

Dear Editor,

We thank you and the two Referees for your time in reviewing our manuscript and your constructive feedback on it. Please find below our point-by-point reply to your comments (italic) and the changes we did in the manuscript to address them (underlined). We also attach the revised version of the manuscript, with tracked changes as well.

Best regards,

Giulia Bruno and co-authors

Editor:

*Dear Authors,*

*As suggested by both referees, the manuscript requires major revisions to address the key concerns raised during the review process. These revisions are critical to ensuring that the manuscript meets the expected standards of quality, rigor, and clarity. The feedback provided by the referees includes constructive criticism and detailed recommendations that aim to enhance the overall impact and scientific validity of the work.*

*Additionally, the authors have proposed modifications to further refine the manuscript, incorporating new insights, improved methodologies and adding new figures, tables and citations. These changes will not only respond to the referees' comments but also strengthen the manuscript by addressing the identified gaps and ambiguities.*

*The major revision process shall involve the following steps:*

- Carefully addressing each point raised by the referees to ensure all concerns are adequately resolved.*
- Integrating the proposed modifications to improve the structure, clarity, and scientific depth of the manuscript.*

We have now implemented all Referees' suggestions in a revised version of the manuscript. In particular, we introduced additional analyses regarding (i) the multiple linear regressions at the catchment scale as suggested by Referee #1, and (ii) long-term variations in alternative streamflow (Q) metrics and (iii) the evaluation of our approach for the estimation of the dynamic storage ( $S_{dyn}$ ) of the catchments as suggested by Referee #2. Also, we discussed in more detail the reasons for our methodological choices and potential uncertainties in the study, the latter in a new section (4.4. Sources of uncertainty).

Referee #1:

*This paper analysed the summer low flows by using the datasets from 363 small catchments. The relationships among evapotranspiration (E), precipitation (P) and streamflow (Q), as well as the storage (S) were quantified. Results showed that summer low flows decreased significantly, and increased E played the main driver in the eastern catchments. In addition, the P-Q relationship changed in 26% of catchments between 1989 and 1993. Generally, the structure of the paper is clear and well-organized, however, there are some concerns.*

We thank you for appreciating the clarity of our manuscript and for raising interesting points to improve our work.

*Specific comments:*

1. *The coefficients of  $a$  and  $b$  in equation (2) should be vary with wind direction and elevation of gauges. Does authors calculate them for each gauges? If so, please provide the analysis which are the main uncertainty for precipitation datasets.*

The method we used for the correction of precipitation ( $P$ ) for gauge undercatch was developed for point data (Richter, 1995), with coefficients depending on gauge characteristics (wind exposure) and meteorological conditions ( $P$  type, Eq. 2). Here, we used a gridded  $P$  dataset (Sect. 2.2) and we applied the correction procedure at pixel-scale, by assuming all pixels as moderately sheltered with respect to wind exposure. This approach was used also in previous works (Duethmann & Blöschl, 2018; Bruno & Duethmann, 2024). Duethmann & Blöschl (2018) showed that alternative assumptions regarding the corrective coefficients had small influences on long-term variations in  $P$  and water-balance derived catchment evapotranspiration ( $E$ ), we are mostly interested in for our analyses. Thus, we argue that this assumption unlikely affected our main conclusions. We agree, however, that  $P$  data are unavoidably affected by some degree of uncertainty, either per se and following this correction, which we did not discuss in the original manuscript. We modified L139–143 to make clearer our assumption regarding the correction procedure and we added some discussion concerning the uncertainty in  $P$  data in the new Sect. 4.4.

L139–143: We corrected the dataset for gauge undercatch following the method proposed for Germany by Richter (1995):

$$P_{corr} = P_{uncorr} + aP_{uncorr}^b \quad (2)$$

with  $P_{corr}$  as corrected  $P$ ,  $P_{uncorr}$  uncorrected  $P$ ,  $a$  and  $b$  coefficients which vary with wind exposure of the gauges, precipitation type (rain or snow), and season. Here, we assumed for all grid cells the coefficients for moderately sheltered locations in Richter (1995), given the low sensitivity of long-term variations in  $P$  to the selected coefficients (Duethmann & Blöschl, 2018).

