Dear authors,

As you have seen during the discussion phase, your manuscript has received comments from 2 referees. While both referees seem to agree that the study is interesting and potentially suitable for publication in HESS, they also identify a number of issues that need to be addressed. Given the nature of these comments, and the fact that the novelty of the conclusions was questioned, I believe these comments will require substantial revisions. Besides the novelty of the conclusions, I would like to ask you also specifically to pay attention to the many references suggested by Anonymous Referee #2, and refer to them when appropriate. Your revised manuscript will be sent back to the referees for consultation.

Please don't hesitate to contact me in the meantime in case of questions.

Best regards

Ryan Teuling

Response: We would like to thank the Editor for his decision. In the revised version of the manuscript, we have improved the introduction section to better highlight the novelties of the study. In particular, our focus is on multi-year drought periods and on the role played by storage and evaporation during these periods in aggravating runoff deficit. By considering a broad array of catchments located in various places in Europe, we have taken into account different climates and different catchment conditions. We think that the questions whether runoff deficit is

1) exacerbated or not during multi-year drought;
2) how much this exacerbation is;
3) and, under which conditions and what are the potential drivers to this exacerbation;

have not been answered yet, especially for Europe.

In the revised version of the manuscript, we have elaborated on several points of novelty of the study and have specifically put our study in relation with what found already in Europe, both in the Introduction and in the section 4 “Concluding remarks and study limitations”. In the latter Section, we have shortened the discussion about the rainfall-runoff models calibration and uncertainty and provided a process-based understanding of our findings. We did so by underlining the potential drivers of exacerbation like vegetation, the differences found with previous studies especially in relation of the multi-year droughts vs droughts in general, and the role played by storage especially in water-limited regions (as also underlined by the recent study of Peterson et al. 2021). In doing this, we have considered references suggested by reviewer 2 and have included almost all of them as relevant.

We think that the manuscript has improved significantly thank to the useful comments of the reviewers.

In the text below reviewers’ comments are reported in black, our answers are in blue normal text, *references are in italic* and the **text from the manuscript is reported in bold blue.**

Peterson, T. J., Saft, M., Peel, M. C., and John, A.: Watersheds may not recover from drought, Science, 372, 745–749, https://doi.org/10.1126/science.abd5085, 2021.

Answers to the comments of Reviewer 1

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.

Response: We thank the reviewer for their critical evaluation of the manuscript which gives us the opportunity to improve the manuscript and to clarify better the points of novelty it provides. In the revised version we have elaborated more on these points in the introduction by focusing on what is missing and what we have added to the literature. See lines 59-90 of the revised manuscript:

"The relevance of ET and storage during droughts has been recognized for a long time (Teuling et al., 2013; Miralles et al., 2019), but the runoff implications of their response to precipitation changes are still poorly understood (Goulden and Bales, 2019). For instance Orth and Destouni (2018) found that runoff decreases account for 65–80% of the precipitation deficits, while ET reductions are small and only notable in dry climates, accounting for 0–20% of the precipitation deficits. In other words, they found that the decline in runoff during droughts is faster and stronger than that in evaporation. While Orth and Destouni (2018) provided pieces of evidence of the importance of soil moisture and evaporation in the propagation of the meteorological drought into hydrological droughts, they did not focus specifically on multi-year drought periods, nor they quantified explicitly the further aggravation of runoff deficit with respect to precipitation deficit. In this regard, previous studies have shown that prolonged meteorological droughts may indeed result in a larger-than-expected decrease in runoff (Saft et al., 2015; Avanzi et al., 2020; Tian et al., 2018; Mastrotheodoros et al., 2020; Tian et al., 2020; Alvarez-Garreton et al., 2021). This observation not only shows that precipitation deficit is an insufficient predictor for fully characterizing droughts, but also proofs that the coevolution of water basins with climate (in the form of ET and ΔS) may play a non-negligible role in driving meteorological to hydrological droughts, as during these periods, plants adapt differently to water stress (McDowell et al., 2008) and might access water even from very deep water pools (Fan et al., 2017; Klos et al., 2018; Goulden and Bales, 2019; Hahm et al., 2019) thus subtracting water to runoff.
The ultimate cause behind this observed exacerbation of runoff deficit is still unknown, with previous works providing contrasting and not conclusive results related either to the buffered response time of evaporation to precipitation deficit (Avanzi et al., 2020) or to streamflow memory (Alvarez-Garreton et al., 2021). It is also unclear whether this exacerbation takes place only in specific climates, such as Mediterranean regions where precipitation distribution is skewed toward winter and summer is normally dry (Feng et al., 2019), or whether exacerbation of runoff deficit during meteorological droughts is a common feature of water basins across climates. In this regard, previous works in non-Mediterranean regions of Europe showed the evaporation amplifies the impact of summer droughts, but these studies mostly focused on storage rather than on runoff exacerbation (Teuling et al., 2013). Large-scale assessments spanning a variety of climates are still needed to gain further understanding of the runoff implications of meteorological droughts. Assessing the validity and the large-scale occurrence of this behavior especially over a variety of basins with different climate and hydrological characteristics has important implications for defining sustainable water management strategies and understanding potential ecological traits and is becoming more and more urgent due to the increasing frequency and magnitude of drought events (Roudier et al., 2016; Sheffield and Wood, 2008; Samaniego et al., 2018; Wehner et al., 2011; Haile et al., 2020).

In this work, we answered the following research questions:
i) do multiyear droughts in Europe correspond to an exacerbation of runoff deficit compared to precipitation deficit?
ii) If so, how severe is this exacerbation and how much is related to a coupled increase in water allocated to evaporation?
ii) Finally, what are the potential drivers of this exacerbation? "

In particular, our focus is on multi-year drought periods (and not droughts in general) and on the role played by the storage and evaporation during these periods in aggravating runoff deficit (lines 51-58 of the revised manuscript)

"The fact that runoff deficit is not only dictated by atmospheric demand but other factors play a role (often not well represented in Earth system models, e.g., (Fowler et al., 2020; De Kauwe et al., 2015)) is especially true when prolonged precipitation deficits occur such as those experienced during multi-year drought periods. While for a single dry year catchment storage may sustain a higher than usual evaporation rate (Orth and Destouni, 2018), this might not happen during multi-year droughts (Peña-Angulo et al., 2021). During multi-year droughts, precipitation deficit and pre-existing catchment conditions (e.g., storage) modulate the response of evaporation to temperature anomaly in ways that are only partially quantified and predictable and that have direct consequences on runoff volumes. These aspects make multi-year droughts essentially different from individual and shorter dry years and motivate specific research on this topic."

Considering a broad array of catchments located in various places in Europe allowed us to consider a variety of climates and catchment conditions. We think that the questions whether runoff deficit is
1) exacerbated or not during multi-year drought;
2) how much this exacerbation is;
3) and, under which conditions and what are the potential drivers to this exacerbation;
have not been answered yet, especially for Europe.

In the revised version of the manuscript, we have also improved section 4 "Concluding remarks and study limitations". In this section we have shortened the discussion about the rainfall-runoff models calibration and uncertainty and provided a process-based understanding of our findings by underlining the potential drivers of exacerbation like vegetation, the differences found with previous studies especially in relation of multi-year drought periods (and not droughts in general) and the role played by storage especially over water-limited regions. In the discussion we have enriched the literature and provided further connections of this study with evidences found by other recent studies which corroborate further our findings. See lines 342-364 of the revised manuscript:

