Reviewer 2

We would like to thank the referee for the detailed comments. We greatly appreciate the time and effort taken in thoroughly reviewing our manuscript and for providing valuable and insightful perspectives on our research. We will carefully consider these comments while revising the manuscript.

We have separated the different comments (shown in italic) and provide our replies below. Text in the original manuscript is shown in '*italic*' and revised text in '**bold**'. Line numbers mentioned in our reply refer to the original manuscript version.

Reviewer comment:

After a thorough read of the article "Catchments do not strictly follow Budyko curves over multiple decades but deviations are minor and predictable" by Ibrahim et al., I can see the amount of work and understand the main arguments of the authors. The goal is to assess the predictive power of the parametric Budyko curves, usually considered as not suitable for climate projections since they rely on a semi-empirical parameter, and the lack of physical explanation behind it questions whether fixing it to project future behaviours of catchments is pertinent. The authors show that over most of the catchments studied, from one 20-year period to the next, the distribution of deviations to the predictive curve is minimum and stable. This leads them to conclude that the Budyko framework can be used for projections under a changing climate, just considering a stable distribution of deviation around the curve as a shape of uncertainty.

Reply:

We would like to thank the reviewer for a thorough review and for highlighting the objective of our research. We appreciate the acknowledgement of our efforts to evaluate the predictive power of the parametric Budyko curves.

Reviewer comment:

The article is well written, well-illustrated and well-integrated into the current literature. However, I am not sure every steps of the method are pertinent and I am not fully convinced by the conclusions drawn and how new the results are.

Reply:

We highly appreciate the reviewer's encouraging feedback on the writing, illustration and integration of our manuscript into the current literature.

Reviewer comment:

The method compares successive periods of 20 years. The method stays pertinent when looking at a 20-year period and looking whether or not the median deviation from the curve can be considered different from zero or not (step 2). Therefore, the conclusions can only be applied to argue that the Budyko framework can be used for 20-years projections, which is rarely the temporality used for climate projections.

Reply:

We acknowledge the observation that the chosen approach is valid strictly for predictions over 20-year windows, while less so for longer-range predictions. In the early phase of the study design, we have alternatively also considered longer time windows, but eventually, deliberately decided to use windows of 20 successive years as an approach that balances the need for sufficiently long time periods to limit the effect of storage changes dS/dt (Han et al., 2020), while preserving the temporal sequence in the data that allowed us to place each catchment into a specific category (i.e. "Stable", "Variable", "Alternating" or "Shift"). This aspect is one of the major novelties of our analysis, as it has – to our knowledge – never been analysed on global scale before. The use of fewer but longer time windows, such as ~ 30 years, as previously done by others, e.g., Destouni et al. (2012), would have considerably limited a meaningful distinction of systematic shifts from more random fluctuations over time.

We would also like to emphasize that 20-year periods are a not uncommon time-horizon for many water resources management interventions and planning, where such shorter-term predictions are often more relevant for decision-making.

However, we completely acknowledge the limitations of our choice. We will therefore add our reasoning for the 20-years window and a detailed discussion of the implications of this choice in the revised manuscript.

Reviewer comment:

The method also compares successive deviation distribution, for instance to define "stable" catchments as catchments for which the deviation to the curve from one 20-year period to the next has no specific direction. However, if I understood correctly, each distribution of deviation to the curve for each 20-year period is calculated around a different curve (with the actualised parameter fitted over the previous 20-year period). Then, what if there is a trend in this parameter? I understand it is not possible to evaluate such a trend significantly due to the length of the data but it would invalidate the comparison of the successive distributions. Why not use the same curve for all periods and look if the distribution around the curve changes over time? Could the successive fit over the 20-year sliding time periods be used here to assess trends?

Reply:

We greatly appreciate the reviewer's sharp observation and agree that a fixed baseline can provide additional insights. To explore this, we conducted the analysis using the reviewer's proposed approach by fixing the first 20-year period as the baseline and calculated distributions of deviations for the subsequent 20-year periods accordingly.

