Dear authors,

Your revised manuscript has now been evaluated by two referees who also previously acted as referees during the online discussion phase. While one of the referees is happy with the revisions and has no further comments, the other referee still identifies a number of issues that have not been addresses satisfactory. Because I consider this referee a leading expert in the field, I believe these comments carry some weight and should be addressed before the manuscript can be considered for final publication. Given the nature of the comments and the fact that likely some additional analysis or figures are necessary, I classify these in between minor and major. This means that the work can in principle be published, provided that the remaining comments are addressed. I am looking forward to receiving a revised version of your work.

Best regards

Ryan Teuling

We thank the Editor for this decision. We have seriously taken into account the comments of the reviewer and thanks to them the manuscript has been significantly improved. In particular:

1. **we overall agree on the potential influences of land cover changes and human regulation on our analysis. We have made our best effort to demonstrate that in our analyses are not significantly impacted by these influences (see replies to comment 1). However, we agree that some limitations remain, and we highlighted them carefully at the end of the manuscript.**
2. **We have clarified that despite single drought years are important to consider, our analysis focused on multi-year droughts so some of the reviewer's, comments, although helpful and meaningful on a broader perspective, are less significant here.**
3. **We have demonstrated with new figures (classical SPI12 vs precipitation anomalies used here, Figure R2 and R3) and new references and validation, the validity of the multi-year drought period identification. We have also improved Figure 2 (see comments 9, 10 and 11).**
4. **We have provided new evidences and literature references about the quality of our analysis related to the selection of the working datasets (see comments 7, 8, 11 and 15).**
5. **We have followed the suggestion of using vegetation information although this does not bring to clear evidences of improvement of our analysis (Figure R4). On the other hand, we have used rooting depth information (Figure 7b) as additional proofs of our conclusions. The role of vegetation in water limited regions was also better discussed (comments 12, 13, 14, 16 and 17).**
6. **We have highlighted the challenge in conducting intra-annual/seasonal analysis and discussed the validity of our yearly-based analysis (see comments 5 and 12).**

We believe that the manuscript has been further improved. Response to the reviewer are provided in bold gray, text changes in blue **and** references in black.

Response to the reviewer

1. Second review of Review of "Evapotranspiration enhancement drives the European water-budget deficit during multiyear droughts". I my first review of this manuscript I stressed two main issues that could affect the interpretation of the results: i) the null consideration of land cover changes and human management to explain the different changes between meteorological and hydrological droughts, and ii) 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 suggested modifications to consider these issues that were not considered by the authors. They have simply included them as ideas of further research in the last section (lines 390-410). Under my opinion they are issues that affect the interpretation of the obtained results and that should be fully considered to explain the relationship between meteorological and hydrological droughts since on the contrary, the interpretation of the results obtained may not be the correct. I repeat again that the manuscript's topic is interesting and with potential to be published in HESS, but under my opinion the authors are avoiding very relevant aspects that should be considered in a study of this characteristic.

   We thank the reviewer for their critical comments. We also concur with the reviewer that these are very important aspects and, as such, we believe that we have not have overlooked them in our previous revision. Please see additional details below.

   As for the possible impact of human management, we have already pointed out in the revised manuscript that we have considered near natural catchments as in Stahl et al. (2010) and Orth and Destouni (2018). This together with our additionally strict screening of the dataset (visual inspection of dubious patterns and rejection of catchments larger than 50k km2) makes us confident regarding the comparatively small impact of human activities in these catchments. While this was already specified in the previous revision, we have now further detailed this in the newly revised manuscript (see revised section 2.3).

   As for the impact of land cover changes, we agree that they can play a role in the long-term perspective (i.e., by impacting trends for instance). However, here we did not focus on trend calculation, rather, we seek the (significant) differences in streamflow yield over shorter periods (3-5 years depending on the catchment) with respect to the rest of the dataset (30 years). So, despite an influence of the land cover cannot be excluded, these changes cannot alone explain the observed shifts. It should also be noted that we deliberately chose ERA5 because of its relatively large number of observations ingested over Europe, which we assumed to be able to track at least a fraction of the possible influence of land cover changes (note that a recent attempt to include more explicit land cover changes obtained by satellite observations showed neutral impact over Europe,

see Nogueira et al. 2021). Regardless of land cover changes, it is important to realize that these changes do not affect the validity of our results (i.e., the shifting of the precipitation-runoff relationship) and their interpretation. While these pieces of information were already included in the revised manuscript, we have stressed even more this aspect in the revised version of the manuscript (see lines 122-132).

"From ERA5 we also extracted actual and potential evaporation. ERA5 is the latest climate reanalysis produced by the European Centre for Medium-Range Weather Forecasts (ECMWF), providing hourly data on many atmospheric, land-surface and sea-state parameters together with estimates of uncertainty. Actual evaporation 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. The latter is expected to provide an improvement of the accuracy in the simulation of latent heat fluxes determined by vegetation and land cover change. On the other hand, attempts to implement more explicit land cover and vegetation changes in the ECMWF model's land cover characterization (by leveraging on state-of-the-art earth observations) showed mostly neutral impacts on the simulated surface latent and sensible heat fluxes when compared against 51 FLUXNET stations over 1996–2014 over Europe (Nogueira et al., 2021)."