"These results were obtained from an empirical, strictly data-based analysis, but are in line with earlier findings (Saft et al., 2015; Avanzi et al., 2020), as well with those inferred from blending data with mechanistic modelling across the European Alps (Mastrotheodoros et al., 2020). The key role of evaporation was also addressed in Europe by Orth and Destouni (2018) and points to the vegetation as the primary driver (Vicente-Serrano et al. 2014, Peña-Gallardo et al. 2016, Peña-Angulo et al. 2021) caused by enhanced evaporative demand during drought.
Our findings suggest that the increase in actual evaporation can be sustained also for long dry periods (i.e., multi-year droughts) and in typically water-limited environments (while for example Orth and Destouni (2018) highlighted this behavior predominantly in energy limited regions). A potential explanation to this can be given by the capacity of deep-rooted trees to access water from weathered highly porous saprock or rock moisture (Rempe and Dietrich, 2018; Hahm et al., 2019; Carrière et al., 2020) which can go up to 20-30 m beneath the surface (Klos et al., 2018). These mechanisms, which are vital to support the ecosystem during extended drought periods, by bringing large volumes of subsurface water into the atmosphere, might subtract water to runoff potentially determining an aggravation of the hydrological drought (Amin et al., 2020; Carrière et al., 2020; Barbeta and Peñuelas, 2017). Thus, during long and sustained dry periods like those

that have impacted the European continent, not only runoff is reduced faster than evaporation (Orth and Destouni, 2018), but it is also reduced stronger than expected.
The understanding of the propagation of meteorological drought into hydrological drought for long and sustained dry periods is challenging because the overall catchment storage is expected to play a major role in driving runoff deficit. However, understanding the role played by storage is complicated by the difficulty to measure and estimate it (apart from large scale satellite-derived measurements like GRACE (Rodell et al., 2009)) which is seldom addressed at catchment scale (McNamara et al., 2011). This is mainly due to the fact that storage is characterized by marked spatial heterogeneity, which is difficult to measure at the point scale and so extrapolate to the catchment scale (Spence, 2010). We have addressed this by plotting the root depth distribution and TAWC for basins experiencing a significant shift in the precipitation-runoff relationship finding that the latter are characterized by slightly larger values of these two variables. Nonetheless, further evidences are needed to corroborate this finding."

Potential model limitations in a broader context have been added along with implications in the prediction of hydrological droughts (lines 365-384 of the revised manuscript):

"On top of this, the tested hypothesis of the drop in predictive skill in ET during drought found by Avanzi et al. (2020) and the fact that ET seems to be less coupled with soil moisture than models can generally predict (Dong et al., 2020; Qiu et al., 2020) may undermine the comprehension of the response of ET to drought in Earth system models, especially over water-limited environments for long dry periods where, the "shallow" moisture storage, simulated by the latter can be already completely depleted leading to the paradox of null evaporation (Fowler et al., 2020). The suboptimal representation of ET by Earth system models is not rare. For example, many models do not include stomatal response to dry periods, hydrologic regulation of plant rooting depth (De Kauwe et al., 2015; Fan et al., 2017), correct representation of the plant hydraulics (Li et al., 2021; Kennedy et al., 2019) as well as coevolution mechanisms such as vegetation mortality and expansion (Goulden and Bales, 2019). These coevolution mechanisms affect the so-called climate elasticity of evaporation (Avanzi et al., 2020), that is, the capability of ET (and indirectly runoff) to respond and adapt to changes in climate.
The findings of our study highlight the need to gain a better knowledge about the propagation of meteorological to hydrological anomalies across different climatic regions (Lorenzo-Lacruz et al., 2013) (as well as the identification of meteorological drought indices that best reflect streamflow anomalies as suggested by (Tijdeman et al., 2018)). This would provide an important basis for large-scale drought monitoring and early warning information. This is true especially over data scarce regions due to the observed decline of stream gauge observations (Crochemore et al., 2020) where meteorological-based drought indices are normally used to monitor and predict hydrological droughts."

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.**

Response: 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 of the original manuscript). 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 amounts, 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), and this facilitates discussion with related research.

Beside the practical example we have included at lines 191-198 of the original manuscript, which highlights the importance of the problem from a water resource management standpoint, there are significant ecological implications of runoff reduction especially for rivers characterized by low streamflow (and so when the absolute term of the runoff reduction is small). We have added the following text to highlight this better at lines 231-236 of the revised manuscript:

"On the other hand, even if 30-40% of runoff reduction can result in small absolute reduction for drier basins, this reduction can lead to important changes in hydrological connectivity and shifts in streamflow regimes from perennial to intermittent (Fovet et al., 2021) with serious implications for water quality (Armstrong et al., 2012; Addy et al., 2019), river ecosystems and fish survival (Lennox et al., 2019) freshwater biodiversity (Datry et al., 2016) and the quality and diversity of ecosystem services they provide especially over Southern Europe (Vicente-Serrano et al., 2014; García-Ruiz et al., 2011)."

Furthermore, we have rewritten the text describing Figure 6a and b and expressed the change in precipitation, runoff and evaporation terms in absolute terms.

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.
Fovet, O., Belemtougri, A., Boithias, L., Braud, I., Charlier, J.-B., Cottet, M., Daudin, K., Dramais, G., Ducharne, A., Folton, N., Grippa, M., Hector, B., Kuppel, S., Le Coz, J., Legal, L., Martin, P., Moatar, F., Molénat, J., Probst, A., Riotte, J., Vidal, J.-P., Vinatier, F., and Datry, T.: Intermittent rivers and ephemeral streams: Perspectives for critical zone science and research on socio-ecosystems, 8, e1523, https://doi.org/10.1002/wat2.1523, 2021.
Datry, T., Fritz, K., and Leigh, C.: Challenges, developments and perspectives in intermittent river ecology, 61, 1171–1180, https://doi.org/10.1111/fwb.12789, 2016.
Armstrong, A., Stedman, R. C., Bishop, J. A., and Sullivan, P. J.: What's a Stream Without Water? Disproportionality in Headwater Regions Impacting Water Quality, Environmental Management, 50, 849–860, https://doi.org/10.1007/s00267-012-9928-0, 2012.
Datry, T., Larned, S. T., and Tockner, K.: Intermittent Rivers: A Challenge for Freshwater Ecology, BioScience, 64, 229–235, https://doi.org/10.1093/biosci/bit027, 2014.
Datry, T., Vander Vorste, R., Goïtia, E., Moya, N., Campero, M., Rodriguez, F., Zubieta, J., and Oberdorff, T.: Context-dependent resistance of freshwater invertebrate communities to drying, 7, 3201–3211, https://doi.org/10.1002/ece3.2870, 2017.

Addy, K., Gold, A. J., Welsh, M. K., August, P. V., Stolt, M. H., Arango, C. P., and Groffman, P. M.: Connectivity and Nitrate Uptake Potential of Intermittent Streams in the Northeast USA, 7, 225, https://doi.org/10.3389/fevo.2019.00225, 2019.
Lennox, R. J., Crook, D. A., Moyle, P. B., Struthers, D. P., and Cooke, S. J.: Toward a better understanding of freshwater fish responses to an increasingly drought-stricken world, Rev Fish Biol Fisheries, 29, 71–92, https://doi.org/10.1007/s11160-018-09545-9, 2019.
Vicente-Serrano, S. M., Peña-Gallardo, M., Hannaford, J., Murphy, C., Lorenzo-Lacruz, J., Dominguez-Castro, F., López-Moreno, J. I., Beguería, S., Noguera, I., Harrigan, S., and Vidal, J.-P.: Climate, Irrigation, and Land Cover Change Explain Streamflow Trends in Countries Bordering the Northeast Atlantic, Geophys. Res. Lett., 46, 10821–10833, https://doi.org/10.1029/2019GL084084, 2019.
García-Ruiz, J. M., López-Moreno, J. I., Vicente-Serrano, S. M., Lasanta–Martínez, T., and Beguería, S.: Mediterranean water resources in a global change scenario, Earth-Sci. Rev., 105, 121–139, https://doi.org/10.1016/j.earscirev.2011.01.006, 2011.

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.

   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.