Sect. 4.4: Uncertainties in  $P$  can arise from potential inhomogeneities in gauge data over time and gauge undercatch. Here, we used a gridded  $P$  dataset from the interpolation of a fixed number of gauges over time to minimize inhomogeneities and we corrected it for gauge undercatch (Sect. 2.2). This correction procedure may lead to further uncertainties, but Duethmann and Blöschl (2018) demonstrated that assumptions regarding corrective coefficients have little influence on trends in  $P$  and  $E$ , we were mostly interested in here.

2. *It is quite difficult to understand the equation (4), where the dynamics of storage was approximated by  $P_{mam}$  and  $P_{djf}$ . Please provide more explanations. In addition, since baseflow is 0.66 in this area, which means the soil moisture and groundwater both plays important roles in runoff variation. But it seems they are not taken into account in the analysis.*

We agree that soil and groundwater storage are relevant for  $Q$  generation in the study catchments. As predictors of trends in summer low flows, we indeed considered variations in  $E$ , summer  $P$ , and storage ( $S$ ). We accounted for the influence of  $S$  variations using winter precipitation ( $P_{DIF}$ ) and spring precipitation ( $P_{MAM}$ ) as wetness conditions in the seasons preceding the summer and therefore, proxies for  $S$  recharge. We used these proxies since long-term data on soil moisture and groundwater are unavailable for the large sample of small catchments that we analyzed, similarly to what done by previous works (Duethmann et al., 2015; Saft et al., 2016; Laaha et al., 2017). We rephrased L174–176 as follows and discussed this point in the new Sect. 4.4.

L174–176: We used  $P_{\text{MAM}}$  and  $P_{\text{DJF}}$  as proxies of storage recharge in the seasons preceding the summer similarly to what done by previous work to overcome the unavailability of long-term data on soil moisture and groundwater storage for the study catchments (Duethmann et al., 2015; Saft et al., 2016; Laaha et al., 2017).

Sect. 4.4: Finally, as potential predictors of changes in summer low flows we approximated storage processes with  $P$  in the season preceding summer, due to unavailability of long-term  $S$  data for the study catchments. We chose this approach instead of using alternative proxies for  $S$  (e.g., estimates of  $S_{\text{dyn}}$  or baseflow from  $Q$  data) to avoid dependences between predictors and target variable (summer low flows). The satisfactory performances of the multiple linear regressions and the plausible signs of their coefficients suggest the suitability of the selected predictors to represent the long-term dynamic of summer low flows (Table S1 and Fig. S6).

3. For multiple linear regression in predicting summer low flows, the authors showed the  $R^2$  in four clusters which exhibited good performance in Table S1. However, there is spatial variation among different gauges so the coefficients vary at each gauge, does the regression in the gauge scale follow the same trend with cluster?

In a first step, we performed the multiple linear regression at a cluster-scale to minimize uncertainties in  $E$  for specific catchments (see also reply to comment #4 by Referee #2). However, we see that such analysis may not fully reveal potential spatial differences in the predictors of the temporal dynamics of summer low flows. Thus, we repeated the analysis at the catchment-scale. We achieved overall satisfactory results in terms of model performances (median coefficient of determination across the catchments equal to 0.78, Fig. 1a and b here) and in line with those at a cluster-scale. By looking at the predictor with highest contribution to the simulation (primary predictor) for each catchment, we found that summer precipitation ( $P_{\text{JJA}}$ ) was the most recurrent one across all clusters (Fig. 1c).  $P_{\text{JJA}}$  was frequently non-significant though, especially where model performances were relatively low (Fig. 1a) and predictors had a similar relative contribution (not shown). By focusing on significant primary predictors only, the most recurrent ones were  $P_{\text{JJA}}$  in the Pre-Alpine cluster,  $P_{\text{MAM}}$  in the South-Central one, and  $E$  in the Eastern and Northern clusters (Fig. 1d). These are coherent with the conclusions we draw at the cluster-scale (i.e.,  $P_{\text{JJA}}$  dominant predictor in the Pre-Alpine cluster,  $P_{\text{MAM}}$  significant predictor in the South-Central one, and  $E$  in the Eastern and Northern clusters, Fig. 1). To reinforce our trend attribution in the revised version of the manuscript, we added this analysis with Fig. 1 here as new Fig. S6 in the Supplement, and adapted the description of methods and results as follows.