This comparison is illustrated in Figures R1 and R2 below, in which we compare the original approach with a dynamic baseline to the use of a fixed baseline for two of the example catchments of our study. For the first example catchment (Chemung River, Fig. 6a-c in the manuscript), we observed that the results from both methods were almost identical (Fig. R1a-b & R2a-b). However, for the second example catchment (Lee River Fig. 6d-f in the manuscript), the median deviations were somewhat higher when using a fixed baseline (Fig. R1c-d & R2c-d).

We also extended this analysis to all other study catchments. However, please note that we could include the full 100-year period in only 159 out of 2387 catchments. For the other catchments, we used the oldest available 20-year period as the fixed baseline. We have found that the proportion of "Stable" catchments increased from 72% (dynamic baseline - original approach) to 84% (fixed baseline - proposed approach) (Fig. R3), suggesting that dynamic baselines are more sensitive for the detection of systematic changes, i.e. "shifting" (trends) or "alternating" behaviour. In contrast, while the number of stable catchments increased, we also observed that the median deviations of the aggregated marginal distributions of deviations were slightly higher when using a fixed baseline (Fig. R4).

Despite the fact that there are some interesting results, we still prefer to keep the temporally changing (dynamic) baseline in the main part of our analysis for the following reasons:

1. The use of the most recent period as the baseline for assessing temporal stability of catchments is particularly relevant for future predictions, as the most recent data are most likely to provide a meaningful representation of current conditions. Using the oldest period as the baseline may not reflect recent conditions and could result in misleading conclusions, in particular when the first and last periods are far apart.
2. As previously mentioned, only 159 out of 2,387 catchments include the full 100-year period for $\Delta_{1\text{-}2}$, while 889 catchments have data for $\Delta_{2\text{-}3}$, and the number increase to 1916 for $\Delta_{3\text{-}4}$, and 2269 for $\Delta_{4\text{-}5}$. This distribution shows that for most catchments, only two or three 20-year periods are available, making changes to the baseline period less impactful.
3. As the reviewer correctly pointed out, it is difficult to quantify trends in the omega parameter based on only 5 values due to the limited length of the available data records. In many cases, we are working with just 2 or 3 time periods, making it impossible to detect significant trends.

Overall, adjusting the baseline over time is more sensitive to capturing recent shifts and trends in hydrological behaviour, which helps to assign catchments to one of the four categories.

The reviewer's suggestion to use successive fits over sliding 20-year periods to assess trends is indeed interesting. However, this would introduce dependencies between the periods, as each successive period overlaps with the previous one. This, in turn, would compromise the independence of the data from the individual time periods and potentially bias the results.

Based on the above and to provide the reader with a more comprehensive view of our results, we will, as the reviewer has suggested, include the results of using a single fixed baseline (Fig. R3 and R4) in the Supplementary Material of the revised manuscript. In addition, we will include a discussion of both approaches in the revised manuscript, outlining their respective benefits and limitations.