Response: We agree with the reviewer that there is already established knowledge that conceptual models do not account for non-stationarity very well. However, studies using such simple conceptual rainfall-runoff models are still very common. We have revised Section 4 by focusing more on the process-based understanding and potential problems related to Earth system models in general (not only rainfall-runoff models) as also pointed out at point 2 above (see lines 365-384 of the revised manuscript):

"On top of this, the tested hypothesis of the drop in predictive skill in ET during drought found by Avanzi et al. (2020) and the fact that ET seems to be less coupled with soil moisture than models can generally predict (Dong et al., 2020; Qiu et al., 2020) may undermine the comprehension of the response of ET to drought in Earth system models, especially over water-limited environments for long dry periods where, the "shallow" moisture storage, simulated by the latter can be already completely depleted leading to the paradox of null evaporation (Fowler et al., 2020). The suboptimal representation of ET by Earth system models is not rare. For example, many models do not include stomatal response to dry periods, hydrologic regulation of plant rooting depth (De Kauwe et al., 2015; Fan et al., 2017), correct representation of the plant hydraulics (Li et al., 2021; Kennedy et al., 2019) as well as coevolution mechanisms such as vegetation mortality and expansion (Goulden and Bales, 2019). These coevolution mechanisms affect the so-called climate elasticity of evaporation (Avanzi et al., 2020), that is, the capability of ET (and indirectly runoff) to respond and adapt to changes in climate.
The findings of our study highlight the need to gain a better knowledge about the propagation of meteorological to hydrological anomalies across different climatic regions (Lorenzo-Lacruz et al., 2013) (as well as the identification of meteorological drought indices that best reflect streamflow anomalies as suggested by (Tijdeman et al., 2018)). This would provide an important basis for large-scale drought monitoring and early warning information. This is true especially over data scarce regions due to the observed decline of stream gauge observations (Crochemore et al., 2020) where meteorological-based drought indices are normally used to monitor and predict hydrological droughts."

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 have improved the discussion of this part by including examples of studies that attempted to do that (see for example Fovet et al. 2020, Hughes et al. 2021). See lines 385-387 of the revised manuscript:

"Given the projected warming climate and aridity, and the role of ET in driving the exacerbation of runoff during droughts, improving understanding of this elasticity appears an urgent task for future work and attempts to do this are still ongoing (Fowler et al., 2020; Hughes et al., 2021)."

Furthermore, we have added some points related to the limitations of the study (lines 403-410 of the revised manuscript):

"The processes underlying the aggravation of the runoff deficit due to increased evaporation for individual catchments may be related to differences in water storage dynamics, flow paths and evaporation due to changes in the infiltration capacity of soils, the duration of infiltration periods, the timing of infiltration periods, the soil moisture regime, amongst other factors. Given the diversity of catchments in our sample, each with its own internal heterogeneity, the mechanisms connecting precipitation deficit to runoff deficit are likely to result from combinations of factors and may vary from site to site, as well as depending on human influences and topography of the catchment. Further work is needed to clarify which hydrological processes are the main contributors to the findings we have presented."

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?

Response: We thank the reviewer for this comment. We added a discussion point on this and highlighted that this can be a limitation of the study (see lines 393-395 of the revised manuscript):

"… in the study we did not consider the potential impact of trends in temperature and the associated long-term coevolutionary mechanisms of the catchments. In this respect there can be potential reflections of coevolution mechanisms on the runoff reduction for which a simple fitting of the precipitation-runoff relationship might be inadequate."

    - 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?

Response: 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 the topic (Saft et al. 2016) used the same criteria as ours for defining a multi-year drought period (we have now used exactly the same procedure of Saft et al. 2016). 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 have added the following at lines 173-176 of the revised manuscript:

“The above procedure resulted in a satisfactory multi-year drought definition (see Figure 2) that was validated with data found in the literature (Parry et al., 2012) with a minimum of three years to a maximum value of eight years for few basins (median duration of four years). Note that we also tested different thresholds (i.e., larger than 3 years), without noticing any significant difference in the results.”

Saft, M., Western, A. W., Zhang, L., Peel, M. C., and Potter, N. J.: The influence of multiyear drought on the annual rainfall-runoff relationship: An Australian perspective, Water Resour. Res., 51, 2444–2463, https://doi.org/10.1002/2014WR015348, 2015.
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?

Response: The precipitation input we used as representative precipitation was the average annual precipitation for each basin. Now the text reads as follows:

“We detected shifts in the precipitation-runoff relationship by fitting a multivariate regression across annual cumulative streamflow (target variable), basin-averaged annual precipitation, and a categorical variable denoting drought and non- drought year”

- 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?

Response: We agree that more clarity is needed when presenting our datasets. The independency that we mentioned was merely a technical detail for highlighting that the results of the fitting are robust (i.e., independency between I and precipitation accumulations). However, this is not required in the approach as demonstrated in the past by Avanzi et al. (2020) and Saft et al. (2015) and we remove the sentence for avoiding confusion. The main reasons why we used two different precipitation datasets here are i) we wanted to rely upon observed-only higher spatial resolution precipitation for the definition of the precipitation-runoff relationship as we think that this reduces water balance issues and ii) we wanted to make sure that the drought periods identified through ERA5 precipitation were consistent with the periods where evaporation anomalies were calculated from the same dataset. Doing otherwise would have led to questionable results due to an inconsistency between the identified drought periods and ET anomalies. Note that, besides the different spatial resolution, the two datasets are very similar providing 30-years long Pearson annual correlation close to 0.9 (see our reply to point 5 to Reviewer 2) and thus providing almost equal (but not exactly the same) multi-year drought definition.

We have clarified the text already present in the original manuscript and added a consideration about the very close similarities between ERA5 and E-OBS. Now it reads (lines 198-205 of the revised manuscript):

“In this study, I in Eq. (1) was defined based on ERA5 precipitation (I=1 during multi-year drought and I=0 for the other years), while the annual precipitation P was calculated based on E-OBS precipitation dataset. We used two different precipitation datasets because we wanted to rely upon observed-only higher spatial resolution precipitation (i.e., E-OBS) for the definition of the precipitation-runoff relationship to reduce water balance issues. On the contrary, for the drought definition we relied on ERA5 because we wanted to have consistent ERA5-based drought definition evaporation anomalies (both coming from the same dataset). On the other hand, both

ERA5 and E-OBS precipitation are characterized by a relatively high accuracy (Massari et al., 2020) and high similarities over Europe (Pearson annual correlation close to 0.9 during the 1980-2016 period not shown) and interchanging them guaranteed very similar results (not shown here).”