Despite all of the above, we agree that some limitations remain. We have stressed them in Section 4 (see lines 424-429).

"Despite we have carried out a high controlled experiment by employing near natural catchments by screening out basins potentially characterized by human regulations and used 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) which might also not well be represented in the used actual evaporation dataset."

Despite these limitations we think this study brings new evidence on the relationships between precipitation and runoff at the European scale. It also warns about potential runoff aggravation due to evaporation that was missing before in the literature.

In any case, we appreciated the detailed comments to our work which helped to better shape the discussion and make the study much more robust.

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.

Nogueira, M., Boussetta, S., Balsamo, G., Albergel, C., Trigo, I. F., Johannsen, F., Miralles, D. G., and Dutra, E.: Upgrading Land-Cover and Vegetation Seasonality in the ECMWF Coupled System:

Verification With FLUXNET Sites, METEOSAT Satellite Land Surface Temperatures, and ERA5 Atmospheric Reanalysis, 126, e2020JD034163, https://doi.org/10.1029/2020JD034163, 2021.

2. Line 38. The study by Peña-Angulo et al. 2021 is not supporting this statement. Thus, this study demonstrates the opposite: the role of vegetation changes on the partition of precipitation between runoff and evaporation.

   Thanks, we have removed it from the context.

3. Line 39-40. Are really evaporation-precipitation feedbacks driving vegetation composition? What are the evidences on this issue? I wonder if it is not the opposite: vegetation changes in structure and composition drive land atmosphere feedbacks…

   As the reviewer correctly noted, these are complex feedback processes influencing each other, to the extent that it would be very hard to isolate each driver. However, there is a number of studies showing evidence that these feedback mechanisms between precipitation and evaporation do affect vegetation (we cited a few examples as Senf et al., 2020; Choat et al., 2018). Of course, as mentioned, these are complex processes and opposite behaviors might also be observed (e.g., Gouveia et al., 2017; Biederman et al., 2014). We have included two new references on this matter in the text (lines 37-41):

4. Line 50. What are these factors? Maybe vegetation physiology, land cover changes, etc?

   We agree that this passage was a bit vague and we have added details to the sentence (see lines 50-53).

5. Line 54. Again, I do not think this is the best reference to support this issue. In any case, this distinction between annual and multi-annual droughts is also debatable as the severity of drought may be strongly relevant. E.g. the annual drought that affected Europe in 2018 strongly altered ET and this event has not characterized by multiannual characteristics. See special issue on this event at (https://royalsocietypublishing.org/toc/rstb/2020/375/1810). Thus, the statement on multi-annual drought could be also supported by annual droughts.

   We agree with the reviewer that single annual droughts are also important. In this regard, there is no reference in the manuscript that they are not able to alter ET. On the other hand, we focused here on multi-annual droughts to nest our study into a growing body of literature on this matter (Saft et al. 2015, Saft et al. 2016, Avanzi et al. 2020). We have clarified again this at lines 58-63:

   "Note that we are not excluding here that runoff deficit exacerbation can also manifest during single severe drought years as the one impacted Europe in 2018 (see the article collection: https://royalsocietypublishing.org/toc/rstb/2020/375/1810). However, here we will focus on multiyear drought periods as it is our interest to understand mechanisms leading to hydrological drought during long and sustained precipitation deficit periods, as already done by Saft et al. 2015, Saft et al. 2016, Avanzi et al. 2020."

This was also listed as one main limitation of the study (lines 430-435):

"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. Despite this, the study might lack details in resolution (for instance it can reveals periods where the runoff deficit is larger and the impact of the climate of the seasons on the runoff aggravation) but it still valid in terms of water balance perspective."

6. Line 63. I would say that plant coverage and vegetation changes are strongly relevant to explain changes on this issue. See e.g. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021GL094672 and https://www.nature.com/articles/s41586-020-1941-5

   We did not find at line 63 the context of these citations. However, we have included the suggested references in the text at lines 426-429:

   "…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; Teuling and Hoek van Dijke, 2020)."

7. Lines 109-124. Note that E-OBS data does not include homogenisation assessment and there are important differences in the stations ingested in the gridded data, with few densities in southern Europe. You also note that potential evapotranspiration product in ERA5 is affected by important uncertainties (https://confluence.ecmwf.int/pages/viewpage.action?pageId=171414970). It seems that ERA-Land works much better (https://www.nature.com/articles/s41597-021-01003-9). Martens et al. 2020 is not included in the references. I think this is the reference: https://gmd.copernicus.org/articles/13/4159/2020/.

