General comments

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.

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. 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.

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?

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.

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.

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.

Specific comments

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.

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.

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

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?

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.

Technical comments

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

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

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 $IE,i+1$ between subsequent time periods were compared and analysed for their changes over time. We have followed three sub-steps."

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

L329-330: I do not understand this comment.

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