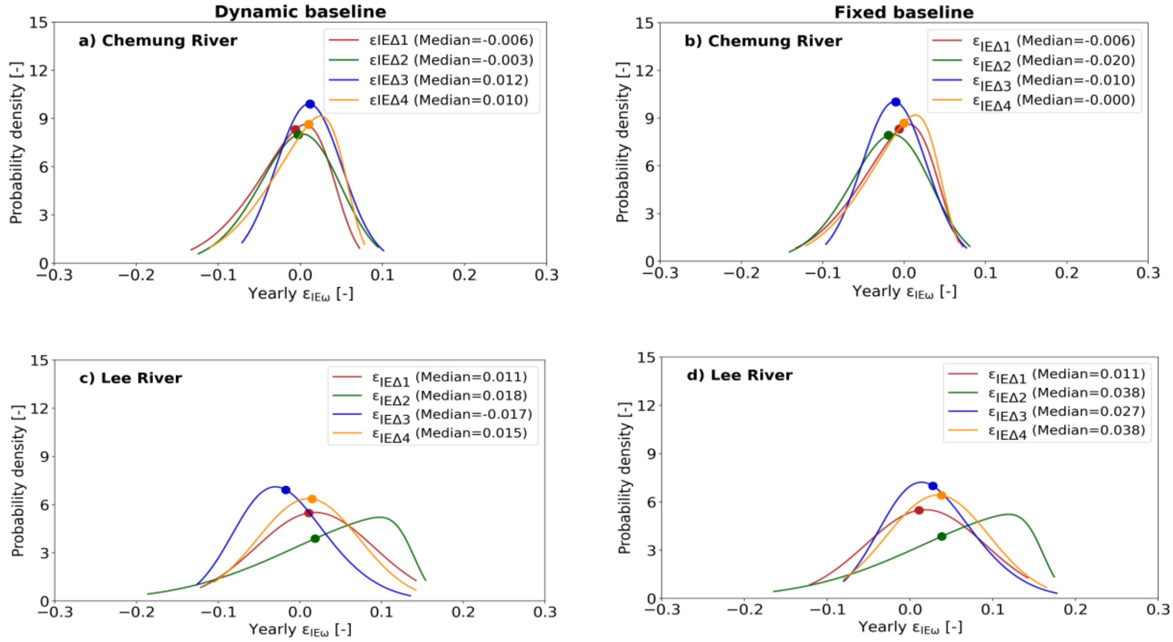

Figure R1: Comparison of individual distribution of deviations ($\varepsilon_{IE\Delta1}$, $\varepsilon_{IE\Delta2}$, $\varepsilon_{IE\Delta3}$ and $\varepsilon_{IE\Delta4}$) between the dynamic baseline (left) and fixed baseline (right) approaches for two example catchments (Chemung River and Lee River) in the Stable category

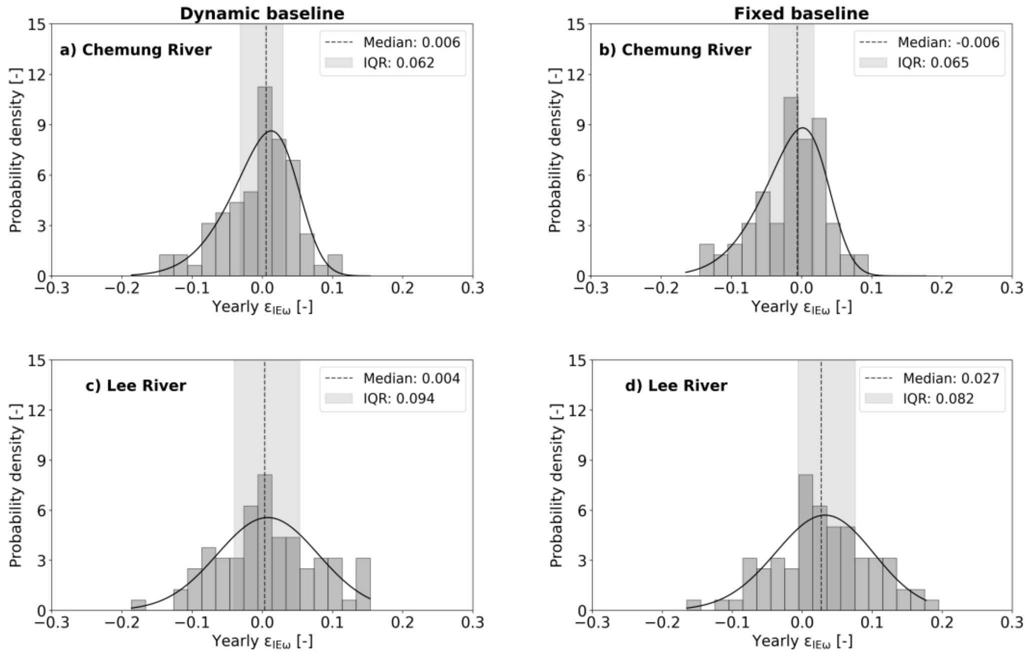

Figure R2: Comparison of long-term marginal distribution of annual deviations between the dynamic baseline (left) and fixed baseline (right) approaches for two example catchments (Chemung River and Lee River) in the Stable category