   We agree with the reviewer and we are aware of this. We think we have already clarified these points in the revised version of the manuscript (see lines 223-225) and in the previous review. In particular:

   - **E-OBS is not the best products but it is one of the best in Europe, especially because it is derived from stations and is expected to be characterized by a small bias (we have also shown that -- at annual scale -- E-OBS dataset and ERA5 Precipitation are quite consistent with each other).**
   - **Regarding evaporation, as stressed also in the revised version of the manuscript we used ERA5 potential evaporation (and corrected for the problem in France which in our case impacted only two basins), and we used it for the sole purpose of the calculation of the aridity index. Actual evaporation from ERA5 did not have this issue, so this issue does not impact the results as the reviewer suggests. We**

also think that <u>ERA-Land actual evaporation</u> is not recommended as it does not ingest any type of observation (it uses only the ERA5 forcing) and might not well represent the important impact of the land cover changes the reviewer has mentioned earlier.

- **Thank you for the relevant suggestion of the reference to Martens et al. 2020. It was indeed included in the manuscript but was missing from the reference list:**

  Martens, B., Schumacher, D. L., Wouters, H., Muñoz-Sabater, J., Verhoest, N. E. C., and Miralles, D. G.: Evaluating the surface energy partitioning in ERA5, preprint, Climate and Earth System Modeling, https://doi.org/10.5194/gmd-2019-315, 2020.

8. A fundamental question to respond here is if ERA5 includes land cover/vegetation changes from observations to account for the possible relevant role of this issue in the relationship between precipitation and runoff during drought episodes. I am curious why GLEAM was not used for this purpose since it is a much more robust model for ET?

   This is a good point, but we think it is beyond the scope of our analysis and would deviate from the main message of the paper, as it is not our objective to validate evapotranspiration products. This could be the focus of another work. Still, we believe ERA5 is a really good reanalysis product and works relatively well over its predecessor (i.e., ERA-Interim, see again Martens et al. 2020), which already showed relatively good performance in comparison to other evaporation products and in closing the water balance for 837 catchments worldwide (Miralles et al. 2016). We also noted that GLEAM has shown to be overcoupled with soil moisture (Qiu et al. 2020) and this tendency could provide suboptimal results during very dry periods (Fowler et al., 2020).

   We have added some more details on these aspects (see lines 122-132):

   "From ERA5 we also extracted actual and potential evaporation. ERA5 is the latest climate reanalysis produced by the European Centre for Medium-Range Weather Forecasts (ECMWF), providing hourly data on many atmospheric, land-surface and sea-state parameters together with estimates of uncertainty. Actual evaporation from ERA5 reanalysis was used because its relatively high quality (Martens et al., 2020) over its predecessor ERA-Interim which already showed relatively good performance in comparison to other evaporation products and in closing the water balance of many catchments worldwide (Miralles et al., 2016). In particular, the good performance is expected over especially over Europe where a substantially large volume of observations is ingested within ERA5. The latter is expected to provide an improvement of the accuracy in the simulation of latent heat fluxes determined by vegetation and land cover change. On the other hand, attempts to implement more complex land cover and vegetation changes in the ECMWF model's land cover characterization (by leveraging on state-of-the-art earth observations) showed mostly neutral impacts on the simulated surface latent and sensible heat fluxes when compared against 51 FLUXNET stations over 1996–2014 over Europe (Nogueira et al., 2021)."

Beside this, we still think that GLEAM is a really good product but we preferred to use ERA5 actual evaporation just for being consistent with the drought characterization that we have performed. Both precipitation and evaporation come from the same dataset and this provides a more robust analysis (this was already underlined in the original version of the manuscript). Indeed, a similar result was also obtained by calculating evaporation as ET = P−Q and thus neglecting the contribution of the change in storage, as in Teuling et al. (2013).

Miralles, D. G., Jiménez, C., Jung, M., Michel, D., Ershadi, A., McCabe, M. F., Hirschi, M., Martens, B., Dolman, A. J., Fisher, J. B., Mu, Q., Seneviratne, S. I., Wood, E. F., and Fernández-Prieto, D.: The WACMOS-ET project – Part 2: Evaluation of global terrestrial evaporation data sets, Hydrol. Earth Syst. Sci., 20, 823–842, https://doi.org/10.5194/hess-20-823-2016, 2016.

Qiu, J., Crow, W. T., Dong, J., and Nearing, G. S.: Model representation of the coupling between evapotranspiration and soil water content at different depths, Hydrol. Earth Syst. Sci., 24, 581–594, https://doi.org/10.5194/hess-24-581-2020, 2020.

9. Lines 125-128. The use of yearly data to assess drought severity is not common and affected by problems to assess monthly and seasonal deficits, which are strongly relevant to understand drought dynamic and the connection between meteorological and hydrological droughts.

Lines 157-169: This is a subjective and confuse identification of drought events. I think a diagram illustrating the different steps would be necessary here. In any case, I would say that the runoff response to precipitation deficits is highly variable and depends on several factors but usually hydrological droughts respond to short time scales of meteorological droughts (e.g. https://hess.copernicus.org/articles/20/2483/2016/hess-20-2483-2016.html, https://www.sciencedirect.com/science/article/abs/pii/S0022169418308813) so probably to determine multiannual droughts characterised by annual data is not the best way to analyse the relationship between meteorological and hydrological droughts.