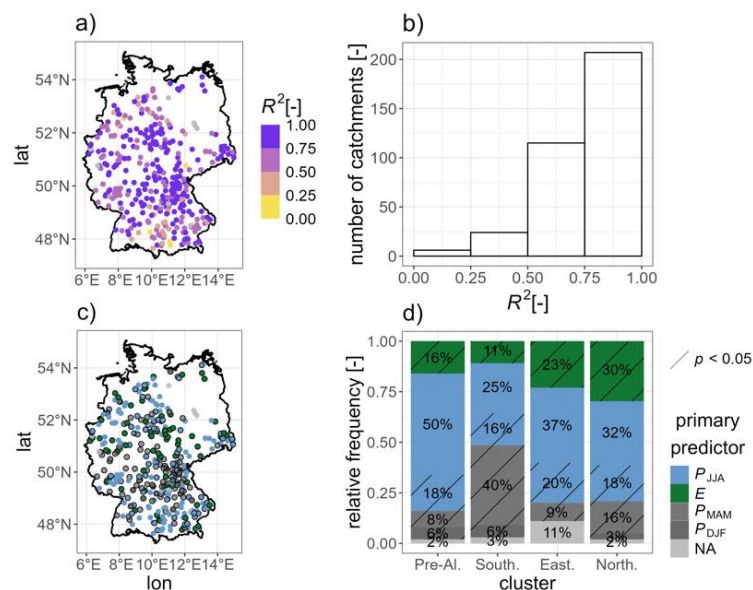


Fig. 1: Model results from the multiple linear regressions on catchment-scale data. (a) Map of coefficients of determination of the models ( $R^2$ ). (b) Histogram of  $R^2$ . (c) Map of primary predictors (black edges if significant). (d) Relative frequency of primary predictors by cluster. Light grey in (a, c, and d) refers to catchments with high multicollinearity of the predictors, and thus excluded from the analysis (Sect. 2.5). In (d), Pre-al. refers to Pre-Alpine, South. to South-Central, East. to Eastern, and North. to Northern cluster.

Methods, L178–183: Firstly, we used multiple linear regression to model the temporal dynamics of summer low flows ( $7dQ_{min, JJA}$  in Eq. (4)), from the dynamics of the predictors, both at a catchment- and cluster-scale:

$$7dQ_{min, JJA} = \alpha_1 E + \alpha_2 P_{JJA} + \alpha_3 P_{MAM} + \alpha_4 P_{DJF} + \varepsilon \quad (4)$$

with  $\alpha_i$  ( $i = 1...4$ ) the regression coefficient for each predictor and  $\varepsilon$  the model residuals. We adopted 5-year averages to focus on long-term dynamics and reduce uncertainties in water balance-derived  $E$ . Moreover, for the cluster-scale analysis we used average time series across the catchments in each cluster to minimize uncertainties in  $E$  for specific catchments, while analysing the main signal at a regional scale.

Results, L256–258: Multiple linear regression for predicting long-term dynamics in summer low flows achieved satisfactory performances both at cluster- and catchment-scale (for  $7dQ_{min, JJA}$ ,  $R^2 > 0.7$  for each cluster and median  $R^2$  of 0.78 across all catchments, Table S1, and Fig. S6a and b).

