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.

R: We thank the reviewer for underlining the importance of this 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 will harmonize the literature and Introduction section by including many of these references in the text.

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.

R: The sentence was too generic indeed. We will rephrase it.

2. Line 39. It should be Van Loon

R: We will correct 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#!).

R: 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

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).

R: 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 the 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:

"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.

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).

R: We agree with the reviewer that 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).

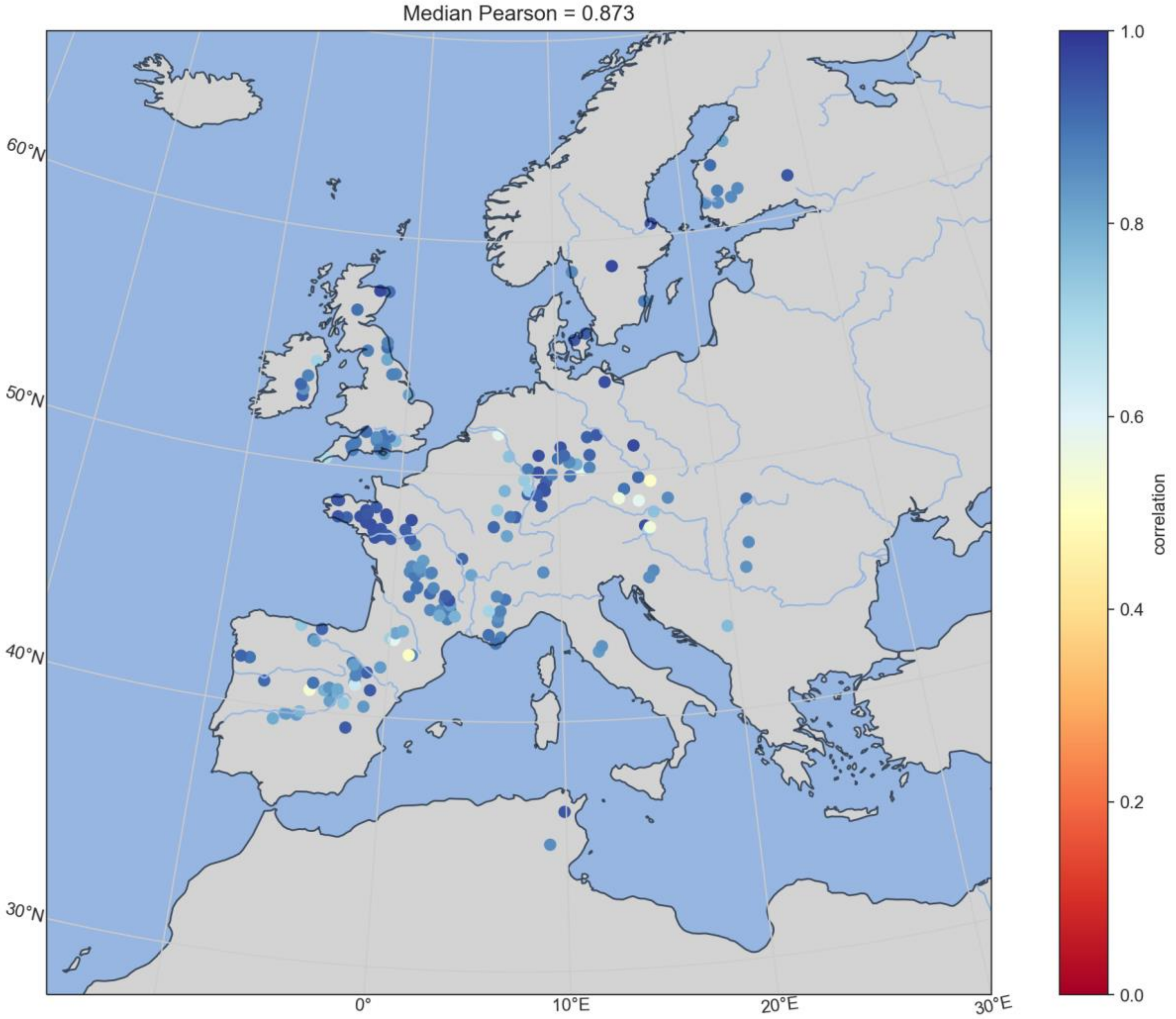


Figure R1: Pearson annual correlation coefficient between ERA5 and E-OBS precipitation (Period 1980-2016). All values of correlation are significant at the 5% level.

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.

R: 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.

R: 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.

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).

R: Please, see our reply to point 2. 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 are 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.

R: 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).

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/).

R: We considered annual cumulative precipitation in the SPI calculation because we are 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 will rely 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.

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.

R: The reviewer is right. We will try to improve the figure or remove it from the manuscript.

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).

R: 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.

R: Please, see our replies to points 3 and 4.

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

R: we thank the reviewer for highlighting this point. This is actually not a new method, sorry for the confusion, we will reformulate 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).

R: We agree with the reviewer, we missed the citations of these important pieces of work. We will include them in the revised manuscript by adding also a discussion.