We understand the concern of the reviewer. However, our methodology to characterize multi-annual drought periods, also used by other studies (e.g., Saft et al. 2016), is not subjective, as it relies on annual precipitation anomalies. We also cross validated our results with previous studies (Marsh et al. 1994, Sheffield and Wood, 2011, Fink et al. 2004, Peters et al. 2006, Spinoni et al. 2015 as well as other drought reports) and found full consistency. We have expanded the validation of our methodology at lines 228-237:

"During the last five decades, Europe has experienced various multi-year drought episodes (Parry et al., 2012; Spinoni et al., 2015; Hanel et al., 2018), which have been perhaps less studied but are as relevant as those that have impacted other world's regions, such as Australia, California, or South America (Dijk et al., 2013; Griffin and Anchukaitis, 2014; Garreaud et al., 2017). For instance, the 1995-1997 multi-year drought impacted almost all Central and North Europe, but unlike episodes prior to 1979 (which were not taken into consideration here), it had a limited initial spatial extent and coherence on a regional basis, with a late exacerbation in terms of severity and extent by 1997 (Parry et al., 2012). The 1989-1991 drought impacted Belgium, France, Luxembourg, The Netherlands, as well as Balkan countries, the Mediterranean (Spinoni et al., 2015) and also UK (Marsh et al., 1994, Peters et al., 2006). The period from 1992 to 1995 was one of the driest in the century for the Iberian Peninsula, and especially for Spain (Sheffield and Wood, 2011; Domínguez-Castro et al., 2019). The 2000-2005 period was also impacted by a severe drought in Northern Italy and Italian Alps (Fink et al., 2004) and Scandinavia

(https://www.geo.uio.no/edc/droughtdb/edr/DroughtEvents/2003_Event.php last access: 07-01-2022). The drought characterization we included in this work, based on precipitation anomalies (Saft et al., 2016, see section 2.4 for details and motivation of this choice), provided consistent results with the above-mentioned studies (see Figure 2). In particular, the 1989-1991 in UK and France and the severe drought that peaked in 1992 in Central Europe (see Table 4 of Spinoni et al. 2015) were identified by our procedure. The same holds for the long drought which hit Spain during 1990-1995 also mentioned by of Sheffield and Wood and the 2000-2005 drought over the Alps and Scandinavia."

Moreover, we would like to stress the fact that our analysis is focused on the yearly water balance and its application to intra-annual or seasonal variations is not straightforward. On the other hand, it does not exclude mechanisms of propagation from meteorological to hydrological drought happening at shorter time scales rather, it bulks them together within the water year. This can cause a loss of resolution but not a loss of water under the water balance perspective.

The analysis repeated with the use of the classical SPI-12 index applied to monthly data as suggested by the reviewer reveals the same temporal pattern of precipitation deficit identified by yearly-based SPI (or precipitation anomalies) and thus lead to very similar results for the multi-year drought (MYD) definition (see Figures R1 and R2 below for the two specific catchments also discussed in the paper in north and South Europe). Therefore, for the purpose of the manuscript we do not see our method to impact significantly our analysis. We have added some text at lines 191-196 of the revised manuscript to clarify the issue:

"Note that we also:

1. tested different thresholds (i.e., larger than 3 years), without noticing any significant difference in the results;
2. cross validated our drought definition (i.e., based on yearly precipitation anomalies divided by the standard deviation) technique with the use of a classical SPI-12 based on monthly data (not shown) and found almost identical patterns between the two. However, given that our analysis focus on annual water balance we preferred to maintain the use of the precipitation anomalies.

