Reply on RC1

Tessa Maurer et al.

Author comment on "Drivers of drought-induced shifts in the water balance through a Budyko approach" by Tessa Maurer et al., Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2021-55-AC1, 2021

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

Review on "Drivers of drought-induced shifts in the water balance through a Budyko approach" by Maurer et al. submitted to HESS

General comments

The authors compiled temperature, potential (PET) and actual (ET) evapotranspiration, precipitation (P) and runoff (Q) data for 14 Californian catchments from a 34-year period with three drought periods to analyze the dependence of ET and Q on P and PET using the Budyko framework. By an innovative approach they quantitatively distinguish drought-induced changes that would be expected within the Budyko framework ("regime changes") from "partitioning" changes that can only be explained by a shift of the curve parameter(s) (in this case, the omega parameter of the Fu equation). They find that regime changes dominate observed changes in ET and Q, while partitioning changes still add non-negligibly to changes especially in some catchments. The topic is relevant to HESS, the methodology sound and original and the results can help understand catchment responses, with the proposed methodology being a potentially useful comparetively simple tool for many other studies in the future. While a number of suggestions for improvement are given below, many of them (hopefully self-explaining which ones) are optional such that from my point of view the manuscript can be accepted after minor revisions. The maybe most relevant single suggestion is avoiding misinterpretations by readers about the degree of novelty of the approach by better acknowledging existing literature on interpreting and decomposing changes in Budyko space (see detailed comments on L84 and L263).

Thank you for your comments and, in particular, your point about existing literature. We believe that appropriately situating our work is fundamental to achieving strong science and therefore agree that it should be better laid out for the reader. Please see our

responses to your specific comments below.

Specific and technical comments

L42 "wetter, monsoon region in China": something seems to be missing in sentence, check

We intended "monsoon" to act as an adjective in this sentence, but understand the confusion. We will change the sentence to "wetter, monsoon-dominated" in the revised manuscript.

L72: be=>been
We will incorporate the comments above in the revised manuscript.

L79: A recent study which among others also briefly looks at drought in a Budyko framework: https://doi.org/10.1098/rstb.2019.0524
Thank you for bringing this article to our attention. We will incorporate it into our discussion of previous drought assessments using the Budyko framework.

L84: "new framing": This is a bit misleading. Although I'm not aware of your exact methodology (way of decomposing) having been applied to your exact question (distinguishing two directions of drought effect) before, the general idea of using movement along vs. perpendicular to curves in Budyko space to distinguish processes (e.g. climate variability from land-use) is quite widespread, occasionally also including quantification efforts. It would be good to re-check the literature, cite a few examples and adapt the wording. Starting points might be e.g. https://doi.org/10.5194/hess-22-567-2018 (which is already cited but not with reference to the decomposition idea) and doi:10.1029/2011WR011586. It would be good to discuss somewhere how your suggested terms "regime shift" vs. "partitioning shift" relate to already introduced terms in such sources. Both, differences in methodology and scientific reasons, can jusitify your choice of terms (e.g. "climate" vs. "residual" in Jaramillo et al. implies a claim about the causes which it seems you could partly disprove for some catchments); but still it is important for readers that not each paper "reinvents the terminology wheel" without referring to past suggestions.
Thank you for bringing this ambiguity to our attention. We did not intend to mislead the reader in suggesting that such an approach is entirely novel, but merely its application to drought conditions. We certainly agree that the papers suggested by the reviewer would provide both more clarity for the reader on the contribution of this paper as well as a sound basis for our approach to droughts. We will clarify the language at both lines 84 and 263, as well as ensure that the abstract, introduction, and discussion are clear on the specific novel contribution of this method to drought contexts.

L104 PRISM may be a well-known climatology dataset in the US but the description focuses on the interpolation/regression method and does not specify the ultimate source of the original data input to the downscaling / interpolation (e.g. station observations or reanalysis?). Please add a sentence on that so readers all around the world can better judge the potential strengths and weaknesses of the data.
Thank you for making this point. Fundamentally, PRISM spatial maps are created based on a regression between digital elevation models (DEM) and a large collection of ground-based precipitation and temperature data, including from the National Weather Service Cooperative Observer Program and Weather Bureau Army Navy stations; U.S. Department of Agriculture National Resource Conservation Service Snow Telemetry (SNOTEL) and snow courses; U.S. Department of Agriculture Forest Service and Bureau of Land Management Remote Automatic Weather Stations; and California Data Exchange Center (CDEC) stations. Depending on the source of the data, different quality control methods

were used. Stations are weighted by a variety of factors, including clustering with other stations, distance to pixel, elevation, coastal proximity, and topographic facet. After initial values have been calculated for each pixel, maps are subject to final steps to ensure spatial consistency, such as bound checks on vertical gradients between neighboring cells. We will add a brief description of these details to the revised manuscript.

