Dear Colleague,

Thank you very much for the thoughtful review of our paper. Please find below our point-by-point reply to comments, including our intended changes to the manuscript. Your comments are in italics, and our replies are in plain text.

Best,
Tessa Maurer on behalf of the co-authors

This paper investigates how the partitioning of precipitation into streamflow and evapotranspiration in 14 catchments in California responds differently to drought versus normal meteorological conditions. The main analysis these catchments undergo is based on the Budyko framework; the response of catchments along a calibrated parametric Budyko curve are assumed to be caused by climate, whereas other movements purely in the vertical direction (i.e. E/P or Q/P) that not follow the Budyko curve are assumed to be caused by a change in the hydrological functioning of the catchment. The results suggest that most runoff changes in catchments are caused by “predictable” shifts along the Budyko curve, but substantial effects of regime shift changes are also observed in many of the catchments. These results are further analyzed by examining the correlation between partitioning shifts catchment properties; low aridity index, high baseflow, shift from snow towards rainfall, and the resilience of high-elevation runoff correlate to increased runoff as a fraction of precipitation during droughts.

Better understanding the effects of drought on the precipitation partitioning across catchments is a relevant research goal. The use of the Budyko framework to study these changes has been widely applied for other aspects of the hydrological cycle (e.g. human versus climatic effects).

While the purpose of the study appears suitable and relevant for HESS, and the methods are largely similar to those widely applied elsewhere in hydrology, I have some questions and comments about the paper that would be good to address.

The paper distinguishes between “regime” shifts, which result from changes in the aridity index along the same Budyko curve, and “partitioning shifts”, which imply a change in the Budyko calibration parameter and thus to the relationships between evaporative demand, precipitation, and ET that govern partitioning of available water. However, what is the physical basis for assuming this? The Budyko framework is developed for characterizing long-term water balances, without any clear theory or evidence that the curve is also appropriate to characterize hydrological change of an individual catchments. I understand...
that many other papers use a similar approach, and it would be unfair to put the burden of proof on you (and not on the dozens of other previous papers), but I struggle to see how application of the framework like this is justified without any clear theoretical or empirical basis that this is a reasonable assumption.

We believe the reviewer makes a valid point that there is no proof of causal connection between changes in the Budyko parameter and basin characteristics, but several studies (e.g., Zhang, Dawes, and Walker 2001; Yang et al. 2007; Jaramillo et al. 2018; Ning et al. 2019) show at least a correlative relationship. Like our approach, these are based on empirical observations, we aim to describe and discuss, although we know that full mechanistic explanations are yet to be formulated. We agree that attempting to establish a quantified, causal relationship between changes to the Budyko parameter (i.e., movement or observed differences in the Budyko space) and specific basin characteristics should be approached with caution. Here, we specifically do not intend our discussion of the mechanisms that possibly influence partitioning shifts to be interpreted as supporting a simple causal relationship, but rather to provide empirical evidence. For this reason, we purposely keep analysis in this section to a minimum, with a calculation of correlation coefficients as the most rigorous quantification. We wish to present a non-exhaustive list of plausible, interrelated causes for unpredictable, nonlinear changes in a basin water balance that are consistent with existing literature on basin drought impacts. We believe these initial investigations can provide the basis for more exhaustive future work aimed at defining a mechanistic framework, but also that it is still valuable to show that droughts may lead to a deviation from a catchment’s non-drought Budyko curve.

We also agree that adjustments to the Budyko framework outside the original intended application must be carefully considered. For this reason, we coupled the Budyko framework with the abcd model to be able to apply it on an annual basis. While this approach is not perfect (see following response), we believe that errors are within the range inherent in any of the other measured data and spatial grids used to characterize the water balance. In terms of spatial scale, previous work (e.g., Bai et al. 2020; Li et al. 2013) that examines the Budyko framework for smaller catchments does show that other physical processes than water demand and availability are more dominant at these scales than at larger ones. However, we believe this is understanding is consistent with our assessment that other mechanisms and basin characteristics influence movement within the Budyko space.

We will clarify the limitations of the Budyko framework in the revised manuscript, as well as the empirical nature of this study and the opportunity for more theoretical research in the future, with specific mention of the potential for the calibration parameter omega to characterize the basin. We will further clarify in Section 4.3 that our discussion of catchment feedback mechanisms is intended to serve as a starting point for further modeling efforts that could establish quantified causal connections using different approaches.