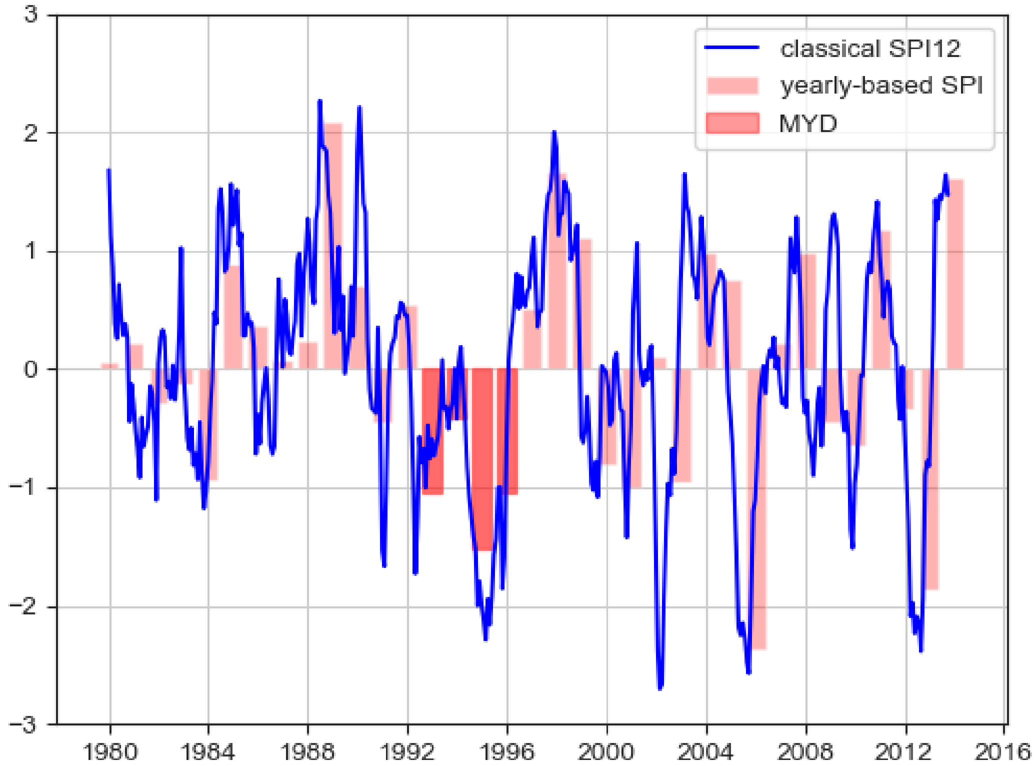

Figure R1: comparison between yearly-based SPI calculated with the procedure described in the manuscript (where the precipitation anomalies divided by the standard deviation were used to identify MYD periods) and the classical SPI-12 for the catchment located in Northern Europe.

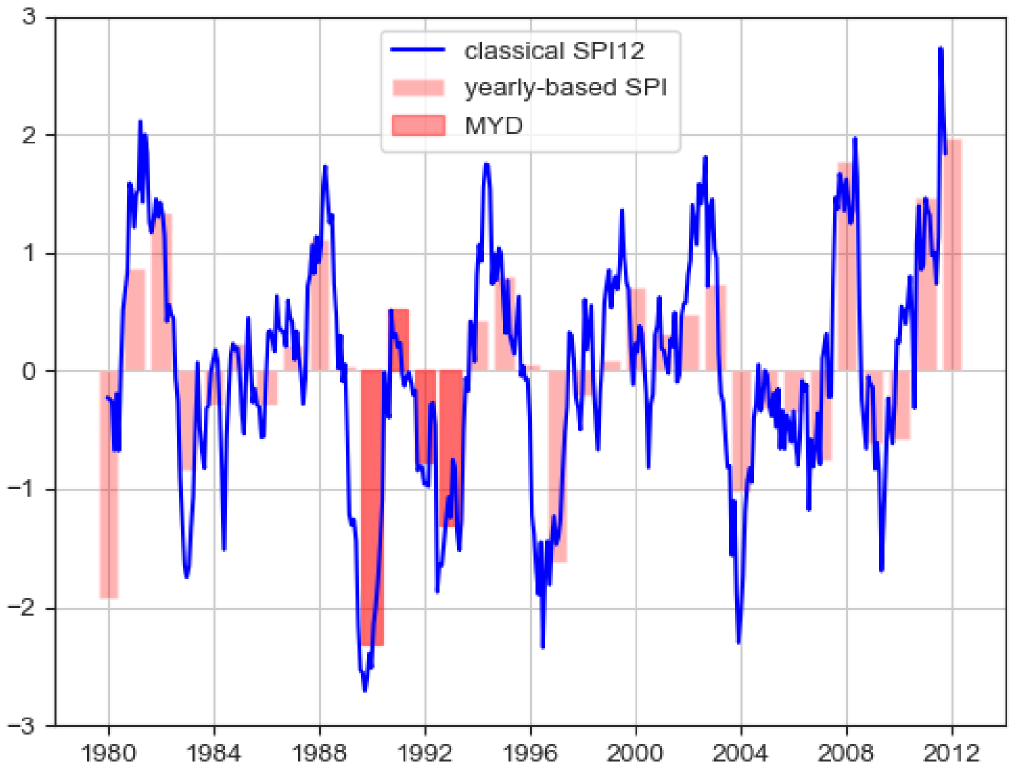

Figure R2: comparison between yearly-based SPI calculated with the procedure described in the manuscript (where the precipitation anomalies divided by the standard deviation were used to identify MYD periods) and the classical SPI-12 for the catchment located in Northern Europe.

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

Marsh TJ, Monkhouse RA, Arnell NW, Lees ML and Reynard NS. 1994. The 1988–92 Drought. Institute of Hydrology: Wallingford, UK.

Peters, E., Bier, G., van Lanen, H. A. J., and Torfs, P. J. J. F.: Propagation and spatial distribution of drought in a groundwater catchment, J. Hydrol., 321, 257–275, https://doi.org/10.1016/j.jhydrol.2005.08.004, 2006.

