Summary

In this study, Massari et al identify a shift in the rainfall-runoff relationship during multi-year droughts compared to normal years. In addition, they attribute this shift to an enhancement in the evaporation, especially in dry, warm, and water limited environments. The manuscript is well structured, and the conclusions are based on sufficient analysis of results. However, I have major concerns about the novelty of the study, the importance of the study's results, and some methodological details. I elaborate on this below, including some minor comments.

Major Comments

1. ***Novelty: In my opinion the results and the conclusions drawn are fairly logical and obvious. For example, precipitation droughts do not occur in isolation. They are generally accompanied by enhanced temperature (not all the time). In that scenario, an increased evaporation is expected, and it follows that runoff would reduce compared to normal years. The same applies for the results presented for energy-limited and water-limited environments. I request the authors to specifically elaborate on the novelty of their findings and explain how these findings move forward catchment hydrology.***

We thank the reviewer for this comment and we will elaborate in the revised manuscript on several points of novelty we see in this research. All these points of novelty start from the overarching evidence that the role played by evaporation during droughts is far from the simplistic assumption that higher temperatures lead to higher evaporation. The reasons for that -- and so the main points of novelty we see in the manuscript -- are provided below:

1) Meteorological Droughts (i.e., we mean here a deficit of precipitation) are not necessarily warmer than wet periods. A good example here is Figure 6b, where we show that 1992-1993 in Spain experienced both a negative anomaly in terms of potential evaporation and a negative anomaly in terms of precipitation. This points to the precipitation-temperature-evapotranspiration relationship being more nuanced than assuming that precipitation deficit will concur with higher temperatures.
2) Temperature is not the only driver of evaporation, especially in water-limited environments. For example, the volume of water that is actually evaporated into the atmosphere can diverge from potential evaporation as dictated by the atmosphere, especially in conditions of soil water stress. This is because over drier soils, despite higher temperatures there is no more water to transpire or to evaporate. This is true especially when prolonged precipitation deficits occur, such as those experienced during multi-year drought periods. While for single dry years catchment storage can sustain a higher evaporation rate, this is not true in general for multi-year droughts and is very much related to the role played by catchment storage and by the vegetation response to water stress (see point 3 and 4 below). In summary, precipitation deficit and pre-existing catchment conditions (e.g., storage) modulate the response of evaporation to temperature anomaly during multi-year droughts in ways that are only partially quantified and predictable at this stage. This makes multi-year droughts essentially different from stand-alone dry years and motivates specific research on that topic.
3) In this framework, the role of catchment storage is often overlooked. Evaporation enhancement during multi-year droughts occurs only if the catchment has sufficient storage to sustain evaporative demand for a prolonged time and vegetation has the

ability to access this storage. For instance, weathered bedrock and rock moisture have been identified as significant water sources for plant transpiration in addition to soil water in Mediterranean climates (Goulden et al. 2014, Hahm et al. 2020). If access to the whole of the regolith is not possible, or regolith moisture has already been depleted, then transpiration might even decrease as a consequence of plant mortality (Karban et al. 2017, Senf et al. 2020). At least, ET might have a very non-linear behaviour as a consequence of how different vegetation species respond to water stress (McDowell et al. 2008, Domec & Johnson 2012, Gentilesca et al. 2017). In our study, we provided evidence that the increased evaporation demand is a common feature over different climates during multi-year droughts (so for very long dry periods) and quantified its impact on runoff reduction. Both these results were not fully expected a priori and provide valuable information for further understanding (and managing) the relationship between meteorological and hydrologic droughts.

4) Last but not least, the role played by increased evaporation during droughts and its impact on soil moisture have been largely demonstrated (Teuling et al. 2013, Miralles et al., 2019). However, knowledge of their impact on runoff is still very limited. Here, we show that this impact can be significant both over water- and energy-limited regimes.

In the revised version of the manuscript we will accordingly revise the introduction section to better highlight the novelty of this study.

Karban, R., & Pezzola, E. (2017). Effects of a multi-year drought on a drought-adapted shrub, Artemisia tridentata. *Plant Ecology*, *218*(5), 547-554.
Senf, C., Buras, A., Zang, C. S., Rammig, A., & Seidl, R. (2020). Excess forest mortality is consistently linked to drought across Europe. *Nature communications*, *11*(1), 1-8.
Goulden, M. L., & Bales, R. C. (2014). Mountain runoff vulnerability to increased evapotranspiration with vegetation expansion. Proceedings of the National Academy of Sciences, 111(39), 14071-14075. https://doi.org/10.1073/pnas.1319316111
Hahm, W. J., Rempe, D. M., Dralle, D. N., Dawson, T. E., & Dietrich, W. E. (2020). Oak transpiration drawn from the weathered bedrock vadose zone in the summer dry season. Water Resources Research. 56 (11): e2020WR027419, 56(11). https://doi.org/10.1029/2020WR027419
Gentilesca, T., Camarero, J.J., Colangelo, M., Nolè, A., Ripullone, F., 2017. Drought-induced oak decline in the western Mediterranean region: an overview on current evidences, mechanisms and management options to improve forest resilience. iForest - Biogeosciences and Forestry 10, 796. https://doi.org/10.3832/ifor2317-010
Miralles, D.G., Gentine, P., Seneviratne, S.I., Teuling, A.J., 2019. Land–atmospheric feedbacks during droughts and heatwaves: state of the science and current challenges. Annals of the New York Academy of Sciences 1436, 19–35. https://doi.org/10.1111/nyas.13912
Teuling, A. J., Van Loon, A. F., Seneviratne, S. I., Lehner, I., Aubinet, M., Heinesch, B., Bernhofer, C., Grünwald, T., Prasse, H., and Spank, U.: Evapotranspiration amplifies European summer drought: EVAPOTRANSPIRATION AND SUMMER DROUGHTS DROUGHTS,
Geophysical Research Letters, 40, 2071–2075, https://doi.org/10.1002/grl.50495, 2013