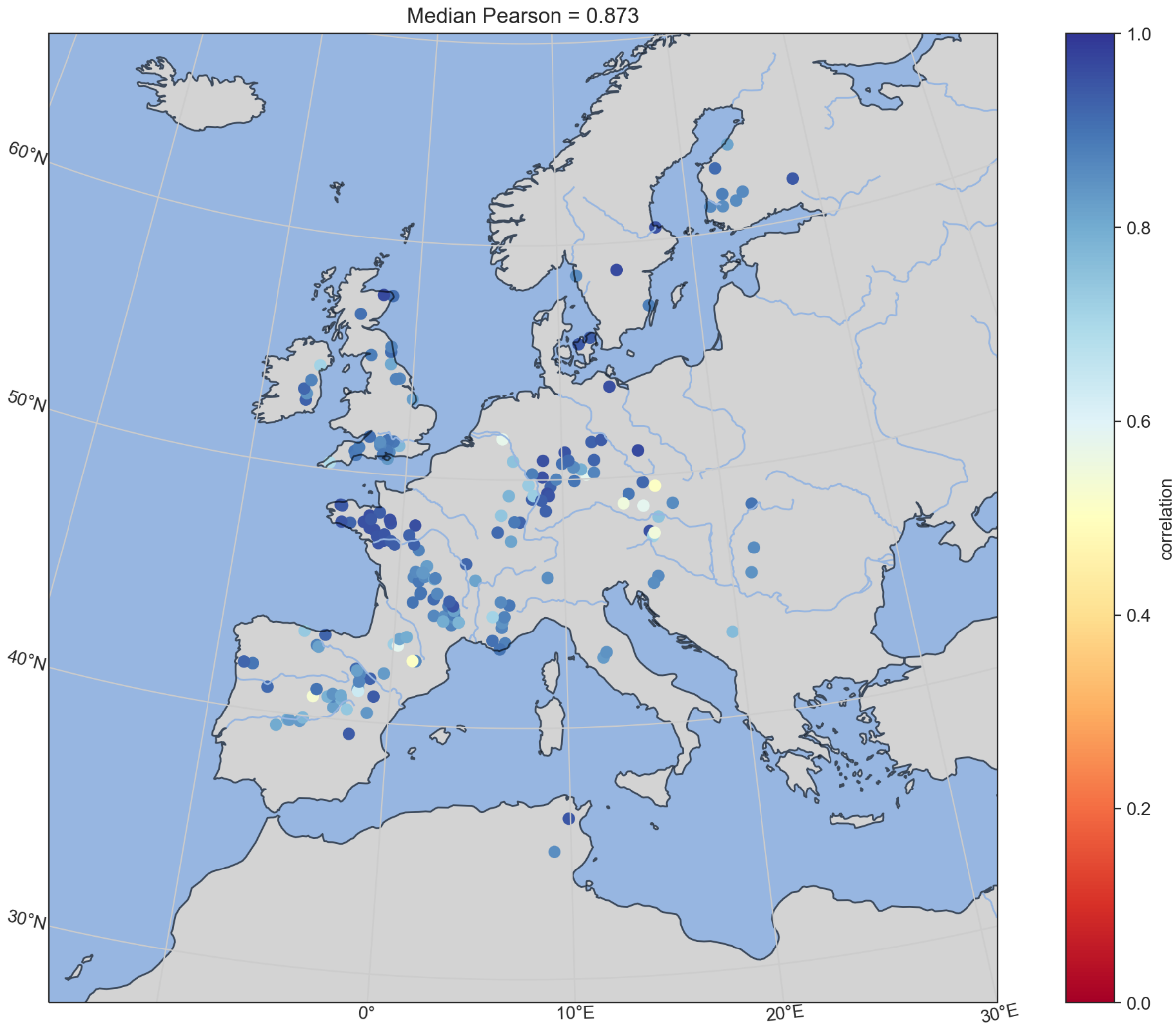


Figure R1: Pearson annual correlation coefficient between ERA5 and E-OBS precipitation (Period 1980-2016). All values of correlation are significant at the 5% level.

- 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.

Response: This was done based on long-term PET, P, and ET, as suggested in the literature to define (qualitatively) 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 have added the following sentence at lines 132-135 of the revised manuscript to clarify this:

“The climate of the specific basin was defined based on the aridity index which we calculated as the ratio between long-term average annual potential evaporation and precipitation (both from ERA5). Note that this is index was used for the mere climate classification while departures from it during multi-year drought periods and its impact on the precipitation-runoff relationship are beyond the scope of the study.”

- 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.

Response: 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.

We have improved the discussion section to better highlight this point. See lines 356-364 of the revised manuscript:

"The understanding of the propagation of meteorological drought into hydrological drought for long and sustained dry periods is challenging because the overall catchment storage is expected to play a major role in driving runoff deficit. However, understanding the role played by storage is complicated by the difficulty to measure and estimate it (apart from large scale satellite-derived measurements like GRACE (Rodell et al., 2009)) which is seldom addressed at catchment scale (McNamara et al., 2011). This is mainly due to the fact that storage is characterized by marked spatial heterogeneity, which is difficult to measure at the point scale and so extrapolate to the catchment scale (Spence, 2010). We have addressed this by plotting the root depth distribution and TAWC for basins experiencing a significant shift in the precipitation-runoff relationship finding that the latter are characterized by slightly larger values of these two variables. Nonetheless, further evidences are needed to corroborate this finding"

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…".

Response: We have carefully revised the manuscript. We hope that now the quality of the writing has been improved significantly. 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 have modified 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.

Response: 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 have removed the sentence to avoid confusion.

Answers to the comments of Reviewer 2

Review of "Evapotranspiration enhancement drives the European water-budget deficit during multi-year droughts". This is an interesting study focused on a very relevant research topic as availability of water resources is essential during dry years and determining the possible role of ET is particularly important during these periods. The study covers catchments in different regions of Europe and related streamflow and precipitation observations. I find two main issues that prevent for a positive assessment of the ms. The first one is related to the null consideration of land cover changes and human management to explain the different changes between meteorological and hydrological droughts, and the second one the consideration of multiannual meteorological droughts as driver of hydrological droughts since the majority of the small catchments analysed should be insensitive to long time scales of meteorological droughts. I think the authors should be carefully consider these two issues in order to provide more robust results on this very relevant topic.

Response: We thank the reviewer for underlining the importance of our work. We have carefully considered the issues raised by the reviewer and have provided below our considerations. We apologize for missing some important references and thank the reviewer for the wealth of references included in this review, which help to put our work in a broader scientific context. We have harmonized the literature and Introduction and discussion sections by including many of these references in the text along with a discussion.

1. Line 37. I would not say runoff response to precipitation changes has been rarely studied. There is a vast scientific literature on this issue considering both modelling and empirical studies. A quick search in Scopus with (TITLE-ABS-KEY (runoff) AND TITLE-ABS-KEY (precipitation AND change)) returns more than 8000 documents so I think authors should reformulate this sentence.

Response: The sentence was too generic indeed. We have rephrased it.

2. Line 39. It should be Van Loon

Response: We have corrected it.

3. Introduction in general. What I miss in this section is a particular focus of the role of land cover and human management/demand. Vegetation characteristics and land cover changes are primary drivers of the enhancement of hydrological droughts during periods of precipitation deficits (see e.g. https://www.nature.com/articles/s41467-018-06013-7). The partition of precipitation between blue and green water may be strongly relevant during periods of water deficits in which water consumption by vegetation would enhance, reducing runoff production. For this reason, vegetation changes may be determinant and much more important than changes in temperature and demand, ultimately affecting ET. This has been observed for example in Mediterranean catchments (https://www.sciencedirect.com/science/article/pii/S2213305421000321#!).

Response: We thank the reviewer for pointing this out. In the study we used actual evaporation (ET) and not potential evaporation. As such, ET incorporates bare soil evaporation, evaporation from interception and transpiration from vegetation (in other words, what the reviewer calls "vegetation changes" should be included already in the ET we considered). Unlike Orth and Destouni (2018), ET was derived here only from ERA5 (thanks to its higher spatial resolution and its higher quality with respect to ERA-Interim, Martens et al. 2020) which ingests a large number of observations and thus should guarantee a better representation of the true atmospheric fluxes. Temperature and demand (i.e., potential evaporation) was only used in Figure 6 to understand the overall climate of the years, in agreement with previous studies. In practice, our assumption is that the change in vegetation activity determined by the increased radiation during drought reflects in enhanced ET values as long as water is available for plants. We agree that this aspect was not clear enough in the manuscript so we will make it clear in the revised version and cite a recently published paper confirming this (Peña-Angulo et al. 2021).

Orth, R., Destouni, G., 2018. Drought reduces blue-water fluxes more strongly than green-water fluxes in Europe. Nat Commun 9, 3602. https://doi.org/10.1038/s41467-018-06013-7

Peña-Angulo, D., Vicente-Serrano, S.M., Domínguez-Castro, F., Noguera, I., Tomas-Burguera, M., López-Moreno, J.I., Lorenzo-Lacruz, J., El Kenawy, A., 2021. Unravelling the role of vegetation on the different trends between climatic and hydrologic drought in headwater catchments of Spain. Anthropocene 100309. Martens, B., Schumacher, D.L.,

Wouters, H., Muñoz-Sabater, J., Verhoest, N.E.C., Miralles, D.G., 2020. Evaluating the surface energy partitioning in ERA5 (preprint). Climate and Earth System Modeling. https://doi.org/10.5194/gmd-2019-315 https://doi.org/10.1016/j.ancene.2021.100309

We have improved the introduction section by following the reviewer's suggestions. The suggested references have been added and discussed in relation of the objectives of the study.

Please, see the new Introduction (section 1) and final remarks section (section 4).

4. The other important factor is human demand, which represents higher percentage related to the total available water during dry years, exacerbating hydrological droughts in highly regulated basins (https://www.sciencedirect.com/science/article/pii/S0341816216301291?via%3Dihub). I agree with authors that enhanced temperature and atmospheric demand increase ET, and this is particularly relevant during dry periods. Thus, certain role has been already identified in Mediterranean areas (https://iopscience.iop.org/article/10.1088/1748-9326/9/4/044001), nevertheless, the strong increase of hydrological droughts observed in large regions of southern Europe cannot be explained exclusively by climate/runoff relationships. The problem is much more complex and climate/land cover changes/water regulation and demand should be considered in this kind of analysis to explain why hydro droughts are exacerbated in relation to precipitation deficits (see https://www.sciencedirect.com/science/article/pii/S0012825211000134).