Fink, A. H., Brücher, T., Krüger, A., Leckebusch, G. C., Pinto, J. G., and Ulbrich, U.: The 2003 European summer heatwaves and drought –synoptic diagnosis and impacts, Weather, 59, 209–216, https://doi.org/10.1256/wea.73.04, 2004.

10. Line 173. I think Figure 2 is not showing the validation of the methodology. It is only showing the date of drought onset for the different events. It is also curious that no relevant drought events are recorded for the decades of 2000 and 2010 although strong droughts were recorded during the period (e.g., in 2012 and 2017 in southern Europe and 2018 in central and northern Europe).

We have changed Figure 2 again and reported only the year of the most intense drought year within the MYD period identified as well the MYD duration (see new Figure 2 below). The fact that single drought years were not considered is again because we were not interested in single (even severe drought), but we focused on multi-year droughts. This does not mean that they were not identified as drought years, but simply that we did not consider them in our analysis of the precipitation-runoff relationship. This of course does not imply that the shift cannot show even during single years.

Furthermore, more recent drought periods were not considered as our runoff dataset does not extend to 2018 for many catchments.

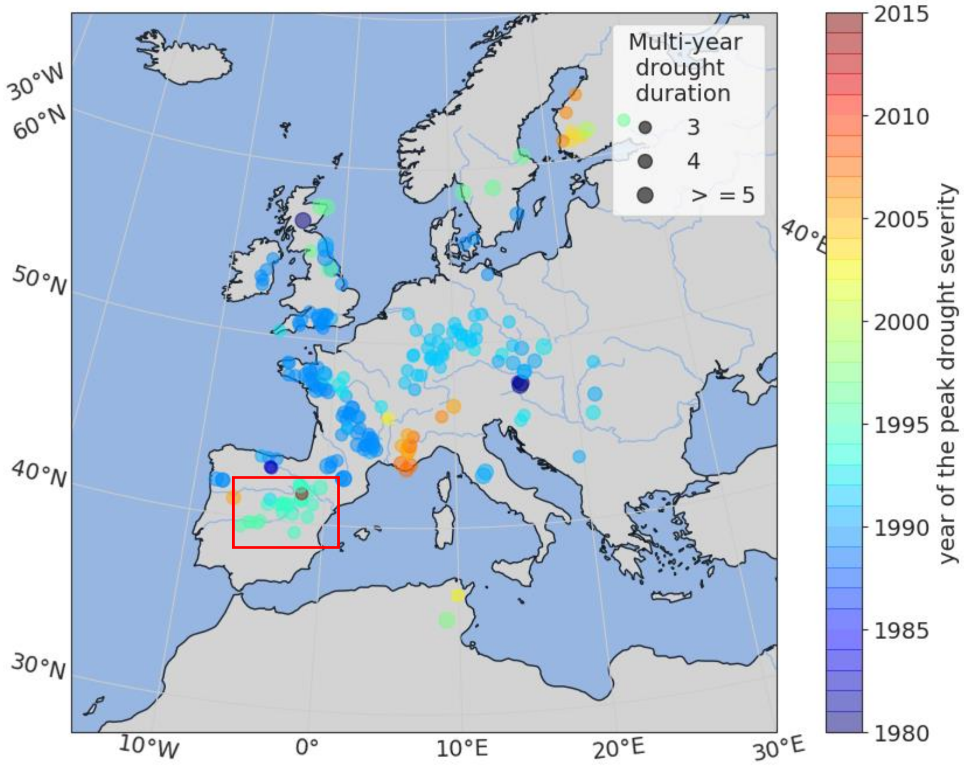

Figure R3: multi-year drought identified with the procedure described in section 2.4.

"Figure 2: Year of the most intense drought (the most negative precipitation anomalies) within the multi-year drought period identified as well as its duration for each basin. Note that other years might have shown more severe droughts but they were shorter than the 3 years period we have defined."

11. Line 214: 2000-2004 were not particularly dry in the Iberian Peninsula. Only 2005 was a extreme dry year. See https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.6126 Under my opinion the method used by the authors to identify the dry periods based on multiannual data may generate these artefacts identifying long-term droughts that are only associated to particular annual drought conditions. Main multiannual drought period in the Iberian Peninsula was recorded between 1990 and 1995, which caused strong hydrological deficits and socioeconomic impacts.

Indeed, our multi-year drought definition is consistent with the statement of the reviewer. That is, in Spain we identified exactly the multi-year drought period cited by the reviewer between 1990 and 1995 (see new Figure 2 of the revised manuscript, or even the Figure 2 in the original manuscript, or alternatively Figure R3 above with the square red box pointing to Spain).

12. Line 217-222. Probably selecting annual periods (instead multiannual) and particularly summer streamflow deficits, this pattern would be reinforced. I have the impression that the use of the multiannual criteria is probably masking this stronger reduction of streamflow in response to precipitation deficits (simply given the alteration of the partition between blue and green water). It is in summer when stronger transpiration by plants is recorded so probably considering both cold and warm seasons together is reducing the basins in which this effect is expected as in winter the

relationship between precipitation and streamflow is higher given little influence of water consumption by plants (see e.g.. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021GL094672)

Line 236. I suggest to include an annual and seasonal analysis in order to determine this issue. Thus, from a conceptual point of view this should be much more correct as the multiannual criteria may mask very different behaviour between years and seasons.