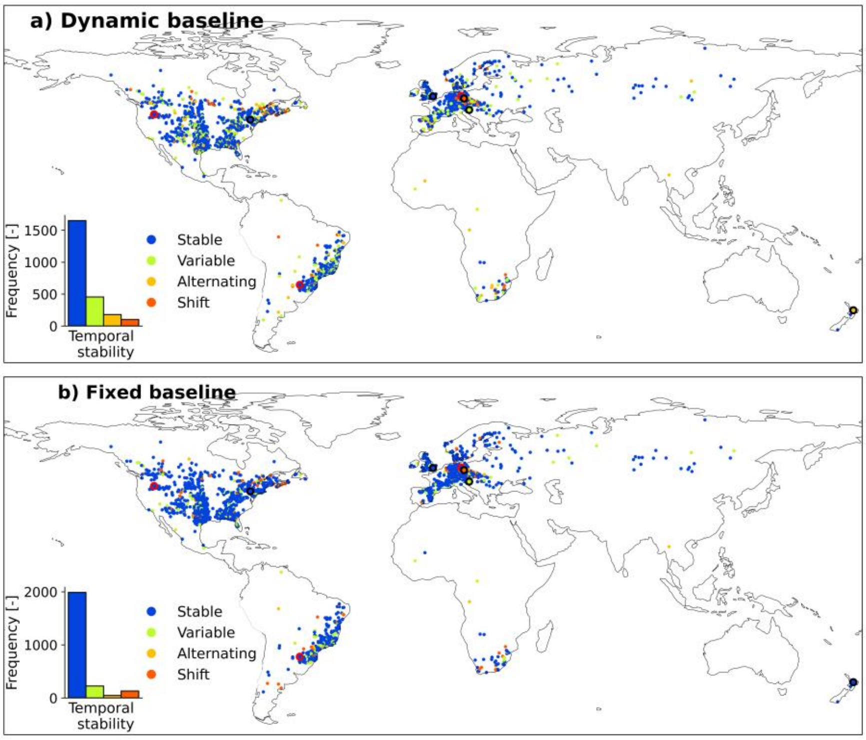

Figure R3: Comparison of temporal stability of the studied catchments using the dynamic baseline (top) and fixed baseline (bottom) approaches. Catchments highlighted with a black border represent the 5 selected examples from Fig. 6 (of the original manuscript), while those outlined in red denote three additional selected example catchments shown in the supplement (Fig. S4 of the original manuscript)

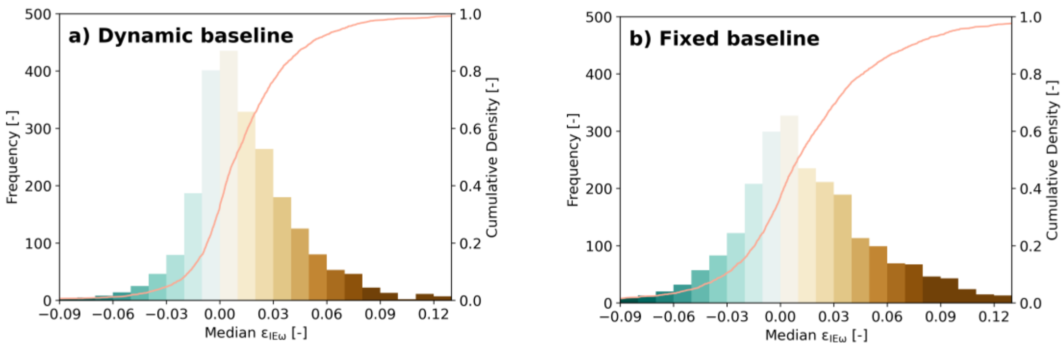

Figure R4: Comparison of long-term median $\epsilon_{IE\omega}$ values of aggregated marginal distribution of $\epsilon_{IE\omega}$ across all catchments (2387) comparing the dynamic baseline (left) and fixed baseline (right) approaches

Reviewer comment:

The authors argue that the distribution around the curve is just a natural variation around the curve ("stable" catchments) or due to regular climatic cycles ("variable" catchments). However, not all catchments fit in these categories, and since there seems to be no homogeneity in the spatial distribution or climatic characteristics of these catchments, it undermines the conclusion that the framework can be used for prediction in most catchments. It is not a generality, since such a study would need to be lead first in a catchment to check that it fits in a "stable" distribution, and whether or not it will seems arbitrary.