Half of the catchments seem to violate the conservation of mass (i.e. \( ET > P + \Delta S \)) in drought conditions. Does this not suggest that there is something off with the estimates on which all conclusions are based?

Thank you for bringing up this point. We do not expect that the abcd calibration model used to estimate subsurface storage is a perfect model, and underestimations in the level of subsurface withdrawal in a given year would result in \( ET > P + \Delta S \). However, for the vast majority of years, this method was successful in accounting for subsurface storage and bringing the available water within the theoretical bounds of the Budyko framework. Overall, we believe the errors fall within a similar level of uncertainty to that already extant in the spatial datasets and readings of full natural flow also used in the analysis.

Since the errors may also originate in the underlying spatial datasets, we opted not to
estimate withdrawal from the subsurface as the residual of the other components of the water balance for which we have spatial data (ET, P, and Q). This method would prevent any water year from falling above the water limit line, but doing so would assign all uncertainty and error to the $\Delta S$ component and also assume that all water in the system is available for plant use (ignoring, for example, percolation to deep groundwater). This decision was also supported by Reviewer 1 (see comment to Line 107). In light of these concerns, we felt it would be overly simplistic to use the water balance residual. In addition, since as the reviewer notes, most years that fell above the water limit line were drought years, that it would bias the results to discard those years. We note in lines 202-3 that the years that violate the water limit line amount to less than 3% of all basin-water years.

We will provide more detail about our decision-making process when selecting the $abcd$ model for estimating subsurface storage and provide more explanation of the sources and expected ranges of uncertainty. We will also discuss the implications of those years violating conservation of mass for the results of the analysis, including that the current calibrations of $\omega$ may lead to slight overestimations of the partitioning shifts in basins where data fall above the water limit line.

Detailed comments

L15: would it be possible to say something more precise than this very generic closing part of the statement?
We agree this sentence could be more specific. In the revised manuscript, we will specify that this work has relevance for water resources managers (e.g. dam operators, utility companies, and water agencies) to be better able to 1) forecast changes to runoff during droughts based on available climate data and 2) understand under what circumstances and to what extent their forecasts may be less reliable due to nonlinear basin-climate feedbacks. We will further specify that this work is of particular benefit in arid, drought-prone regions like California.

L26-27: I think this statement needs to be backed up by some references that support this is a widely accepted fact. Personally, I am aware of the possibility this is true, this stating it as an almost universal fact seems like a bit of a stretch (to me).
Thank you for pointing out this oversight. Several of the papers currently cited in the manuscript, including Saft et al., 2016; Avanzi et al., 2020; Tian et al., 2020; and Potter, Petheram, and Zhang, 2011 have observed a shift in the precipitation-runoff relationship during drought periods in arid regions, including but not limited to California. We will add these citations to this sentence in the revised manuscript.

L201: How can the relative error have units mm, and are these relative errors calculated adding up over and underestimations of runoff, causing the overall relative error to be small?
Thank you for catching this. The label of millimeters was an error on our part; this should be a unitless number. We will correct this in the main text and supplement in the revised manuscript. Further, we double checked the water balance error values, and they did sum both over- and under-estimations. We have calculated the relative error using the absolute value of the residual. Across the basins, most relative errors in flow were less than 15% with a maximum value of 36%; we recognize that the maximum relative error value has increased at this point, but still believe that the results are functional to our scope given the high Nash-Sutcliffe Efficiency values (see Table S3 in the Supplement) and the fact that we did not use this method for all water balance components, but only as a way to decouple $\Delta S$ values from ET and P, which we believe was important to do (see response to the reviewer’s second major question).

L201: is the accurate simulation of runoff an assumed indication that the other fluxes ($ET$
and delta S) also are reliable?
Thank you for asking this clarifying question. Yes, the abcd model is calibrated to streamflow as suggested in Thomas 1981 and it is assumed the model then performs well for internal variables. We specifically did not attempt to calibrate the abcd model to ET because, as noted in the previous responses, our aim was to estimate ΔS in a way that was decoupled from the other water balance components in the Budyko framework.

**References**