Our results do not go against what stated by the reviewer, and the found exacerbation of runoff at annual time scale does not exclude that this exacerbation can be even stronger seasonally or for specific summer months. On the other hand, this exacerbation will surely have an impact on the annual water balance, both because of the increased ET due to the higher evaporative demand during dry years and because winter precipitation (even when substantial) is used to replenish empty storage as a result of summer droughts. So, we believe that calculating the overall impact on runoff of ET and storage on a yearly basis is not wrong, although we agree that it only tells part of the story.

We also want to stress again that working at seasonal scale can be very difficult in a long-term perspective, as the definition of dry and wet seasons 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

13. Line 246-249. In this kind of relationships between climate and water availability, probably it is much more important to consider vegetation characteristics than a simple aridity index (see e.g. https://hess.copernicus.org/articles/11/983/2007/hess-11-983-2007.pdf I stress again the need of considering vegetation characteristics and vegetation changes in the analysis to interpret well the relationship between meteorological and hydrological droughts.

The reviewer is right and we overall agree with this statement. However, considering vegetation (for instance a classical vegetation index like NDVI or LAI) is not always straightforward because phenology, type of vegetation, specific response of plants to water stress and hydrological regulation of rooting depth all might exert a role. We also have attempted to find relation between vegetation type (obtained from Corine Land cover), rooting depth (from the European Soil Database) and the tendency of basins to shift, but we did not find robust results except for the rooting depth (Figures R3, R4).

Our explanation to this is that especially over Europe specific land cover classes are hardly really dominant and the basins are characterized by multiple non dominant land cover classes.

Figure R4 displays box plots representing the percentage of coniferous and broadleaf forests against the tendency of basins to shift (0 → no shift, 1→ shift). It can be seen that those land cover percentages are well below 50% in general (so they are not really dominant land cover classes) and that – specifically for broadleaf forests – there is a tendency for shifting basins to have larger percentage of this class (despite this is not statistically significant according to the Kolmogorov-Smirnoff test).

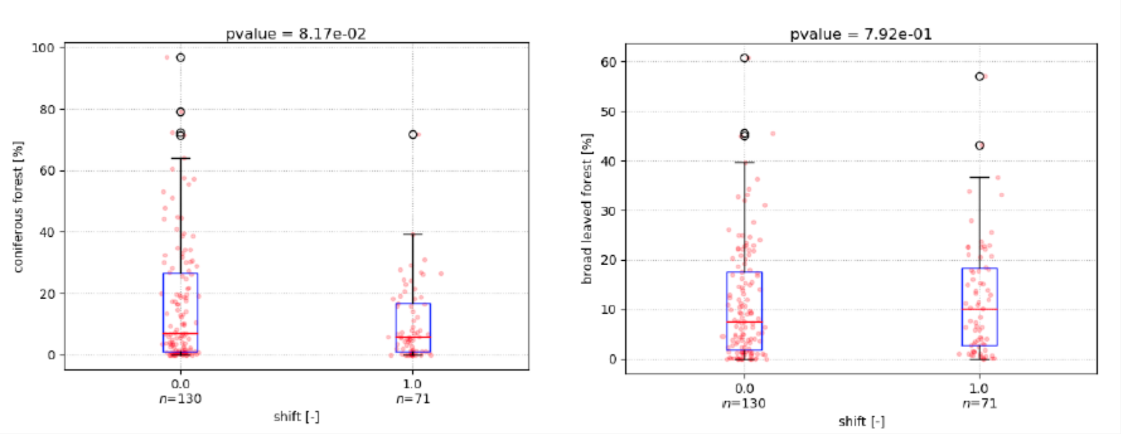

Figure R4: land cover class of broad leave forests and coniferous forests as a function of the shifts of the basins.

Figure R5 plots the same separation between shifting and not shifting basins as a function of the rooting depth. It can be seen here that larger rooting depth are observed for shifting basins, supporting our hypothesis that access to deep water storages by plants might support high evaporative demand during dry periods.

