like data availability or convenience often play a role in the selection of a potential evapotranspiration formulation. There is no uniform way of dealing with potential evapotranspiration in hydrology, and often, different concepts (such as reference evapotranspiration and potential evapotranspiration) are used interchangeably. Therefore, studies like the one by Vishal Thakur et al., analysing the effects of different formulations of potential evapotranspiration are of great value to shed light on this often-neglected topic. The manuscript presents an analysis of the influence of different potential evapotranspiration formulations on different simulated components of the hydrological cycle. To do so, the authors make use of a large-scale modelling approach and analyse the modelling results with effective methods. They come to the important conclusion that the choice of a potential evapotranspiration method has an influence on the results when studying the hydrological cycle and its components.

My main concern is that the majority of the chosen formulations (or methods) are temperature-based. Since temperatures were rising in the studied period between 1980 and 2019, the potential evapotranspiration methods based on temperature do show a positive trend. So far, it is not clear if temperature-based methods are still reliable under the conditions of a warming climate (see for example studies on the so-called "pan evaporation paradox": Li et al. (2013, 10.1002/wrcr.20202); Wang et al. (2017, 10.1002/wat2.1207)). With eight out of twelve methods being temperature-based, a possible overestimation may strongly influence the results of this study.

Specific comments and suggestions that will hopefully help to improve the different parts of the manuscript are listed below.

Specific comments

Introduction

- While Thornthwaite (1948) was the first to introduce the term "evapotranspiration", according to Miralles et al. (2020, 10.1029/2020WR028055), the concept itself was already used earlier. This should be specified. Furthermore, as the different concepts regarding evaporation (including "reference crop evapotranspiration") are often used interchangeably in hydrology, I would suggest to add a statement about this problem, e.g., referring to the Miralles et al. paper mentioned above as well as the already cited paper by Xiang et al. (2020). In addition, I would argue that there should be some indication if you consider evapotranspiration to include interception or not in your study. See also the commentary by Savenije (2004, 10.1002/hyp.5563).
- In the first sentence of the paragraph starting at line 41, PET is stated to directly influence AET. In the second sentence, an alternative to this direct influence is presented. Therefore, the first sentence should be adjusted so that it becomes clear that this is just one possibility. Furthermore, depending on how interception is dealt with, this may be a necessary component to add in the list of fluxes that make up AET (in the second sentence of this paragraph).

Methods and data

- In the beginning of section 2.2, you give the study by Hersbach et al. (2020) as a reference for ERA5-Land. I would argue that the suitable reference there is the paper by Muñoz-Sabater et al. (2021), that you cite later in the manuscript. Furthermore, in line 97, there is a reference

missing for EM-Earth and SC-Earth, or it is not clear to me if the Tang et al. (2022) reference belongs to this. If the EM-Earth data are based on ERA5 data (and not on ERA5-Land data), I would suggest to include the reference to the Hersbach et al. (2020) paper there.