Reply:

While we agree with the reviewer's observation that not all catchments fit into the categories of "Stable" or "Variable", it is important to note that these catchments ("Alternating" or "Shift") constitute only a small minority, comprising 12% of the study catchments. The remaining 88% of catchments are either "Stable" or "Variable".

As pointed out by the reviewer, there is no clear spatial pattern. The regional distributions of $\epsilon_{IE\omega}$ remain, with medians of ~ 0 – 0.02 (Fig.9 in the original manuscript), broadly consistent with the global distribution (Fig.8) but also with each other across most spatial and climatic classes. This indeed suggests that the overall pattern is rather homogenous, and regional/local effects remain limited. The presented distributions (Figs. 8, 9) are in the absence of further information nevertheless useful to quantitatively estimate the uncertainty for any specific catchment based on past information in a probabilistic way, as clearly pointed out by Montanari and Koutsoyiannis (2014): "*[…] If a deterministic description of the process statistics along time, applicable to future times, is not available, which implies that non-stationarity is impossible to define, the only way for making predictions is through assumptions of stationarity*". Whereby "*if a deterministic description […] is not available*" in our cases corresponds to the impossibility to identify "shifting" and "alternating" out-of-sample catchments and "*assumption of stationarity*" corresponds in our case to the assumption that out-of-sample catchments are largely "stable" and "variable".

We will discuss these points in more detail and revise our conclusions to reflect that while the framework works well for the wide majority of catchments, it cannot take into account systematic shifts or alternating behaviour in out-of-sample catchments.

Reviewer comment:

I feel the results would benefit from a different presentation, to help show their impact. As briefly presented on the discussion, I feel it would be more pertinent to express the deviation to the curve by how much it shifts the predicted aridity or discharge (%), rather than present changes in an abstract parameter. The impact of the shift in the parameter is different depending on the aridity of the catchment, which could be interesting to analyse and could shed the results in a different light.

Reply:

We thank the reviewer for this excellent suggestion. We agree that expressing the deviation from the curve in terms of the percentage change in the evaporative index or discharge (or runoff coefficient) provides a better understanding of the impact. We have calculated the percentage change in discharge for each catchment and will include this analysis in the revised manuscript (Fig. R5). Additionally, we have analysed these changes across four different aridity classes to further highlight how the impact of the parameter shift varies with catchment aridity (inset in Fig. R5). This approach has allowed for a more meaningful presentation of the results, and it will be incorporated into the revised manuscript. This figure will be added as Figure 7c in the revised manuscript.

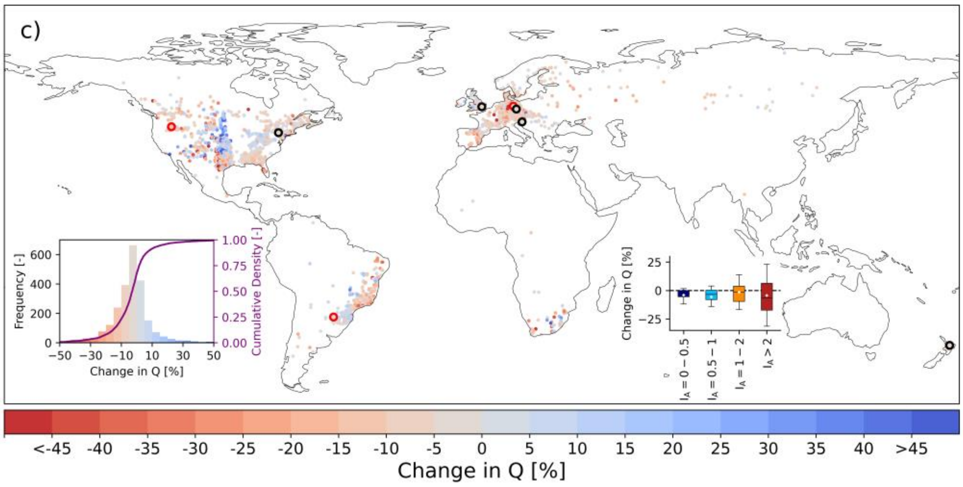

Figure R5: Spatial distribution map of changes in Q as a result of long-term median $\epsilon_{IE\omega}$ values. Histogram and cumulative density of changes in Q and change in Q across different I_A bins are presented as two insets. Catchments highlighted with a black border represent the 5 selected examples from Fig. 6 (of the original manuscript), while those outlined in red denote three additional selected example catchments shown in the supplement (Fig. S4 of the original manuscript). The boxes represent the 25th to 75th percentiles, while whiskers extend to the 10th and 90th percentiles. Diamonds denote the arithmetic mean, and outliers are not shown.