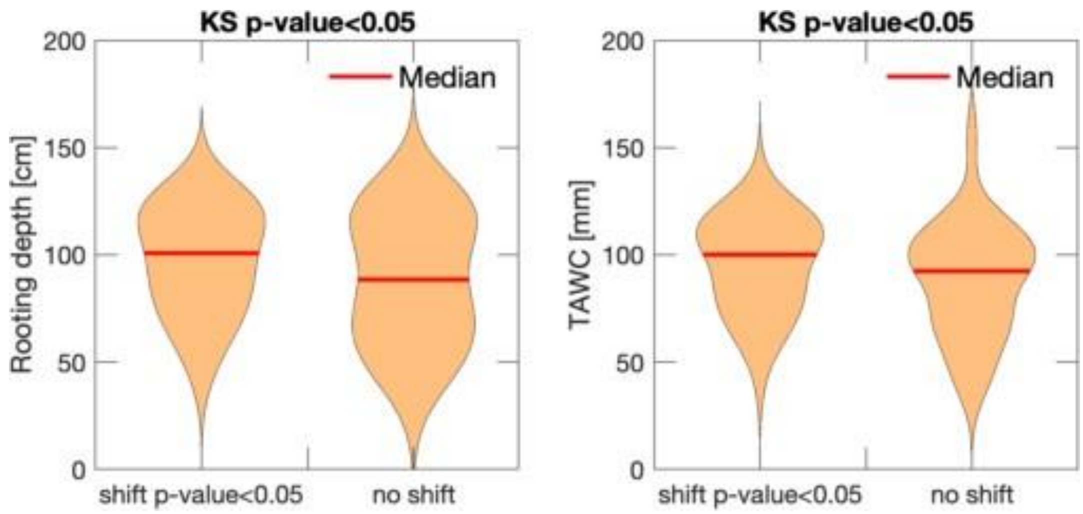

Figure R5 (Figure 7 of the manuscript): Average rooting depth and total available water (TAWC) content for basins characterized by a significant shift in the precipitation-runoff relationship and those where shift was not significant. KS refer to the two- sample Kolmogorov-Smirnov test between the distribution of basins with shift (p-value<0.05) and basin showing not statistically significant shift.

14. Lines 250-262. Again, I think these results could be affected by vegetation coverage. Why not to use a simple average vegetation index, LAI or forest coverage from Corine land cover instead of an aridity index?

    See our reply to point 13.

15. Lines 259-261. I think this results is coherent with what should be expected as atmospheric demand is higher in arid basins. Nevertheless, is this statement contradicting the message that shifts are obtained in both humid and dry basins? In any case, I think that the interpretation of this figure can be biased by the basin area. To determine if the catchments that generate more or less runoff are affected by larger shifts the area of each basin should be considered (e.g. by means of the ratio between Q/Area in the x axis).

    We agree with the suggestion of the reviewer and indeed we have already considered this. All the calculations since the original version of the manuscript were carried out in terms of runoff and not in terms of river discharge. We have pointed this out this at lines 164-166:

    "To avoid influence of the catchment area and to be coherent with the units of precipitation and evaporation streamflow measurements have been expressed in terms of runoff (i.e., we have considered the catchment area to transform m^3/s to mm/year)."

16. Line 270. For figure 5 I am really curious on how to explain in water limited regions an increase during drought periods. In arid lands precipitation is the limitation for ET so the majority of precipitation is evaporated as atmospheric evaporative demand is very high (e.g. > 1200mm/year in southern Europe), so if precipitation is reduced during multiannual droughts, how this enhancement of ET suggested by Figure 5 is explained? I would expect a reduction of streamflow in these cases but not ET enhancement.

    This is another example of complex feedback mechanisms between precipitation and evaporation. We think that a major role here is played by the plasticity of plants that can access water from deeper reservoirs in water-limited regions and/or during dry periods, and tend to adapt to the specific climate to survive, therefore enhancing evaporation. This behaviour has been found in many studies (e.g., Amin et al., 2020; Rempe and Dietrich, 2018; Hahm et al., 2019; Carrière et al., 2020, Klos et al., 2018) and was discussed at lines 375-381 of the newly revised manuscript:

    "A potential explanation to this can be given by the capacity of deep-rooted trees to access water from deep reservoirs and even weathered highly porous saprock or rock moisture (Rempe and Dietrich, 2018; Hahm et al., 2019; Carrière et al., 2020, Amin 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 especially over water limited regions, 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).

We have also tried to address this by plotting the root depth distribution (see Figure R5) for basins experiencing a significant shift (vs non shifting basins) in the precipitation-runoff relationship. We found that the latter are characterized by larger values of this variable. Nonetheless, further evidence is needed to corroborate this finding.

Our findings are also in line with what found by Destouni and Orth (2018), who stated that drought reduces blue-water fluxes more strongly than green-water fluxes in Europe and the latter are often seen an increase especially at the beginning of the drought period.

17. Line 344. But vegetation is not included in the analysis of this article, so how this statement is supported by the authors based on their results? Or maybe are they assuming the importance of this factor but they did not address in their analysis

Thank you for pointing this out. We have reformulated the sentence. We meant that the increased evaporation can be likely caused by vegetation (given that large part of the evaporation fluxes is caused by transpiration we think this is a reasonable assumption). The sentence was reformulated anyway (see lines 370-373).

"The key role of evaporation was also
addressed in Europe by Orth and Destouni (2018) and points to the vegetation as the potential driver (Vicente-Serrano et al. 2014, Peña-Gallardo et al. 2016, Peña-Angulo et al. 2021) causing enhanced evaporative demand during drought. Despite this was not clearly demonstrated in this work, the cited literature and Figure 6b suggest that this likely the case."