2. ***Implications: I am not very sure that the change in the rainfall-runoff relationship (even the maximum of -40%) is very significant in terms of absolute terms. This is because, it is fair to assume that during multi-year droughts, especially in arid watersheds, the amount of rainfall is reduced significantly (upto -185% according to the findings of this study). Therefore, how does this translate to any significant change in runoff? It would be useful if the authors provide an idea of the change in***

runoff in absolute terms and not only percentages, as I feel that the implication of the change in rainfall-runoff relationship may not be significant at all.

We agree with the reviewer that analyzing changes in precipitation-runoff relationships in absolute terms would be informative at specific locations and to support decision making (we also provided an example at lines 192-198). At the same time, we note that focusing on absolute terms would be challenging given the variety of catchments we considered. This is because absolute changes in runoff are strongly dependent on local precipitation values, an effect that we bypassed by focusing on percentage changes. Also, percentage changes have already been used in previous papers on this topic (Saft et al. 2016, Avanzi et al. 2020, Tian et al., 2020, Alvarez Garreton 2021, Peterson et al. 2021), which facilitates discussion with related research

Saft, M., Peel, M. C., Western, A. W., and Zhang, L.: Predicting shifts in rainfall-runoff partitioning during multiyear drought: Roles of dry period and catchment characteristics: RUNOFF SHIFTS DURING DECADAL DROUGHT, Water Resources Research, 52, 9290–9305, https://doi.org/10.1002/2016WR019525, 2016.

Avanzi, F., Rungee, J., Maurer, T., Bales, R., Ma, Q., Glaser, S., and Conklin, M.: Climate elasticity of evapotranspiration shifts the water balance of Mediterranean climates during multi-year droughts, Hydrology and Earth System Sciences, 24, 4317–4337, https://doi.org/https://doi.org/10.5194/hess-24-4317-2020, 2020.

Tian, W., Bai, P., Wang, K., Liang, K., and Liu, C.: Simulating the change of precipitation-runoff relationship during drought years in the eastern monsoon region of China, Science of The Total Environment, 723, 138 172, https://doi.org/10.1016/j.scitotenv.2020.138172, 2020

Alvarez-Garreton, C., Boisier, J. P., Garreaud, R., Seibert, J., and Vis, M.: Progressive water deficits during multiyear droughts in basins with long hydrological memory in Chile, Hydrology and Earth System Sciences, 25, 429–446, https://doi.org/https://doi.org/10.5194/hess-25-429-2021, 2021.

Peterson, T. J., Saft, M., Peel, M. C., & John, A. (2021). Watersheds may not recover from drought. *Science*, *372*(6543), 745-749.

In the revised manuscript, we will expand on the already reported implications of absolute runoff reduction at lines 192-198 by providing more concrete examples.

3. ***Conclusions: Likewise, the two main conclusions regarding the need for a) better calibration of rainfall-runoff models and b) better representation of different processes in the conceptual models, are not very novel. It is well known that conceptual models do not account for non-stationarity very well, and the response of the watersheds during multi-year droughts is a specific case of non-stationarity.***

We agree with the reviewer that this is well known. However, the need for improved calibration protocols and a better representation of the variety of processes involved in the water balance have been usually demonstrated by analysing the performance of conceptual (and even "physically-based") models in reproducing floods. On the other hand, few studies (Saft et al. 2016, Avanzi et al. 2020) highlighted the sub-optimal performance of these models in reproducing total runoff during droughts. Thus is of paramount importance for water resource management.

We also note that the main findings of this manuscript are based on data, and we did not consider any hydrologic model. Thus, we will take this comment as an opportunity to revise section 4 and focus more extensively on process-based conclusions rather than modeling implications.

Similarly, how feasible is the inclusion of complex coevolution mechanisms in simple rainfall-runoff models. I suggest the authors focus on the implications of the study beyond improvement of conceptual models.

This is an interesting point for future work. In the revised version of the manuscript, we will improve the discussion of this part by including examples of studies that attempted to do that (see for example Hughes et al. 2021). See also our reply to point #3 above.

Hughes, J., Potter, N., Zhang, L., & Bridgart, R. (2021). Conceptual Model Modification and the Millennium Drought of Southeastern Australia. Water 2021, 13, 669.