L262–265: Catchment-scale results showed similar patterns, despite being unavoidably more affected by noise than those at the cluster-scale (Fig. S6c and d). By focusing on significant primary predictors only (i.e., predictors with highest contribution to the simulations),  $P_{JJA}$  was the most recurrent one in the Pre-Alpine cluster,  $P_{MAM}$  in the South-Central cluster, and  $E$  in the Eastern and Northern clusters (Fig. S6d).

4. *Could authors discuss why the  $r = -0.49$  in  $P_{djf}$  for the easter cluster in lines 258?*

In the Eastern cluster, we indeed detected a negative correlation between trends in summer low flows and trends in  $P_{DJF}$ , which may sound counterintuitive under our assumption of  $P_{DJF}$  as a proxy for storage recharge. While we acknowledge that spurious effects in the correlation analysis may play a role here, we see a mechanistic explanation, related to  $E$ -storage feedbacks (Boeing et al., 2024) and the observed changes in this cluster. In particular, catchments in the Eastern cluster showed both positive and negative trends in  $P_{DJF}$ , but generally negative trends in  $P_{MAM}$  and positive trends in  $E$ . This means that increases in  $P_{DJF}$ , and thus in storage conditions at the beginning of the growing season, might have buffered the decreases in  $P_{MAM}$  in sustaining the increases in  $E$  in some catchments. Increases in  $E$ , in turn, contributed to decreases in summer low flows in this cluster (Fig. 6 and 7). Thus, increases in  $P_{DJF}$  may have indirectly contributed to decreases in summer low flows in the Eastern cluster. In the revised manuscript, we expanded the mechanistic explanation on this point in the Discussion as follows.

L347–349: These three mechanisms (i.e., widespread increases in  $E$ , local decreases in  $P_{JJA}$ , and local decreases in  $P_{MAM}$ , in the previous sentence not reported here) also overcompensated local increases in storage recharge during winter (approximated by  $P_{DJF}$ ), possibly through storage depletion by  $E$  (Boeing et al., 2024) for instance in the Eastern cluster (Fig. 7).

5. *For 363 catchments, there is only 15 catchment with negative  $Cp-q$  rel values which distributed sparsely in Fig.8(a). I am curious about the possibility caused by data process uncertainty.*

Having no changes in the  $P$ - $Q$  relationships even during prolonged dry periods was common expectation for humid catchments until recently (Massari et al., 2022) and a sparse occurrence of these changes is positive for water management (see e.g. Fowler et al., 2022 for practical implications of these changes). We see that our study revealed less widespread changes in the  $P$ - $Q$  relationship during the multi-year drought in Germany in the early 1990s than previous works for other case studies. We discuss that these differences may relate to the characteristics of the multi-year droughts (i.e., severity and duration of the  $P$  deficits) and to the hydro-climatic properties of the catchments. We agree that the unavoidable uncertainty in  $P$  and  $Q$  data (see reply to comment #1) may also play a role, despite we here aimed at minimizing uncertainty in  $P$  data by using a dataset specifically tailored to long-term consistency (Sect. 2.2). To add this point to the Discussion, we rephrased L390–395 as follows.

L390–395: According to this analysis, the multi-year drought in Germany in the early 1990s had less severe impact on  $Q$  generation than the Millennium drought in Australia (changes in 56 % of the catchments, with median decrease of approximately -50 %, Saft et al. 2015), the 2012–2016 event in California (mean decreases of -28 % across three catchments, Avanzi et al. 2020), and the 2010–2020 drought in Chile (changes in 61 % of the catchments and mean decrease of -19 %, Alvarez-Garreton et al., 2021). These differences may be related to the characteristics of the  $P$  anomalies (severity and duration), potential uncertainties in the underlying data (Sect. 4.4), and the hydro-climatic characteristics of the catchments.