L105: I guess that inavailability of radiation data was the reason to choose a comparatively crude, less known, semi-empirical PET approach such as Hamon? Here or later e.g. in the discussion, it would be good to comment on the effect it might have had on results.
You are correct that the Hamon method was selected since it was usable with the relatively limited spatial data available. We felt that interpolating and distributing very sparse radiation data could lead to even more uncertain estimates, despite the method itself being more sophisticated. We therefore resolved to use a standard approach that has already been applied to environments in the Sierra Nevada (Rungee, Bales, and Goulden 2019). We will include more background on this decision in the revised manuscript.

L107: Please add one or few sentences on the cornerstones of the ET estimation methodology of Roche et al. 2020. Together with the runoff mentioned in the next sentence, you have everything you need to "close" the water budget (i.e. check for gaps and surpluses in P = ET + Q) and / or quantify the Budyko input parameters P, PET, ET and Q without determining any of them residually, which is good; however, this is only perfectly true if the methodologies to quantify each of them do not implicitly use one or more of the other parameters. As far as I can judge from a quick glance into Roche et al. 2020, this is not a (big) problem here but readers should be put in the position to get a first idea without reading the reference.
Thank you for raising this important and valid point. Our data sources are not perfectly independent as the ET regression we used from Roche, et al. uses both NDVI and precipitation. Despite this, we still felt this dataset gave more reasonable ET values for our study area than the calibration based only on NDVI because large portions of the northern Sierra Nevada are significantly wet in winter and including precipitation as a predictor improves ET estimates in such regions (see Roche et al. 2020 for a discussion on this). We have addressed these uncertainties previously (Avanzi et al. 2020), and we found in this study that estimates of the four water balance components tally with expectations and previous work (Avanzi et al. 2020; Rungee, Bales, and Goulden 2019; Roche, Goulden, and Bales 2018; M. L. Goulden et al. 2012; Michael L. Goulden and Bales 2014). This approach is, thus, comparatively established at this stage and provides among the best data-driven estimates available for the region. We will add these details and discussion of the implication to the revised manuscript.

L113: To build further on the comment before, it would be good to report (here, results section or supplement) how large the needed corrections to P were and how much they differed between basins, to give an idea of the overall quality of the dataset - or rather, it's weakest (most assumption-dependent) parameter, which might actually have been ET rather than P.
Thank you for making this point. The largest (i.e., highest magnitude) adjustment factor for precipitation was 85.7 mm in Shasta, which also had the largest adjustment as a percentage of long-term average precipitation (7.6%). Wetter basins tended to have higher adjustment factors. (The minimum adjustment factor was in the Stanislaus, at 2.35 mm and 0.3% of average annual precipitation). We will add the full set of adjustment factors, in depth and as a percentage of precipitation, to the supplement.

L119: Mention both PET/P and ET/P consistently as symbols, in words, or both.
We will make this revision in the revised manuscript.

L123: (Du et al., 2016) => Du et al. (2016). Same at L126 for Thomas and possibly more places.
We will make this correction in the revised manuscript.

L125-133: Difficult to follow. Consider rewording and/or showing the equation(s), if needed in the supplement.
Thank you for bringing this to our attention. We will revise the section to include the equations for the *abcd* model so the description is more concrete.

L152: Remain consistent about writing omega as a symbol or a word.
We will make this revision in the revised manuscript.