Response: We thank again the reviewer for raising this important point. We overall agree that land cover changes and human influence may play a major role in the response of any catchment to drought. In particular, land cover change should alter ET fluxes especially in the long-term perspective (>10 years or longer), so any ET datasets should contain this information either directly (by using satellite dataset of dynamic land cover changes) or indirectly via ingestion of observations to overcome the inherent simplified model parameterization of soil and land cover. Our choice of ERA5 reanalysis was driven by these considerations, its long-term availability, and its relatively high quality with respect to its predecessors (Martens et al. 2020), as stated at lines 77-81 of the original submission

"Evapotranspiration from ERA5 reanalysis was used because its relatively high quality (Martens et al., 2020) especially over Europe where a substantially large volume of observations is ingested."

For the problem of water regulation, we used exactly the same dataset of Orth et al. (2018) and Stahl et al. (2010), which shall guarantee that we are using the state of the art in terms of smallest human influence. In addition to that, we performed a very strict preprocessing of the runoff time series. See our point 1 to section 2.2 of the original manuscript:

"From an initial number of more than 3,900 stations, 1,043 stations were retained by excluding (via visual inspection) those with evident dubious patterns due to human regulations (such as constant flows), inhomogeneity, problems in low flow range, missing values for a long period of time (> 2 year) (as suggested in Kundzewicz and Robson, 2004), or an observation period below 20 years. Although care was taken in identifying these issues, some human-induced alterations are likely to be still present in these time-series. Nevertheless, a certain degree of disturbance can be tolerated (Murphy et al., 2013), considering also the annual granularity of our analyses."

Stahl, K., Hisdal, H., Hannaford, J., Tallaksen, L.M., van Lanen, H. a. J., Sauquet, E., Demuth, S., Fendekova, M., Jódar, J., 2010. Streamflow trends in Europe: evidence from a dataset of near-natural catchments. Hydrology and Earth System Sciences 14, 2367–2382. https://doi.org/10.5194/hess-14-2367-2010

Orth, R., Destouni, G., 2018. Drought reduces blue-water fluxes more strongly than green-water fluxes in Europe. Nat Commun 9, 3602. https://doi.org/10.1038/s41467-018-06013-7.

We have included the suggested references in the context of our research and have added a discussion of potential limitations of our analysis at lines 396-400 of the revised manuscript:

"Despite we have carried out a high controlled experiment employing full natural runoff measurements by screening out basins potentially characterized by human regulations and used the some of the best available precipitation and evaporation products, we cannot exclude that the observed runoff deficit exacerbation might have been driven by other factors related to the climate/land cover changes/water regulation interactions than the simple increase of actual evaporation (Vicente-Serrano et al., 2019; Teuling et al., 2019)."

and lines 403-410:

"The processes underlying the aggravation of the runoff deficit due to increased evaporation for individual catchments may be related to differences in water storage dynamics, flow paths and evaporation due to changes in the infiltration capacity of soils, the duration of infiltration periods, the timing of infiltration periods, the soil moisture regime, amongst other factors. Given the diversity of catchments in our sample, each with its own internal heterogeneity, the mechanisms connecting precipitation deficit to runoff deficit are likely to result from combinations of factors and may vary from site to site, as well as depending on human influences and topography of the catchment. Further work is needed to clarify which hydrological processes are the main contributors to the findings we have presented."

5. Line 70-75: See limitations of e-obs in some regions of Europe in which this dataset reinforces artificially negative precipitation trends (https://iopscience.iop.org/article/10.1088/1748-9326/ab9c4f). Note that this data does not include homogeneity testing and it may have limitations in areas in which low percentage of meteo stations is used (e.g., southern Europe).

Response: We agree with the reviewer that the E-OBS dataset is not perfect. In particular, we found several issues in Southern Italy and Northern Africa (where catchments of the datasets were excluded from our analysis). However, the overall agreement between E-OBS and ERA5 is very good elsewhere (see Figure R1), with median Pearson correlation coefficient between annual precipitation values close to 0.9 (n=210).

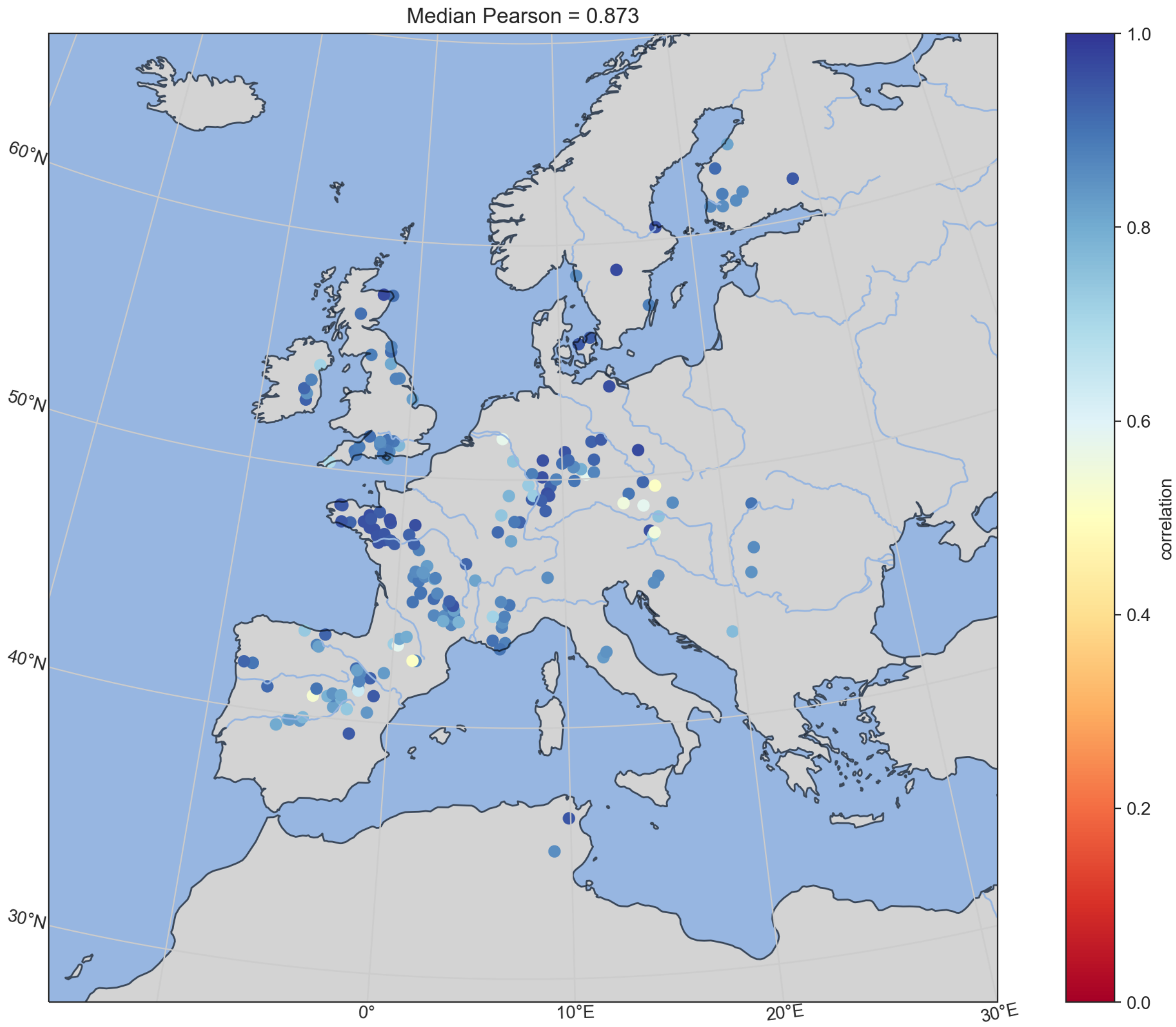


Figure R1: Pearson annual correlation coefficient between ERA5 and E-OBS precipitation (Period 1980-2016). All values of correlation are significant at the 5% level.

We have added some comments at lines 203-205 of the revised manuscript:

"On the other hand, both ERA5 and E-OBS precipitation are characterized by a relatively high accuracy (Massari et al., 2020) and high similarities over Europe (Pearson annual correlation close to 0.9, not shown) and interchanging them guaranteed very similar results (not shown here)."