Reviewer comment:

Furthermore, with raw values of the shift in the parameter, it is difficult to understand whether it is a negligible change or not, as argued in the conclusion. Having more understandable orders of magnitude of this shift and the associated uncertainty would help argue that there is a potential in using a parametric equation for projection, with an inevitable associated uncertainty, which could be not be wider than the uncertainty associated to climate projection or to physical-based models. This is however still not a very new argument, and should be made with an understanding of the counter-arguments, that we are never sure that empirical models will respond reasonably when faced with unprecedented changes in climate. I believe this study would be interesting in that regard, as, if it doesn't introduce completely new concepts, it has a broader perspectives and a more targeted objective to quantify the uncertainty associated to the deviation from the parametric curve for a catchment in the Budyko framework. It would benefit from being formulated as such.

Reply:

We agree that it can be difficult to interpret the relevance of change based only on raw values of the shift in the parameter. The reviewer's suggestion to show the magnitude of the shift in forms of a magnitude change such as % change in discharge is indeed very helpful and we will include this in the revised manuscript. Furthermore, we also fully agree that it is an inherent weakness of empirical models that they typically cannot deal with previously unseen changes of the underlying distribution of a specific variable. Identifying the sensitivity of catchments to such changes in underlying distributions to address exactly this issue was our main intention with the classification of the study catchments into four categories. Following both, the shorter term climatic variability but also the long-term changing climate over the past century, catchments classified as "Alternating" or "Shift", are more likely to have experienced changes in the underlying distributions and are thus plausible to remain sensitive to change in the future. For such catchments, the empirical model will thus be less robust for predictions. However, for the other categories i.e., "Stable" or "Variable", the catchments exhibited much less sensitivity to past climatic variability. In the absence of statistical evidence for changing distributions it is thus not implausible to assume that they remain, at least for the near-future, rather insensitive to change and that the empirical models can thus provide plausible predictions. However, we explicitly acknowledge that issue remains a limiting characteristic of statistical models. We will include these points in the discussion section of the revised manuscript.

Reviewer comment:

Abstract, l11: I think "behaviour" is not the right term. You consider in your study parametric curves, where the parameter is generally considered to represent the specific behaviour of a catchment. A move along the curve is supposed to represent the changes in the catchment responses under a changing climate but with a fixed behaviour.

Reply:

Indeed. We will modify Line 11 as follows:

"***A movement along a specific Budyko curve with changes in the aridity index over time has been used as a predictor for catchment responses to changing climatic conditions***"

Reviewer comment:

L176-179: Here your two sentences are contradictory. If I understood correctly, for each 20-year sub-period, you fitted the curve to the set of n=20 values, not to the 20-year average directly. Therefore, you need to change the first sentence of that paragraph which says the exact opposite.

Reply:

We thank the reviewer for pointing this out. We will modify the lines 176-179 as follows:

"For each catchment and each individual 20-year time period T_i, the catchment-specific parametric Budyko curve $I_{E,i}$ defined by parameter ω_i is obtained by fitting Eq.(2) to the set of 20 annual values of each catchment in the Budyko space, as computed from the observed water balance data".

Reviewer comment:

I really like Figure 3, it helps understanding the steps of the method.

Reply:

We are glad that the reviewer found this figure helpful.

Reviewer comment:

Paragraph 3.1: I am not sure I understand the pertinence of that part of the results. Is there a point in comparing the changes in climate variables at the global scale? Would it not be more pertinent to look at these changes in different groups of catchments, for instance looking to see if they relate to the categories of "stable", "variable", "alternating" or "shifting" catchments? Or geographically?