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.

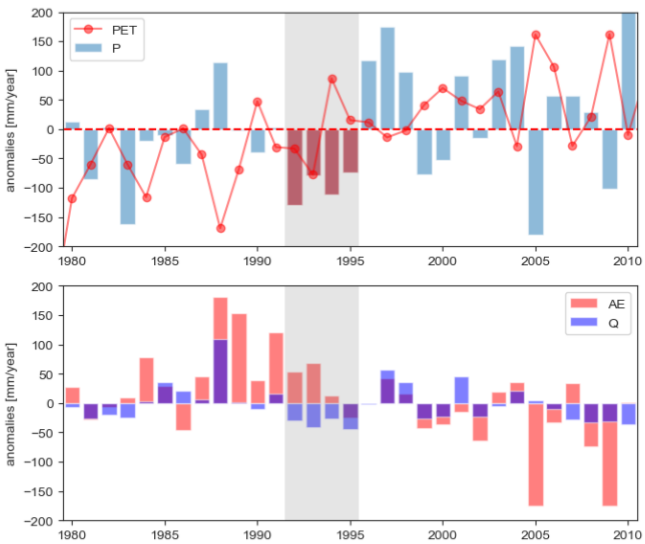
R: The reviewer raises an important and interesting point. Notwithstanding the fact that we think that it is important to consider the annual time scale because of the paramount role of the catchment storage in explaining the annual runoff response of the basin and thus not only ET (see response to remark 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).

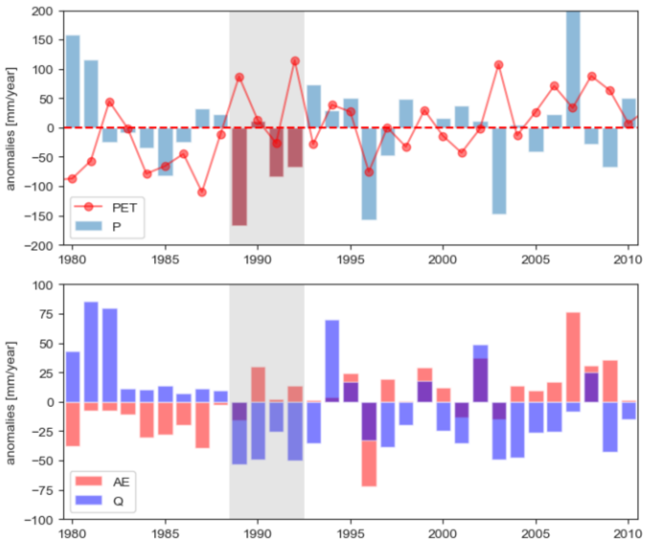
- Feng, X., Thompson, S.E., Woods, R., Porporato, A., 2019. Quantifying Asynchronicity of Precipitation and Potential Evapotranspiration in Mediterranean Climates. Geophys. Res. Lett. 46, 14692–14701. https://doi.org/10.1029/2019GL085653

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.

R: 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).



(North Europe, LON 6.8O, LAT 53.6N)



(South Europe, LON 5.4O, LAT 39.8N)

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.

R: 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. We double checked the position of the basin and it is located in forested region (see Figure R3) on a hilly terrain and with mean monthly air temperature from E-OBS dataset not particularly cold (see in Figure R4 the time series temperature). Although we cannot exclude effects of micro climate it is very likely that this catchment is representative for water limited regime. Regarding the human impact the considerations done above holds, specifically the selected basin does not seem highly anthropized (again see Figure 3).