6. *For Fig.5 and Table S1, it showed that  $P_{jja}$  played a major contribution to  $Q$  changes in Pre-Alpine and South-Central cluster, while  $E$  played much more contribution in Eastern and Northern area. Does it relate to elevation changes?*

Catchments in the Pre-Alpine and South-Central clusters have indeed relatively higher mean elevations than others (Fig. 3 in the manuscript). Furthermore, catchments in the Pre-Alpine cluster generally experienced decreases in  $P_{JJA}$  and mild increases in  $E$ . Catchments in the South-Central cluster showed similar behaviors, but overall small trends. On the contrary, catchments in the Eastern and Northern clusters largely had increases in  $P_{JJA}$  and  $E$ . Therefore, we argue that differences in the main drivers of decreases in summer low flows between the Pre-Alpine and North-Eastern areas can be ascribed to differences in the variations in the drivers themselves, rather than to differences in catchment characteristics like elevation. To make the differences in hydro-climatic changes among the clusters easier to grasp, we added the following summary table in Sect. 4.2.

Table 1: Summary of long-term variations in summer low flows and in their potential predictors (annual evapotranspiration,  $E$ , and precipitation over summer,  $P_{JJA}$ , spring,  $P_{MAM}$ , and winter,  $P_{DJF}$ ) over 1970–2019, and drivers of variations in summer low flows (only significant ones according to both attribution analyses, Sect. 2.5) for each cluster. Red arrows refer to strong decreases, light red arrows to mild decreases, light blue arrows to mild increases, and blue arrows to strong increases at the median level, with  $\pm 2$  % decade<sup>-1</sup> as thresholds for strong increases/decreases.

	Cluster			
	Pre-Alpine	South-Central	Eastern	Northern
Variations in summer low flows	↓	↓	↓	↓
Variations in $E$	↑	↑	↑	↑
Variations in $P_{JJA}$	↓	↓	↑	↑
Variations in $P_{MAM}$	↑	↓	↓	↓
Variations in $P_{DJF}$	↑	↑	↑	↑
Drivers of variations in summer low flows	$P_{JJA}$	-	$E$	$E, P_{JJA}$

Referee #2:

*Bruno et al. analyzed the decreasing trend of low flow in 363 small catchments in Germany. They also attributed the decrease to the increase of ET. They further unraveled that the change of P-Q relationship during drought produces lower flow which is generally due to the increased ET. They conducted this work based on observations of P and Q, empirical expression of subsurface storage, and ET derived based on water balance and statistical analysis. I think studying the decrease of low flow and trying to find the major drivers is very important in the climate change background. The data and analysis are generally reliable, the structure and the writing are good. However, there are the following concerns which need clarification from authors for further review.*

We thank you for your overall positive evaluation of our manuscript and for raising interesting points to improve our work.

- 1. I was more or less confused by the overall idea of the authors. If you wanted to check if the decrease of the flow is caused by increased ET, and you also have calculated the water balance, why not do a straightforward analysis of the overall change of P, ET, Q, and S. Then it is easy to get if the decrease of flow is mainly driven by ET. Then you can do the analysis in your manuscript as a follow-up. Otherwise, I feel the conclusions are even not that convincing as, for example, the decrease of low flow might be just because of the shift of the timing of streamflow.*