L156-163 and Figure 2: The description in the text at the end of section 2.3 and the graphical description in Figure 2 b do not seem to match. I believe the text is "correct" in the sense that the regime shift is consistent with its definition ("what would be expected according to Budyko/Fu") and the partitioning shift is the rest such that both add up to the total observed shift in ET/(P-deltaS). However for the Figure to match this, the vertical blue "partitioning shift" arrow would need to start near the tip (not the foot) of the red arrow / near the centroid of the + symbols, and its tip and the triangles (which are not to refrrred to in the text, I think it should be the true observed data of the drought years?) should be further to the (upper) right on the omega=3 line. The difference between these two ways of illustration matters because the distance between the two Fu lines changes with aridity index.
Thank you for pointing out the lack of clarity here. In the conceptual figure, we were trying to distinguish the changes due to one type of shift specifically. The triangles therefore represent the hypothetical points that would be observed if only a partitioning shift had occurred. The summation of these changes would be off to the upper right and is what we see in the observed data. We will clarify this point and also add a fourth cluster of points on the conceptual figure to represent the true observed data of the drought years, as the reviewer said.

Figure 2: Compute more nodes of the Fu equation to make the lines smoother
We will make this change in the revised manuscript.

*L167: with respect *to* runoff?*
We will make this correction in the revised manuscript.

L169: How did "amount of available storage" and the methodology used to estimate it relate to the deltaS values and abcd model used to estimate it earlier?
Thank you for asking this; we agree that the phrase "amount of available storage" is vague. In the cited studies, it refers to plant-accessible water storage, in other words, the water that is available to buffer ET against precipitation deficits. The cited papers quantify this value in different ways, but it is conceptually the same as the value we estimate using the *abcd* model. We will change the wording of the phrase to "amount of plant-accessible storage (here, the value estimated as ΔS)" in the revised manuscript.

*L175: is estimated *from* average...?*
We will make this correction in the revised manuscript.

L192-193: Unclear: If you refer to changes between droughts (as opposed to between drought and non-drought), then why is there only one difference value per basin given although there were threee drought periods?

Thank you for bringing this lack of clarity to our attention. We are reporting average differences for each regions between all drought periods collectively and all non-drought

periods collectively. We see how this is confusing when compared with Fig. 3 and will clarify this point in the revised manuscript.

L201 / S2: How can a relative error still have units of mm? In case of doubt, specify relative to what / briefly explain the methodology.
This was an error on our part; it should be a unitless number, calculated as the summation of error divided by the sum of the observed values. We will correct this in the text and supplement.

L202-203: Were these years excluded from the calibration? Not that I'd like to suggest to do so, it's just that the curent wording almost seems to suggest so.
Thank you for asking about this; those years were not excluded from the calibration, so we appreciate knowing about that lack of clarity. We will explicitly state that they were included in the revised manuscript.

Figure 5: If regime shifts and partitioning shifty behave strictly additive (without any nonlinear/interactive terms), which it looks like and would be consistent with the methodology description near L163, wouldn't it be more intuitive to use stacked columns? E.g. plotting partitioning shift on top of regime shift - if they have the same sign, the total column length is the total shift, if not the resulting total shift could still be a point marker within the column?
This is a good suggestion. Our thinking in plotting them separately was so the reader could quickly distinguish the basins where they shared the same sign from those where they did not, but we believe the reviewer's suggestion may make that even easier. We will test this option during manuscript revisions and incorporate it assuming it meets this need.

Sect. 3.2 in general: While excessive, or rather wrongly interpreted, significance testing is meanwhile disputed (e.g. https://www.nature.com/articles/d41586-019-00857-9), could you think of a simple methodology to roughly transfer what is said from the K-M-tests about changes in the two indices (L211) to the importance comparison between regime and partitioning shifts? While the text qualitytively already tries to convey this message, inspection of the point clouds in Fig. 4 seems to suggest even more that only few catchments (maybe only Kaweah, Kern and Tule) saw a "significant" partitioning shift, while the shifts in all other catchments might be within the range of uncertainty indicated by the scatter of annual data, and thus statistically indistinguishable from "pure regime shifts". Maybe a simple way to try to do this could be to compare the partitioning shift to what would have been significant at p=0.01 or 0.05 in the total ET/PET shifts. A more complex way could be a Monte Carlo type approach where years are randomly removed from the drought / non-drought subset, but maybe this would be overdoing it.
Thank you for raising this point; we agree that the implications of p-values and statistical significance should be better articulated in the literature. We further agree that we can clarify our use of the K-S tests and their implications. Since only two *omega* values were calculated per basin, it was not possible to directly establish significance with regards to the partitioning shifts. Instead, we used the K-S tests to determine if there was, as a baseline, observable movement along each axis. We found that shifts in all basins and along both axes were significant to the *alpha*=0.01 level with the exception of the ET/P – ΔS shift in the Feather, which was significant to the *alpha*=0.05 level (see Table S4 in the Supplement). We believe that this application of the K-S tests as a simple way to compare the distributions of each value (ET/P – ΔS and PET / P – ΔS) during drought and non-drought periods was appropriate. However, we did not mean to imply that the significance of these shifts along each axis are equivalent to significance in the partitioning shift. We will clarify this point in the revised manuscript. We will also calculate and report the results of a K-S test for the shift in ET/PET values for each basin, again recognizing that this is not equivalent to a significance test for the partitioning shift. We will further re-evaluate the language in Section 3.2 to ensure that we do not misstate the implications of