Reply:

We thank the reviewer for the valuable feedback and suggestion. The rationale for showing periodic changes in climatic variables, at a global scale throughout the study period, is to provide an overall view of how the key variables that determine the Budyko coordinates (E_A/P and E_P/P), have changed over time. This global perspective helps to understand the trends and catchments' movement within the Budyko space. However, we also agree with the suggestion to provide a more detailed description based on the catchment categories for this analysis. We have carried out the proposed analysis and will incorporate it in the revised manuscript by replacing Figure 4 with Figure R6 (see here below).

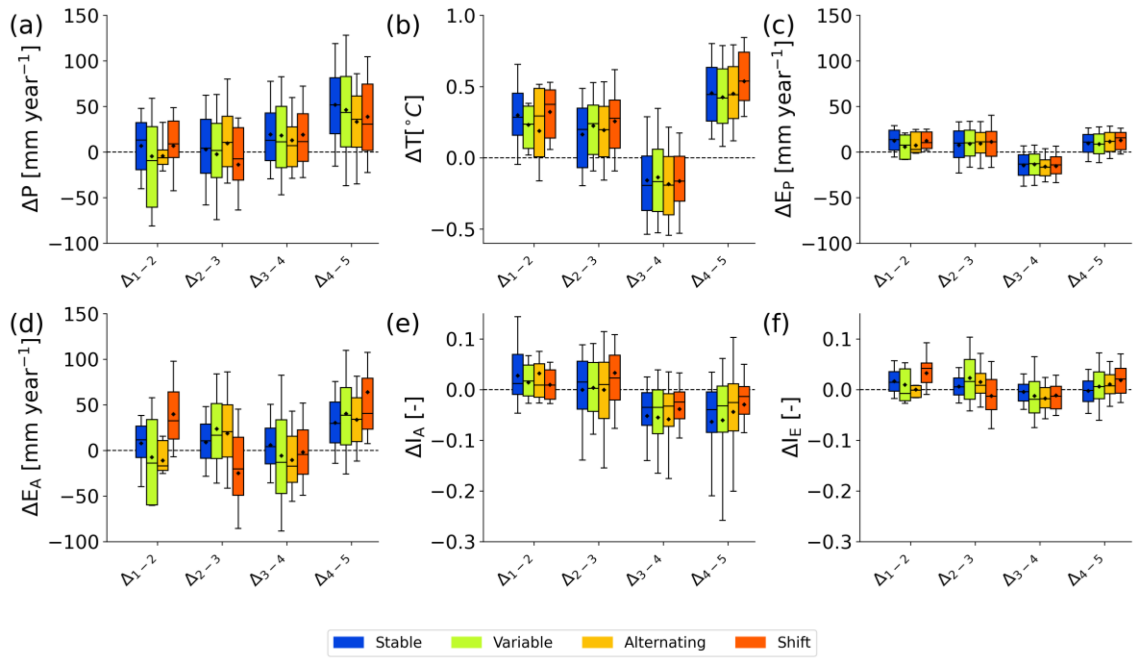

Figure R6: Temporal stability category-wise mean 20-year changes in hydro-climatic variables for the studied catchments between two consecutive periods. a) Precipitation *P*, b) Temperature *T*, c) Potential evaporation E_P, d) Actual evaporation E_A, e) Aridity index I_A, and f) Evaporative index I_E. The boxes represent the 25th to 75th percentiles, while whiskers extend to the 10th and 90th percentiles. Diamonds denote the arithmetic mean, and outliers are not shown

Reviewer comment:

L535: You make the argument here that "the spread around the regional medians consistently decreases with increasing IA across all latitude bands" and therefore that "catchments in more humid regions across the study 535 domain are subject to more pronounced annual water storage fluctuations". However, as you say yourself, the impact of the shift is different depending on the aridity of the catchment. Here this argument would beneficiate from presenting relative changes in discharge or IE rather than changes in the parameter.

Reply:

We completely agree that presenting relative changes in discharge or I_E rather than changes in the parameter would provide a clearer understanding. We have conducted the analysis as suggested and will include the results in the revised manuscript by adding Figure R5 as Figure 7c in the revised manuscript.

Reviewer comment:

L50: The sentence could be reformulated. Maybe the word "described" is unnecessary.