We understand that you suggest evaluating here the impact of changes in  $E$  on  $Q$  at an annual time scale. This has already been well documented for several regions and periods, partly overlapping with our case study as well (Teuling et al., 2009; Fischer et al., 2023; Renner & Hauffe, 2024). Here, we rather aimed at assessing the impact of increases in  $E$  on  $Q$  during dry periods specifically, to complement these previous findings. We see, however, that assessing changes in additional  $Q$  metrics, such as annual  $Q$  and its timing, may provide useful context regarding hydrological changes in the region and the decreases in the magnitude of summer low flows that we observed. Therefore, we quantified changes in annual  $Q$  and its timing, the latter through the center of mass date of  $Q$  ( $CMD_Q$ , Court, 1962; Han et al., 2024). Annual  $Q$  and  $CMD_Q$  generally showed decreases across the catchments between 1970 and 2019 (Fig. 2g-l here). However, these decreases were less significant than those in the magnitude of summer low flows, with a median trend in annual  $Q$  ( $CMD_Q$ ) of  $-1.3 \text{ \% decade}^{-1}$  ( $-0.006 \text{ month year}^{-1}$ ) across all catchments and significant negative trends in 10 (3) % of them (Fig. 2g-l here), as compared to a median trend in  $7dQ_{\min, JJA}$  of  $-3.7 \text{ \% decade}^{-1}$  across all catchments and significant negative trends in 31 % (Fig. 4).