6. Lines 77-83. Note the problems of the potential evapotranspiration product in ERA5 (https://confluence.ecmwf.int/pages/viewpage.action?pageId=171414970). I would recommend authors to calculate the atmospheric evaporative demand using the FAO-56 equation from ERA5 inputs instead of using the product in ERA5. It solves the existing problem.

Response: We thank the reviewer for pointing this out. We indeed found this problem too (zero PET in central France) during our analysis, but only in the potential evapotranspiration variable data over a small area compared to the whole Europe as indicated in the link. So, our main results were not affected.

On the other hand, this problem undermined the calculation of the aridity index, so we replaced those incorrect ERA5 PET with PET estimated through the simple Hamon potential evapotranspiration approach (which uses only temperature data and a calibration coefficient, see Lu et al. 2005). In doing that, the calibration of the coefficient of the Hamon potential evapotranspiration for the problematic area in France was performed with nearby data. We also did a cross validation by using basins where ERA5 data were correct and found differences below 5% with our method. Besides, we want to underline that this problem was limited to only 3 basins over the total of 210 basins so it does not impact the conclusions of our study.

- Lu, J., Sun, G., McNulty, S.G., Amatya, D.M., 2005. A Comparison of Six Potential Evapotranspiration Methods for Regional Use in the Southeastern United States1. JAWRA Journal of the American Water Resources Association 41, 621–633. https://doi.org/10.1111/j.1752-1688.2005.tb03759.x

7. Lines 85-87. If the purpose is to characterize drought events I would encourage authors to use more detailed temporal resolution. Annual data could mask dry conditions at the seasonal/monthly scales that could be strongly relevant to explain hydrological drought response to water deficits. In the introduction it is stressed the complexity of runoff generation processes and to restrict the analysis to the annual scale could limit the identification of snow effects, high precipitation events, etc., which are determinant to evaluate the factors of occurrence of hydro droughts.

Response: In principle we agree with the reviewer that intra-annual variations are important. However, our analysis focuses on the annual water balance, thus intra-annual variations were not taken into account. An annual time scale is necessary to correctly compute the water balance of catchments were precipitation and demand are highly seasonal (for example, Med catchments). In these situations, a monthly resolution would misrepresent the $Q = P - ET + \Delta S$ equation. This annual perspective was also used by previous literature (see Saft et al. 2016 and Avanzi et al. 2020, Peterson et al. 2021 for example), which enables us to directly compare our results with previous works. We have planned to go into more details in a future work, but for the sake of the study annual variations are sufficient to explain the variability of the meteorological forcing and properly solve the water balance.

We have added a comment on this at lines 401-402 of the revised manuscript:

"Our analysis was based on annual time scale. However, intra-annual variations of the water balance components could exert an important role to explain hydrological drought response to precipitation deficits."

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

Avanzi, F., Rungee, J., Maurer, T., Bales, R., Ma, Q., Glaser, S., Conklin, M., 2020. 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/10.5194/hess-24-4317-2020

Peterson, T.J., Saft, M., Peel, M.C., John, A., 2021. Watersheds may not recover from drought. Science 372, 745–749. https://doi.org/10.1126/science.abd5085

8. Section 2.2. I see the authors have restricted the analysis to catchments below 50.000 km2 in Europe. Nevertheless, this does not reduce the role of possible human disturbances (e.g. highly regulated catchments in the headwaters).

Response: Please, see our reply to point 4. Figure 1b shows that catchments of the size close to 50000 km2 are a very small part of the entire dataset.

9. It seems that series with data gaps are used. I think this is strongly problematic to assess dry severity as high flows in one day that correspond to data gaps may entirely alter annual flow, particularly in small catchments of water limited regions. If the analysis is applied at the annual scale, why not to use directly monthly streamflow and then try to fill the gaps? Filling monthly data is less problematic than filling daily streamflow information**.**

Response: As mentioned above in our response to comment n.4, we relied on the GRDC and EWA runoff datasets as in Orth et al. (2018) and Stahl et al. (2010) which guarantee the least human impact on the dataset. On the one hand, these datasets are available at daily resolution and any monthly runoff is obtained from daily values, so we did not see any differences by starting from monthly values as an interpolation should have been carried out also to obtain the latter. On the other hand, we performed a very strict screening process by removing all the years and basins where missing data were larger than 15 days (sorry there was a typo in the manuscript stating 315 days at line 106, whereas we retained all years with at least 350 days of discharge data). Please, see our points 3 and 4 of the section 2.2 of the original manuscript (lines 105-106).

We added some text about potential limitation at lines 396-400 of the revised manuscript:

"Despite we have carried out a high controlled experiment employing full natural runoff measurements by screening out basins potentially characterized by human regulations and used the some of the best available precipitation and evaporation products, we cannot exclude that the observed runoff deficit exacerbation might have been driven by other factors related to the climate/land cover changes/water regulation interactions than the simple increase of actual evaporation (Vicente-Serrano et al., 2019; Teuling et al., 2019)."

9. Lines 124. Calculation of SPI is not correct. It must be calculated from monthly series and data must not be smoothed previously. Smoother series are obtained selecting long SPI time scales. There are WMO guidelines to calculate SPI. Drought definition is very arbitrary in the methodology and it strongly may affect the obtained results. Streamflow drought usually responds to varied time scales of meteorological drought according river regimes, catchment characteristics, climate, human disturbances, etc. Sometimes hydrological droughts respond to short time scales of meteorological droughts but in other cases they respond to longer time scales. There is a vast scientific literature on this issue that should be considered by authors (https://hess.copernicus.org/articles/20/2483/2016/hess-20-2483-2016.html, https://www.sciencedirect.com/science/article/pii/S0022169418308813, https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2017WR022412). Ideally, the authors should consider to determine what is the time of response of hydro droughts to meteorological droughts and then to select the most suitable time scale for each basin. In any case, three years in meteo drought is a too long period to explain the occurrence of drought in small catchments, which usually respond to shorter time scales (http://www.int-res.com/abstracts/cr/v58/n2/p117-131/).

Response: We considered annual cumulative precipitation in the SPI calculation because we were specifically interested in detecting precipitation anomalies within the hydrological year. This approach provided a more straightforward comparison with evaporation anomalies. We did not apply any smoothing to the original SPI time series, rather we smoothed data only for the identification of the length of the multi-year drought period as also performed by Saft et al. (2016). Considering either SPI or strict precipitation anomalies led to the same multi-year drought definition and exactly the same results. In the revised version of the manuscript, we have relied upon precipitation anomalies normalized by the standard deviation to avoid further misunderstanding on the terminology. Furthermore, our validation of the main drought events that hit Europe in the last 30 years confirm the reliability of our identified drought periods. Please see the revised section 2.4.

10. Line 185. The figure is not informative. In the majority of catchments analyzed, different meteorological droughts have been recorded so the year of drought onset is not informative. Does it refer to the onset of the first drought identified? I do not think this information is useful to assess the characteristics of meteorological droughts in the region.