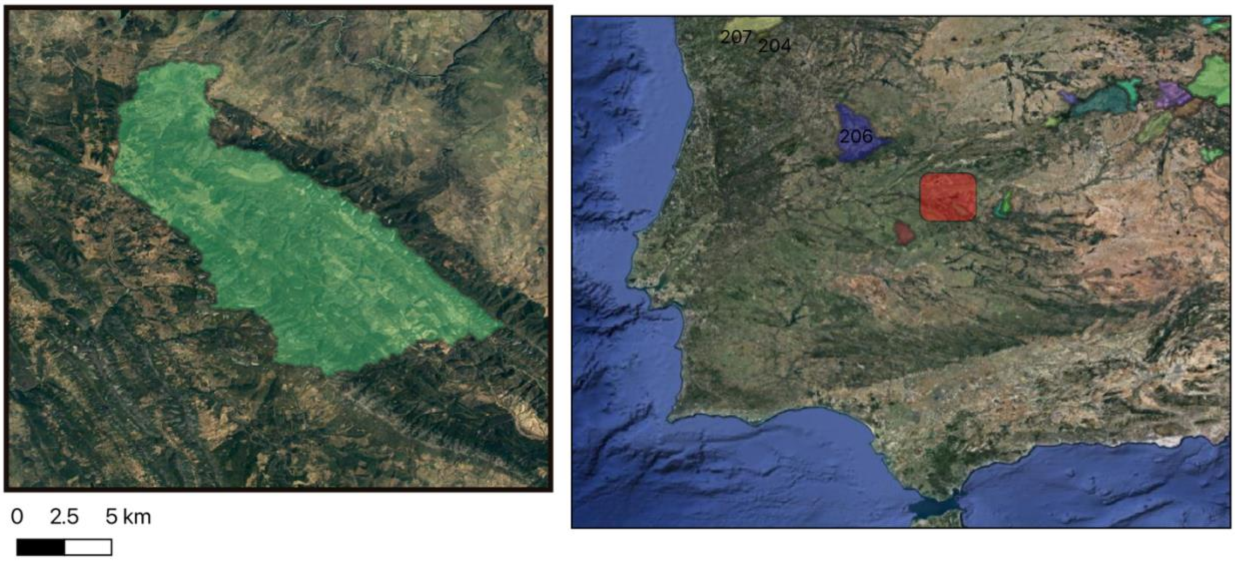


Figure R3: Location of the catchment in southern Europe.

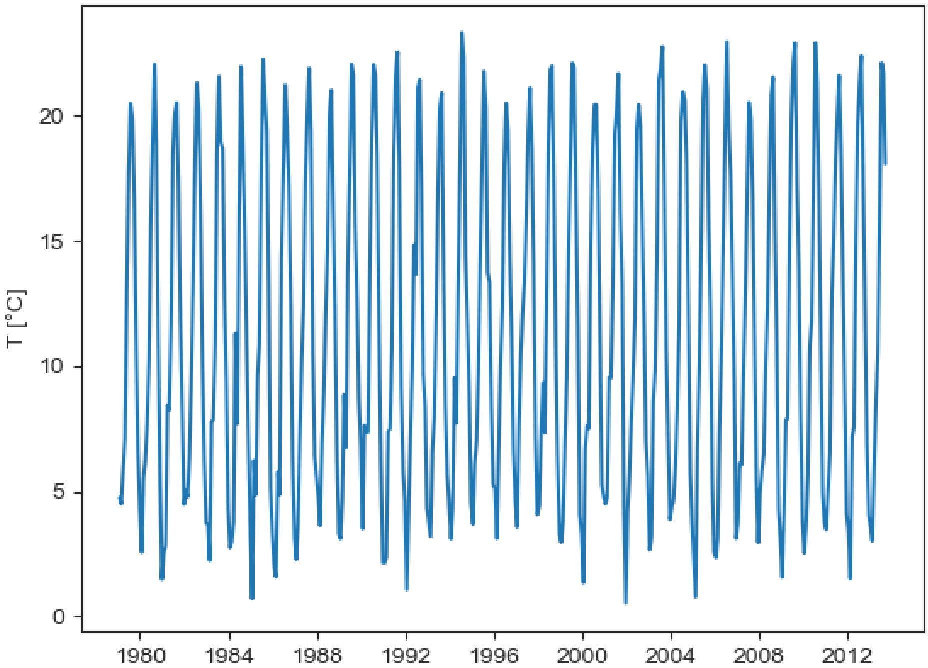


Figure R4: Monthly E-OBS temperature.

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.

R: 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 R6) and it does not seem such a highly populated area to explain significant urban water consumptions.

We will discuss better the limitations of the study in the revise version of the manuscript.

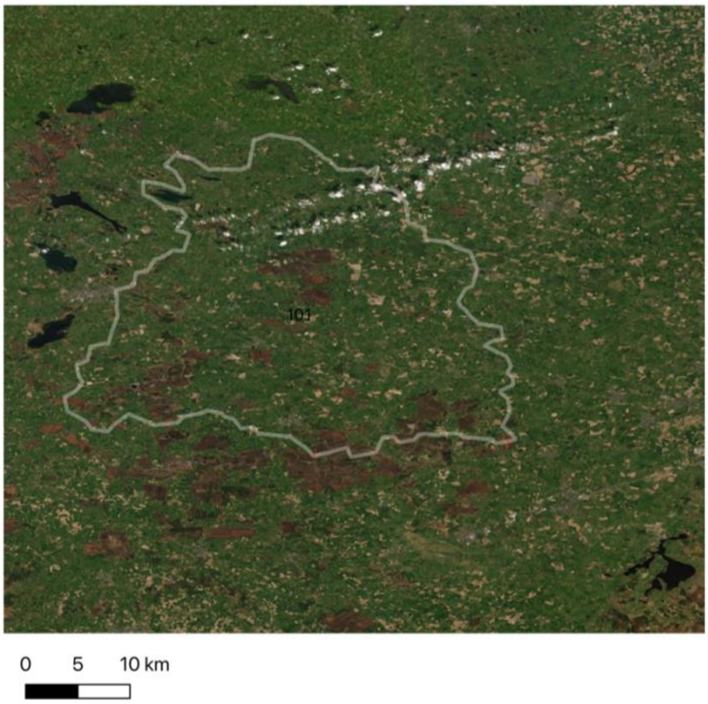


Figure R5: catchment located in Northern Europe.