- I assume the AET, Q, and TWS data in Figure 1c to be simulated values. In my opinion, this needs to be mentioned in the figure caption as it only becomes clear after reading further.
- In line 122, you state that some of the temperature-based methods also use some extraterrestrial radiation term calculated based on latitude. Potentially, this could already be mentioned earlier, when the definition of the temperature-based category is given (i.e., it could become clearer that also radiation terms can be required and a method is still considered to be temperature-based).
- When you list the terms that are required for the Penman-Monteith and the Priestley-Taylor method, I would suggest to list all the terms (as it's only two equations), instead of giving some examples and concluding with "etc.". Alternatively, you could give one or two examples and then place a reference to the table where you list all the input data required for each method.
- For me, the description of the modelling part (paragraph starting on line 139) is hard to follow. Therefore, I ask you to give more information about this part of your study: You do one model run per catchment and PET method (553 x 12 = 6636). Is there any model calibration? If yes, please elaborate on that: What objective function(s) did you use? What was the spatial resolution? If no, where do you take the model settings from? Are these the same for all catchments? How do you make sure that the parameterization that you use matches your catchments? Did you think about equifinality and how other possible parameterizations could influence your results? How well did the model perform for the different catchments? In the very end of the discussion, the reader learns that a default parameterization was used. Please add this to the methods part and make sure that the questions listed above become clear.
- Related to the comment above, where does this default parameterization come from? Was there a method for potential evapotranspiration involved when this default parameterization was obtained? If so, may this have varying effects on the results based on this (or similar) methods (that can potentially profit from compensating effects) and the results based on different methods (that cannot profit from any compensating effects)? Potentially and if possible (may be limited due to data availability constraints) it would be interesting to compare the current results to the results of a calibrated model (i.e., calibrated for each of the different methods) – you also include a similar remark in the discussion. With additional calibrations that make use of the different methods, it could become clearer if the (assumed) use of one method for the default parameterization affects the results. If additional calibrated model runs are not possible, I would suggest you to include some text on the potential effects of the default parameterization in the discussion.
- I suggest you to clarify in lines 158 and 159 that you compare winter data to winter data, spring data to spring data, etc. so that it is immediately clear that you do not compare all four seasons with each other.
- In section 2.3.4, you describe the modification of the DCI by including also non-significant trends to be able to include all trend estimates. Are you sure that you are not including too much noise by also counting very weak positive and negative trends? I suggest to at least show that your findings do not change if only reasonably significant trends are considered.

Results

- In the first paragraph of section 3.1, you describe the different trends in PET for the three categories of catchments. As PET describes a potential, i.e., the maximum ET that could be achieved if there was no water-limitation, the trend in PET should not depend on if a catchment is water-limited or energy-limited. Thus, I suggest to formulate this paragraph differently, such that it is clear that the PET trend is influenced by other factors, but not by the availability of water.

- In the last paragraph of section 3.2, you suddenly write about December and not about the winter season (line 234). If you did monthly analyses, this needs to be stated clearly in the methods section. Otherwise, December may be the wrong term here.
- As already mentioned in the general comments: With the rising temperatures, all the temperature-based methods tend to show positive trends. Since most PET methods included are temperature-based, they have the largest influence on the DCI. It would be good to test if this unequal distribution of the different types influences your results.
- To make the part about the combinations of hydrological cycle component changes (section 3.4) easier to understand also before studying Figure 6, I suggest you to add an example of a "possible combination of hydrological cycle component changes" after you introduce this term on lines 269 and 270. Similarly, it may be good to inform the reader what you mean with "the first five hydrological cycle combinations" (lines 270 and 271).
- In lines 272 and 273 you state that the temperature-based methods account for more catchments with positive trends across all hydrological cycle components than combinational methods (and you put this in comparison to the Blaney-Criddle method leading to positive trends in all components for most catchments). This statement is not clear: Do you want to say that the temperature-based methods lead to the "all positive" combination in more catchments than the combinational methods? If so, this is not true for the Baier-Robertson method, if I interpret Figure 6 correctly. Please reformulate this statement and reconsider if this should be formulated as a comparison to the first part of the sentence.
- The sentence regarding the combinations with AET+ and TWS- (lines 280 and 281) should come before the statement about the last five combinations to be consistent with the order in Figure 6.
- I think that Figure 6 should be improved so that it can be grasped more quickly and easily:
  - I suggest you to put the lower part (in the following called "combinations") to the top and the upper part (in the following called "catchment count") to the bottom. The reader first wants to know what combination we are looking at, and then wants to look at the results for this combination. I see that you used the "combinations" part as an "axis label", but this is not so clear when looking at the plot in the beginning as it looks more like two different parts.
  - For the "combinations", consider displaying them differently. For example, a simple table-like graphic in which you could even work with additional colours would in my opinion be easier to interpret than the current way to display the combinations, see below. If you decide to do so, do not forget to change the caption.