Response: We have modified the figure and the caption and we have plotted only the onset of the most severe multi-year droughts identified. See the updated figure below:

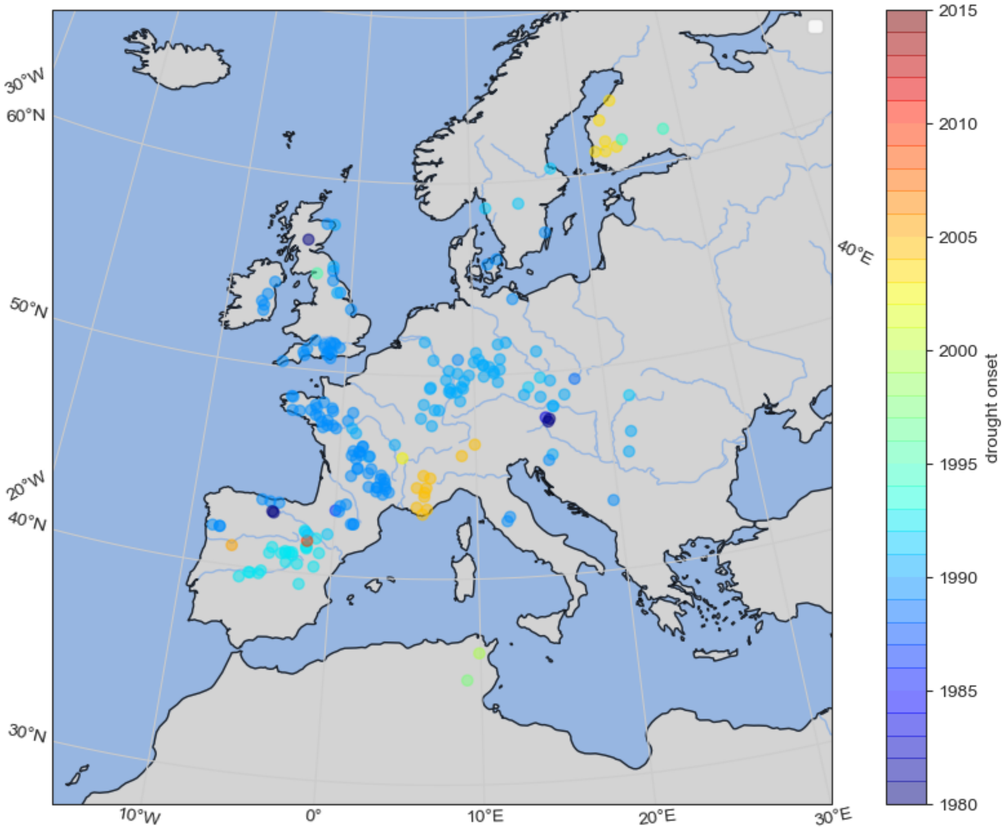


Figure 2. Onset of the most severe multi-year drought period observed in Europe from 1980 to 2015. The most severe multi-year drought period was selected based on the largest negative precipitation anomalies (averaged on the period) we have observed among the multi-year drought periods (with length ≥ 3 years) we have identified for each basin.

11. Lines 186-208: I think these results can be strongly affected by the proposed method (e.g. the selection of meteorological droughts based exclusively on multiannual information).

Response: Please, see our reply to point 9.

12. Lines 219-241: The spatial distribution of the mismatch between meteorological and hydrological droughts may be essentially explained by the role of vegetation changes and human water demand.

Response: Please, see our replies to points 3 and 4, the new introduction and discussion sections were almost all of the suggested references have been included and discussed in relation to the results of our study. In particular, the potential role of the vegetation has been highlighted at lines 38-39 of the revised manuscript:

"Depending on the direction of precipitation change, evaporation-precipitation feedback mechanisms may comprise vegetation expansion and/or mortality (Senf et al., 2020; Choat et al., 2018; Peña-Angulo et al., 2021)…"

lines 342-346:

"These results were obtained from an empirical, strictly data-based analysis, but are in line with earlier findings (Saft et al., 2015; Avanzi et al., 2020), as well with those inferred from blending data with mechanistic modelling across the European Alps (Mastrotheodoros et al., 2020). The key role of evaporation was also addressed in Europe by Orth and Destouni (2018) and points to the vegetation as the primary driver (Vicente-Serrano et al. 2014, Peña-Gallardo et al. 2016, Peña-Angulo et al. 2021) caused by enhanced evaporative demand during drought."

lines 348-355:

"A potential explanation to this can be given by the capacity of deep-rooted trees to access water from weathered highly porous saprock or rock moisture (Rempe and Dietrich, 2018; Hahm et al., 2019; Carrière et al., 2020) which can go up to 20-30 m beneath the surface (Klos et al., 2018). These mechanisms, which are vital to support the ecosystem during extended drought periods, by bringing large volumes of subsurface water into the atmosphere, might subtract water to runoff potentially determining an aggravation of the hydrological drought (Amin et al., 2020; Carrière et al., 2020; Barbeta and Peñuelas, 2017). Thus, during long and sustained dry periods like those that have impacted the European continent, not only runoff is reduced faster than evaporation (Orth and Destouni, 2018), but it is also reduced stronger than expected."

lines 369-372:

"The suboptimal representation of ET by Earth system models is not rare. For example, many models do not include stomatal response to dry periods, hydrologic regulation of plant rooting depth (De Kauwe et al., 2015; Fan et al., 2017), correct representation of the plant hydraulics (Li et al., 2021; Kennedy et al., 2019) as well as coevolution mechanisms such as vegetation mortality and expansion (Goulden and Bales, 2019)."

13. Line 219. This method has not been previously explained. It should be detailed in the methods section.

Response: we thank the reviewer for highlighting this point. This is actually not a new method, sorry for the confusion, we have reformulated the sentence.

14. Line 214-226. I mostly agree with these results, but the role of vegetation changes and human demand are determinant to explain the suggested temporal pattern. See e.g. https://hess.copernicus.org/articles/23/3631/2019/hess-23-3631-2019.html, https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2019GL084084).

Response: We agree with the reviewer, we missed the citations of these important pieces of work. We have included them in the revised manuscript by adding also a discussion. See also our reply to point 12.

15. 233-241. I think that the annual focus is masking stronger influence of ET on runoff as authors are mixing the behaviour of the humid and dry season, in which the magnitude of ET is very different. I wonder why not to focus on the dry season independently in which most of the ET is recorded to identify this kind of impacts. Probably the role of ET would clearly reinforce in comparison to the annual resolution but mostly in energy-limited regions. In dry regions in summer it is expected small role of ET given the low water availability.

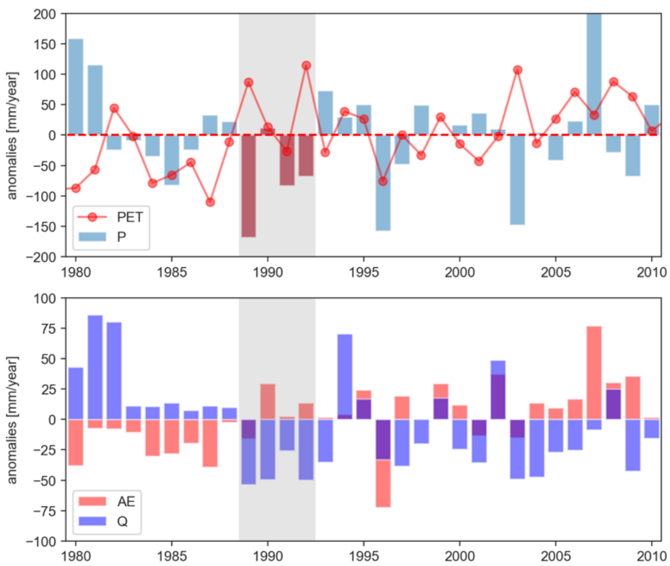
Response: The reviewer raises an important and interesting point. Notwithstanding the fact that we think that it is important to consider the annual time scale because of the paramount role of the catchment storage in explaining the annual runoff response of the basin and thus not only ET (see response to remark 7), the definition of dry and wet seasons in a long-term perspective might be challenging because of climate change (Feng et al. 2019).

For instance, Mediterranean climates are experiencing a shift in the synchronicity in the precipitation and PET that arise from their relative magnitudes and timing and result in the modification of the seasonal interaction of precipitation and PET in time, determining geographically expansion or retreat of typically Mediterranean climates (Feng et al. 2019).