significance to a given p-value (as described in the article cited by the reviewer).

L246: "Tule. northern..." => "Tule. Northern..."
We will make this correction in the revised manuscript.

L263: See comment on L84.
Please see our response to comment on L84.

L2278: 10 times less: Is this mentioned anywhere in the results section or at least supplement? Sorry if I overlooked it, but it seems to come a bit out of nowhere here.
Thank you for pointing out an oversight on our part in not including the underlying data. This is based on the annual estimates of change in storage from the *abcd* model and the annual precipitation from PRISM. To avoid a large table with thirty years of data for each basin, we will add a table to the supplement with the maximum ratio of subsurface storage change to annual precipitation for each in basin.

L353: "become drier" - specify in which sense (e.g. less runoff?)
We agree this is unclear; in this case, "drier" means more arid (as measured by the aridity index). We will make this change in the revised manuscript.

Figure 6 and Table 2: Maybe I overlooked something but other than for the aridity index threshold, which was explained and discussed at L233, the thresholds for the other three parameters are poorly or not connected to the manuscript text (both in terms of explaining how they were determined and of discussing their implications).
The thresholds discussed were identified manually from the data. They were meant to serve as estimates and are somewhat subjective. For example, we erred toward selecting round numbers, but in most cases, they can be changed up or down slightly and will give the same results in terms of classifying basin behavior. In the case of the aridity index, the threshold we identified happened to coincide with those identified independently in existing literature. In the revised manuscript, we will adjust in the language in Sections 3.3 and 4.3 to specify that the thresholds are estimated in order to identify tendencies in basin behavior, not hard-and-fast cutoffs. In Section 4.3, we will make explicit reference to the threshold values to better connect them to our discussion of basin behavior.

References

Avanzi, Francesco, Joseph Rungee, Tessa Maurer, Roger Bales, Qin Ma, Steven Glaser, and Martha Conklin. 2020. "Climate Elasticity of Evapotranspiration Shifts the Water Balance of Mediterranean Climates during Multi-Year Droughts." *Hydrology and Earth System Sciences* 24: 4317–37. https://doi.org/10.5194/hess-24-4317-2020.

Goulden, M. L., R. G. Anderson, R. C. Bales, A. E. Kelly, M. Meadows, and G. C. Winston. 2012. "Evapotranspiration along an Elevation Gradient in California's Sierra Nevada." *Journal of Geophysical Research*: Biogeosciences 117 (3): 1–13. https://doi.org/10.1029/2012JG002027.

Goulden, Michael L., and Roger C. Bales. 2014. "Mountain Runoff Vulnerability to Increased Evapotranspiration with Vegetation Expansion." *Proceedings of the National Academy of Sciences of the United States of America* 111 (39): 14071–75. https://doi.org/10.1073/pnas.1319316111.

Roche, James W., Michael L. Goulden, and Roger C. Bales. 2018. "Estimating Evapotranspiration Change Due to Forest Treatment and Fire at the Basin Scale in the Sierra Nevada, California." *Ecohydrology*, no. March: 1–10. https://doi.org/10.1002/eco.1978.

Roche, James W., Qin Ma, Joseph Rungee, and Roger C. Bales. 2020. "Evapotranspiration

Mapping for Forest Management in California's Sierra Nevada." *Frontiers for Global Change*. https://doi.org/10.3389/ffgc.2020.00069.

Rungee, Joseph, Roger Bales, and Michael Goulden. 2019. "Evapotranspiration Response to Multiyear Dry Periods in the Semiarid Western United States." *Hydrological Processes* 33 (2): 182–94. https://doi.org/10.1002/hyp.13322.