PRE	+	-	+	+	-	-	-	+	-	-
AET	+	+	+	+	-	-	+	+	+	-
Q	+	-	-	+	-	-	-	-	+	+
TWS	+	-	-	-	-	+	+	+	+	+

  - For the "catchment count", I think it would improve the readability of the figure if not all bars were the same height, i.e., if the height of the bar would decrease from left to right, proportionally to the higher number of catchments that are contained in the bars on the left than in the ones on the right. This way, it would become clearer which combinations occur how often. For that, you could consider using horizontal instead of vertical bars to gain more space (this would also allow you to use the "combinations" part as "axis labels", but then of the vertical axis).
  - Please double-check the catchment count for each method. To my understanding, the number of catchments per method contained in this plot should always sum up to 553. However, for example for the Thornthwaite method, the number of catchments only sums up to 240+156+59+70+24+2=551.
  - As the Thornthwaite method and the Blaney-Criddle method are displayed in neighbouring fields for the seventh combination, please consider not using red and green for

these two methods as they are hard to distinguish for colour-blind people. Alternatively, please use patterns in addition, to make it possible to distinguish the two colours.

Discussion

- In line 302, you compare your results to the ones in the study by Hanselmann et al. (2024). Please note that while the authors of this study are affiliated in Poland, the research has been conducted for Spitsbergen.
- In the paragraph starting on line 336, you discuss your results as well as the results of the study by Teuling et al. (2019). Based on the Teuling et al. paper as well as based on the PET methods that you considered in your study, I assume that you mean the Penman-Monteith method in this paragraph when you write about the Penman method. Please double-check and correct. The same issue occurs again in line 384.
- Later in the same paragraph, where you discuss a possible drying hydrological cycle, I think that a discussion of the study by Milly & Dunne (2016) that you refer to elsewhere in your manuscript is lacking as they studied this topic.
- In line 368, you write about methods that consistently overestimate PET. It is unclear to me how you define overestimation here: Do you just consider the methods leading to the highest estimates to be overestimating PET? If so, I would not consider it to be surprising that the catchments shift to the energy-limited category (as all the higher estimates are excluded). If all the low estimates would be excluded, the catchments would probably shift to the water-limited category. Please elaborate more on what you mean there.
- When you discuss the limitations of your study, you state that on the one hand, the PET methods were found to be sensitive to the input data, and on the other hand, that hydrological models are sensitive to the precipitation input. Based on this, I think it would be good to write something about the data quality of the input data that you used (as you basically state that your results are highly sensitive to these data).

Summary and conclusions

- In the very last part of the summary, you recommend an ensemble of PET formulations instead of one single formulation. While agreeing with this statement, I think that it should be supported by the study and not only occur in the end without being mentioned before. Thus, I suggest you to use the ensemble of the different methods as a thirteenth option of the PET calculation in your manuscript.

Appendix

- Please add the references from where you obtained the formulations for the different methods to Table A1.

Technical corrections

- The words of the title should not be capitalized.
- In the abstract, the third category is called “combination type”, later it is called “combinational type”. Please improve for consistency.
- Please note that “ERA5-Land” is written with a capital L (not “ERA5-land”). This is not consistently correct throughout the manuscript.
- In line 75, you specify two of the components again after the abbreviations have often been used in the preceding introduction. Either, leave this away, i.e., just give the abbreviations, or give the full names of the components plus the abbreviations for all three components consistently. Furthermore, I think the sentence starting at the end of this line is redundant (stating the same as the preceding sentence in other words), please double-check and correct.
- On line 117 there is a full stop after “estimation”, but then the sentence seems to continue. Please double-check and correct.