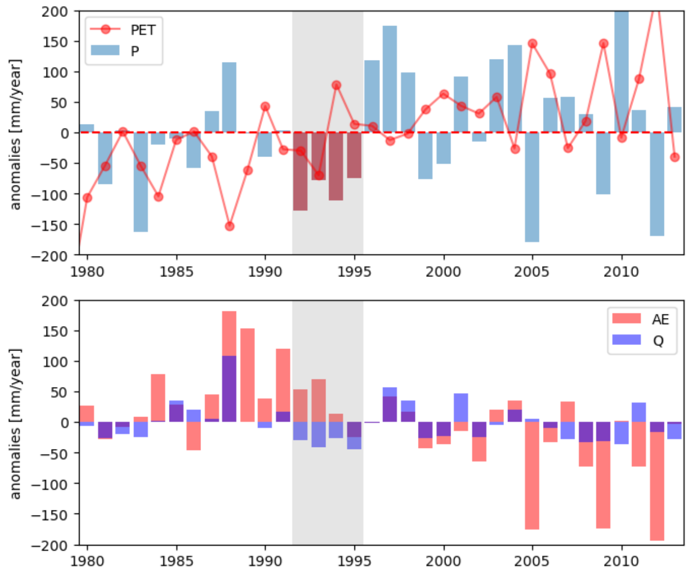
Feng, X., Thompson, S.E., Woods, R., Porporato, A., 2019. Quantifying Asynchronicity of Precipitation and Potential Evapotranspiration in Mediterranean Climates. Geophys. Res. Lett. 46, 14692–14701. https://doi.org/10.1029/2019GL085653

16. 253-255. These are quite strange results. A reduction of precipitation by 95% should not increase ET, I would expect a reduction in ET as consequence of the very low water availability. In any case, a 95% of reduction seems to be too large to be reasonable.

Response: We apologize for the confusion, there was an error in the scale bar. We will report values expressed in mm/year for both plots, which are more physically understandable (see Figure R2).



(a) - North Europe



(b) South Europe

Figure R2: new Figure 6.

17. 263-265: It seems that small catchments selected for the Iberian Peninsula are located in mountain areas, and they are not water limited, by mostly energy limited so conclusions on this issue should be carefully reinterpreted. Probably the coarse spatial resolution of the data 36 km2 shows limitations to assess this kind of issues in complex mountainous areas of Spain. In any case, water demands by agriculture and vegetation changes are main driver in this area and I think that authors should consider these factors in some way in their analysis in order to obtain a reliable explanation of the observed changes. I wonder if in North Europe (e.g. UK water demands by urban supply during dry years) it may be also a substantial influence of external factors to explain different trends between precipitation and runoff during dry years**.**

Response: The spatial resolution of the used dataset can be of course an issue and we will highlight this limitation in the revised version of the manuscript (lines 411-413).

"Despite we selected some of the best possible climate datasets of evaporation and precipitation some uncertainty is unavoidable which is also related to the their relatively coarse spatial resolution with respect to the size of the analysed basins."

We also double checked the position of the basin and it is located in forested region (see Figure R3) on a hilly terrain (elevations below 1000 a.s.l). Although we cannot exclude effects of micro climate, our datasets suggests that this catchment is representative for water limited regime (unfortunately we do not have local scale data to verify this out).

Regarding the human impact the considerations done above holds, specifically the selected basin does not seem highly anthropized (again see Figure R3).

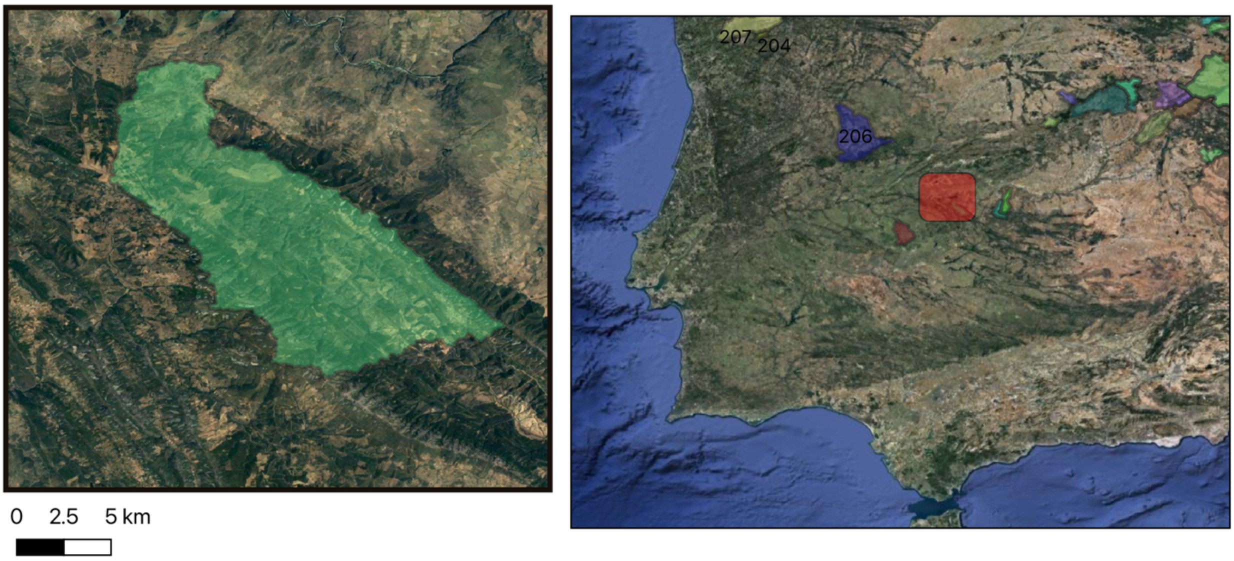


Figure R3: Location of the catchment in southern Europe.

18. I wonder if in North Europe (e.g., UK water demands by urban supply during dry years) it may be also a substantial influence of external factors to explain different trends between precipitation and runoff during dry years.

Response: Water demand is surely an issue, but with over relatively small populated areas like the case of catchments considered here (see our replies above), we doubt that water urban use can have a such severe impact on river discharge (assuming water for urban use is abstracted totally from the river, which is hardly the case). For instance, considering a typical provision of 250-300 l/inhabitant per day for aqueducts planning, reaching 10% of the average discharge the catchment (0.6 m^3/s) implies at least a population of about 200’000 people consuming at maximum rate night and day. We also plotted the basin located in the North of Europe (LON 6.8O, LAT 53.6N, see Figure R5) and it does not seem such a highly populated area to explain significant urban water consumptions.

We have included anyway in the discussion potential problematic related to human impact in the discussion section. See lines 403-410 of the revised manuscript:

“The processes underlying the aggravation of the runoff deficit due to increased evaporation for individual catchments may be related to differences in water storage dynamics, flow paths, and evaporation due to changes in the infiltration capacity of soils, the duration of infiltration periods, the timing of infiltration periods, the soil moisture regime, and the human water use amongst other factors. Given the diversity of catchments in our dataset, each with its own internal heterogeneity, the mechanisms connecting precipitation deficit to runoff deficit are likely to result from combinations of factors and may vary from site to site, as well as depending on human influences and topography of the catchment. Further work is needed to clarify which hydrological processes are the main contributors to the findings we have presented.”

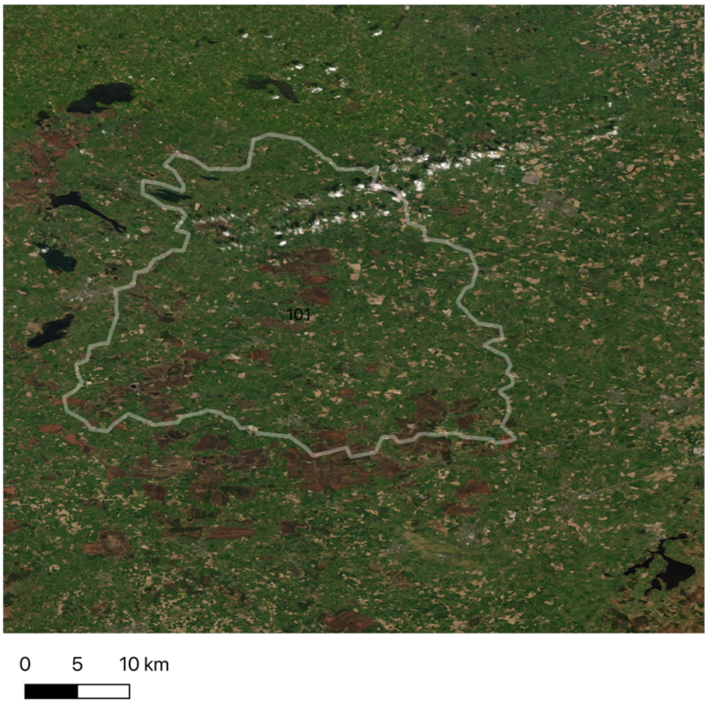


Figure R4: catchment located in Northern Europe.