Reply:

We thank the reviewer for the suggestion. We will modify Line 50 as follows:

"The fact that the long-term water balance exhibits such a relatively consistent behaviour across a wide spectrum of hydro-climatically and physio-graphically distinct environments has led to the hypothesis that the general shape of Budyko curves emerges for natural systems in a co-evolution of climate, soil water storage and vegetation properties".

Reviewer comment:

L80: This sentence is also a little awkward. Especially the last part. Maybe separate it in two sentences.

Reply:

We will modify Line 80 as follows:

"In other words, some level of deviation from the expected parametric Budyko curves is to be expected, as different time periods will never be characterized by exactly the same environmental conditions. However, the mechanistic processes that control these deviations, and thus ω, are not well understood".

Reviewer comment:

L257: sentence unclear. Maybe do two sentences: "To do so, for each catchment the up to j = 4 distributions of deviations $\varepsilon_{IE\Delta j}$ from expected $I_{E,i+1}$ between subsequent time periods were compared and analysed for their changes over time. We have followed three sub-steps."

Reply:

We will split the sentence into two, as the reviewer suggested. The modified sentences are as follows:

"***To do so, for each catchment the up to j = 4 distributions of deviations $\varepsilon_{IE\Delta j}$ from expected $I_{E,i+1}$ between subsequent time periods were compared and analysed for their changes over time. We have followed three sub-steps***".

Reviewer comment:

L325: "Combined this led to …" is an awkward sentence.

Reply:

We will modify Line 325 as follows:

"These combined factors led to slightly more arid conditions in the first half of the 20th century, followed by a considerable reduction of aridity index I_A and thus to a shift towards somewhat more humid conditions towards the end of the century across all of the temporal stability categories (Fig. 4e), in which, on average 78% and 75% of the catchments showed decreases in I_A for $\Delta_{3\text{-}4}$ and $\Delta_{4\text{-}5}$, respectively".

Reviewer comment:

L329-330: I do not understand this comment.

Reply:

We have modified these lines as follows:

"The movement of catchments in the Budyko space due to hydro-climatic changes are illustrated in form of rose diagrams in Fig. S1 (Jaramillo and Destouni, 2014). If these movements were driven only by changes in I_A, catchments would be expected to move within the directional range of $0 < \alpha < 45^o$ or $225 < \alpha < 270^o$ (Jaramillo et al., 2022). However, observed movement of catchments are also found in other directions, indicating deviations $\varepsilon_{IE\omega}$ from the expected I_E, as elaborated in detail in Fig. S1".

Furthermore, Section 3.1 of the original manuscript will be revised in the light of new insights obtained after replacing Figure 4 with Figure R6.

Reviewer comment:

L360: Supplementary material should not be cited before figures from the main article in a given paragraph. Otherwise why not include it?

Reply:

We have removed the citation of supplementary material in this paragraph and inserted it after elaboration of the previous comment.

References

Destouni, G., Jaramillo, F. and Prieto, C. 2012. Hydroclimatic shifts driven by human water use for food and energy production. Nature Climate Change 3(3), 213-217, https://doi.org/10.1038/nclimate1719.

Han, Z., Long, D., Huang, Q., Li, X., Zhao, F. and Wang, J. 2020. Improving Reservoir Outflow Estimation for Ungauged Basins Using Satellite Observations and a Hydrological Model. Water Resources Research 56(9), https://doi.org/10.1029/2020wr027590.

Jaramillo, F. and Destouni, G. 2014. Developing water change spectra and distinguishing change drivers worldwide. Geophysical Research Letters 41(23), 8377-8386, https://doi.org/10.1002/2014gl061848.

Jaramillo, F., Piemontese, L., Berghuijs, W.R., Wang-Erlandsson, L., Greve, P. and Wang, Z. 2022. Fewer Basins Will Follow Their Budyko Curves Under Global Warming and Fossil-Fueled Development. Water Resour Res 58(8), e2021WR031825, https://doi.org/10.1029/2021WR031825.

Montanari, A. and Koutsoyiannis, D. 2014. Modeling and mitigating natural hazards: Stationarity is immortal! Water Resources Research 50(12), 9748-9756, https://doi.org/10.1002/2014wr016092.