- For the McGuiness-Bordne method, first occurring on line 119, you use different ways of spelling. To my knowledge, "McGuiness-Bordne" is the correct spelling, please double-check and adjust accordingly (do not forget figures and captions). Similarly, you don't spell "Baier-Robertson" and "Milly-Dunne" consistently (you sometimes write "Bair-Robertson", and in Table A1 "Milley-Dunne"). Please make sure that you use the correct spelling throughout your manuscript, including also the supplementary material (occurrence of "Milley and Dunne" in text S2). For all methods consisting of two last names, decide if you want to use a hyphen, a space, or an "and" to connect them.
- Double-check the references given in the tables: For example, in Table 2, the reference for the Hargreaves-Samani method is in a different style than the other references.
- It is mathematically problematic to use several letters for the same variable in an equation, e.g., "ND" could be interpreted as "N times D". Please consider reformulating.
- The first sentence of the results (line 175) is a repetition of the methods, please consider deleting.
- In the paragraph starting on line 184, it is not fully clear that the trends that are described are AET trends. I suggest you to formulate this clearer, for example by rewriting the second sentence to "…all PET methods lead to a positive AET trend in terms of median values." Similarly, in the following paragraphs describing the trends in Q and TWS (and in the results section in general), make sure that it is always clear that it isn't the PET methods that have trends but the PET methods that induce trends on the different components (if I understand your statements correctly).
- In line 206, you state: "…but the overall pattern of PET methods matches well with PET in all categories." This statement is unclear, did you mean a good match of PET with AET?
- In the caption of figure 3, you use mm $seas^{-1}$ $year^{-1}$ as a trend unit. I assume that 1 mm $seas^{-1}$ $year^{-1}$ in AET means that each year, 1 mm more goes to AET during the season of interest. As this seems to be an unusual unit (at least for me), I suggest you to explain this in the methods section of your manuscript already.
- In line 2019, the sentence starting with "AET in the summer season" is unclearly formulated.
- The references in lines 309, 310, 313 are not formatted correctly.
- In the supplementary material (Text S1), make sure that each excerpt starts on a new line and that there is always a space between the colon after the study name and the excerpt itself. Furthermore, look for typos in the copied excerpts (e.g., in the second excerpt from the Anabalón & Sharma study, you are missing a lot of spaces). For the Shi et al. (2023) paper, it would be good to indicate which paper you mean (a or b).
- Please include the methods' abbreviations in the caption of Table S1.
- I assume that Figure S1 and Figure S4 show the winter season, please double-check and correct the captions.
- In Figure S4, some of the numbers in the plot are not readable. Please see suggestions to improve Figure 6 of the manuscript (that also apply for all the other figures of this type). This should solve the problem. However, make sure that the different numbers are not written on top of each other.
- For Figures S10 and S11, the correct axis labels would be "Trend in … […]" as the data are showing the trend and not the value of a certain component.

Individual typos

- Line 9: comma missing between "water-limited" and "and"
- Line 13: space between closing parenthesis and comma needs to be deleted
- Line 32: word missing before the comma, probably "components"
- Line 37: should probably be "in hydrological models", rather than "in the hydrological model"
- Line 42: comma missing between "runoff" and "and"
- Caption of Table 1: "Short wave" and "longwave" radiation should be spelled consistently
- Line 133: space missing between "size" and the opening parenthesis

- Table 2: hyphen between “Penman” and “Monteith” is missing for the last method listed
- Line 141: “the” in the beginning of the line needs to be deleted
- Line 124: comma missing after “canopy interception”
- Line 160: parentheses around reference missing
- Line 162: comma missing between “significance” and “i.e.”
- Line 179: full stop missing after “Penman-Monteith[CO_2]”
- Line 228: “and” instead of “are”
- Caption of Figure 2: should be “categories of catchments” (not “categories of catchment”); the same typo occurs in the captions of the other plots of this type
- Line 242: “a strong agreement” (not “an strong agreement”)
- Line 252: comma missing between “Q” and “and”
- Caption of Figure 6: comma missing between “runoff” and “and”
- Line 279: “components” missing after “cycle”
- Line 419: “The” should not be capitalized
- Line 431: “of” missing between “use” and “an”
- Caption of Figure S12: should be “PET, AET, Q, and TWS” instead of “PET, AET, Q and, TWS”. Unit should be mm $seas^{-1}$ $year^{-1}$ instead of mm/seas/year.