4. ***Methodology: There are several unclear methodological decisions which needs to be clarified***

    - ***Was the effect of increasing temperature trends in these watersheds taken into account. Although it may not have an effect in a 3-year period, it may have an effect on the rainfall-runoff relationship in multi-year droughts in the early years (1980s) compared to the later years (2000s). Was this explored?***

We thank the reviewer for this comment. We did not take this into consideration, because -- as the reviewer correctly pointed out -- we have focused specifically on multi-year drought periods. Our multi-year drought periods are mostly concentrated above the 90s, so trends in temperature and the associated long-term coevolutionary mechanisms are likely to play a minor role here. We will provide a discussion on this in the revised version of the manuscript.

- ***The authors need to better justify the use of only 3-years for multi-year drought definition. How was this number arrived at? In addition, how was the threshold of SPI > 0.15 selected?***

To our knowledge, there is neither consensus nor an accepted definition of multi-year drought in the literature. Beside that fact that "multi-year" means "longer than one year", past studies on that (Saft et al. 2016) used similar criteria as ours for defining a multi-year drought period. To validate our methodology we also cross compared the results of our definition with past important multi-year drought events that have impacted Europe (Parry et al. 2012, Spinoni et al. 2015, see lines 178-185 of the original manuscript). We also tested different thresholds (i.e., larger than 3 years), without noticing any significant difference in the results. We will be more explicit on this in the revised manuscript.

Parry, S., Hannaford, J., Lloyd-Hughes, B., and Prudhomme, C.: Multi-year droughts in Europe: analysis of development and causes, Hydrology Research, 43, 689–706, https://doi.org/10.2166/nh.2012.024, 2012
Spinoni, J., Naumann, G., Vogt, J. V., and Barbosa, P.: The biggest drought events in Europe from 1950 to 2012, Journal of Hydrology: Regional Studies, 3, 509–524, https://doi.org/10.1016/j.ejrh.2015.01.001, 2015.

- ***Why was the representative annual precipitation estimated as the mean of average and minimum precipitation and not just average annual precipitation?***

The precipitation input we used as representative precipitation was the average annual precipitation for each basin. We will specify this more clearly in the revised version of the manuscript in section 2.1.

- ***I do not understand why two different precipitation datasets were used. The provided justification does not explain possible discrepancy between drought definition and annual precip which may lead to differing anomalies. Why should they be independent?***

We agree that more clarity is needed when presenting our chosen datasets. We used two different precipitation datasets here because we wanted to rely upon observed-only precipitation and river discharge estimates for the definition of the precipitation-runoff relationship as we think this makes our results stronger (we are not relying on any model for that part of our analyses). On the contrary, for the drought definition only we relied on ERA5, because we wanted to make sure that the drought period identified was the same used to calculate the evapotranspiration anomalies. Doing otherwise would have led to questionable results due to an inconsistency between the identified drought periods and ET anomalies. We will improve lines 166-175 to explain this better.

- ***In Figure 5, was the water/energy limited watersheds classified based on only the drought period? It may be that a generally energy-limited watershed may transition to a water-limited watershed during multi-year droughts.***

This was done based on long-term PET, P, and ET, as suggested in the literature to define (qualititatively) the type of climate that each studied catchment experiences on average. This is of course only a qualitative distinction, as transitions between the two regimes are possible. We will add this comment to the manuscript and will specify that they do not impact the generality of our findings.

- ***I am not sure if discussing the importance of basin storage sustaining ET during droughts is relevant to the discussion here as it focuses on rainfall-runoff relationship. Is the argument that enhanced ET may actually be due to storage changes and not because more precipitation is being converted to ET? It would be helpful if the authors can elaborate.***

We think that the role played by the basin storage is fundamental as it modulates the response of the catchment to water stress (see our response to the point 1 and a also recent publication on Science where a similar procedure used in this study was used, Peterson et al. 2021). Thus, discussing storage is essential to better comprehend changes in the precipitation-runoff relationship and put these changes in a holistic perspective. In other words, we argue that changes in precipitation-runoff relationship cannot be understood without a water-balance perspective that embraces storage and evapotranspiration.

Peterson, T. J., Saft, M., Peel, M. C., & John, A. (2021). Watersheds may not recover from drought. *Science*, *372*(6543), 745-749.

5. ***Minor Comments***

- ***There are several grammatical errors in the manuscript. I request the authors to correct them. For example Line 10, "less than what expected" should be "less than what is expected". Line 120, "...it was not used to define the droughts…" should actually be "...it was used to define the droughts…".***

We will carefully revise the manuscript.

Regarding the second example, that sentence refers to runoff, so it seems correct to us to state that runoff was NOT used to define the drought. We will modify in:
"...runoff response, therefore the latter was not used to define the drought"

- ***Line 207. I do not understand how the fact that having only 2 basins showing a positive shift is an indicator of the "high control" of the experiment or high quality. I do not follow the reasoning here.***

We are sorry for the confusion here. We did not mean that "having only 2 basins showing a positive shift is an indicator of the "high control" of the experiment or high quality". We meant that we cannot exclude that those positive shifts are spurious due to possible errors in the data, as such positive shifts only occur for two catchments. We will revise the sentence.