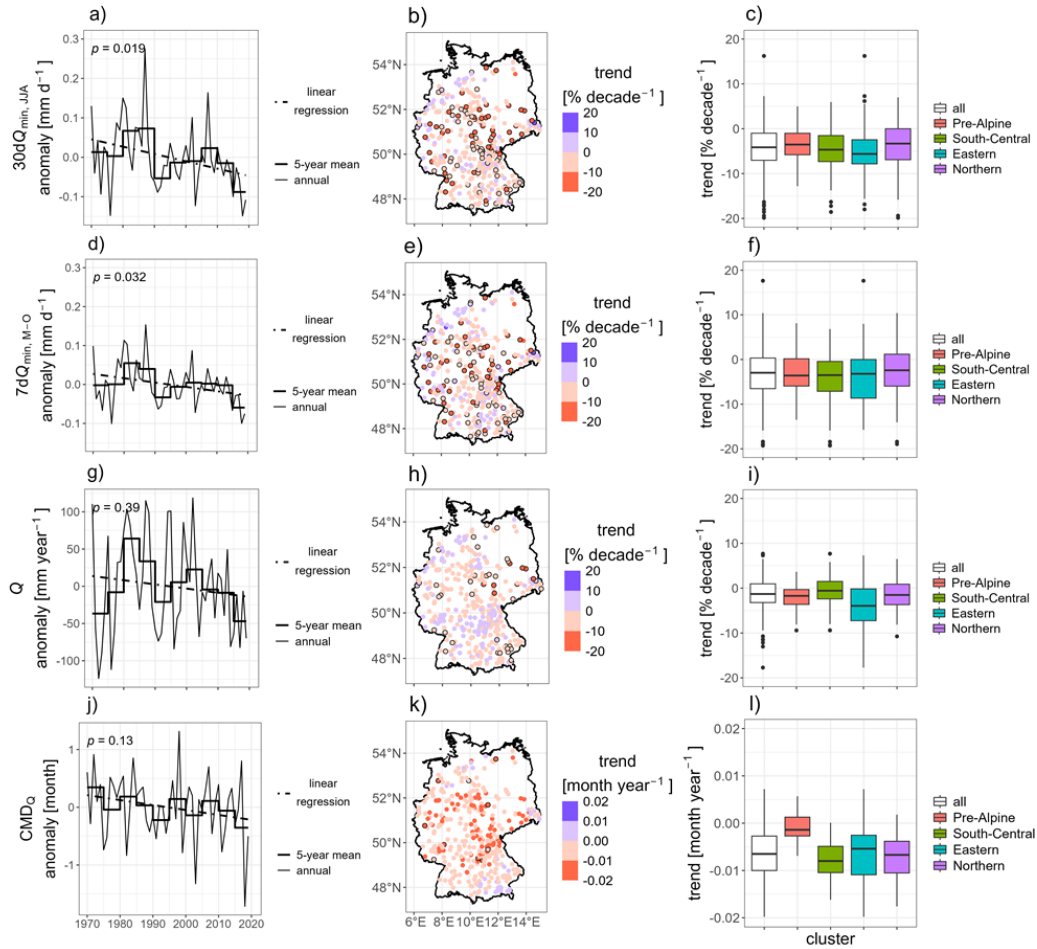


Fig. 2: Long-term variations in streamflow ( $Q$ ) metrics over 1970–2019 ( $30dQ_{\min, JJA}$ , panels a–c, and  $7dQ_{\min, M-O}$ , panels d–f, as metrics for summer low flows, annual  $Q$ , panels g–i, and the center of mass date of streamflow  $CMD_Q$ , panels j–l). (a, d, g, and j) Average anomalies across the study catchments. (b, e, h, and k) Map of catchment-scale trends (black edges if significant). (c, f, i, and l) Boxplots of trends for all catchments and by cluster.

Therefore, summer low flows in particular decreased across the study catchments between 1970 and 2019, as compared to other  $Q$  metrics, and these decreases did not occur simply with an overall decrease in  $Q$  or a shift in its timing. We introduced this analysis in the revised version of the manuscript, by adding Fig. 2 here as new Fig. S4 in the Supplement, and its methodological aspects, results, and discussion in the corresponding sections.

Methods, L166–168: We also characterised general long-term variations in  $Q$ , and considered annual  $Q$  and the centre of mass date of  $Q$  ( $CMD_Q$ ).  $CMD_Q$  is the time of the year when half of the annual  $Q$  has occurred and as such, a metric for  $Q$  timing (Court, 1962; Han et al., 2024).

Results, L230–233: Additional  $Q$  metrics similarly decreased over 1970–2019, even though less significantly than summer low flows (median trend in annual  $Q$  of  $-1.3 \text{ \% decade}^{-1}$  across all the catchments and significant negative trends in 10 % of them, Fig. S4g–i, and median trend in  $CMD_Q$  of  $-0.006 \text{ month year}^{-1}$  across all catchments and significant negative trends in 3 % of them, Fig. S4j–l).

Discussion, L325–328: Furthermore, decreases in annual  $Q$  and in its timing ( $CMD_Q$ ) were more elusive than those in summer low flows (Fig. S4g–l), meaning that summer low flows in particular decreased

across the study catchments over 1970–2019, and likely not simply due to a general decrease in  $Q$  or a shift in its timing.

Furthermore, we added more context around current knowledge on the effect of increases in  $E$  on decreases in  $Q$  at an annual time scale in the Introduction and Discussion as follows, also to highlight better the novelty of our study with respect to previous literature.

Introduction, L55–57: These increases in  $E$  contributed to decreases in annual  $Q$  on the long-term (Teuling et al., 2009; Fischer et al., 2023; Renner & Hauffe, 2024) and during dry years, especially in mountain catchments (Tran et al., 2023).

Discussion, L359–361: Our findings on the role of increases in  $E$  on decreases in summer low flows expand results at annual time scale (Fischer et al., 2023; Tran et al., 2023), and those by Montanari et al. (2023) on concomitant decreases in summer low flows and increases in  $E$  in northern Italy over recent decades.

2. *You mentioned you used Kirchner's approach to calculate  $S_{dyn}$  which needs that  $S$  is the main control of  $Q$  generation. I am wondering if the 363 catchments you used meet this requirement and where is your analysis for this?*

We agree that the assumption of  $Q$  mainly controlled by  $S_{dyn}$  may not always hold true across large samples of catchments, despite previous studies adopted it (Staudinger et al., 2017; Trotter et al., 2024) also in our study region (Stoelzle et al., 2013; Berghuijs et al., 2016). We have now evaluated this assumption similarly to what done by Kirchner (2009). Specifically, we computed the percentage of winter days (November to April, included) with (i) rising  $Q$  ( $dQ/dt > 0$ ) when precipitation ( $P$ ) exceeds  $Q$ , and (ii) decreasing  $Q$  ( $dQ/dt \leq 0$ ) when  $P$  does not exceed  $Q$ . This can be thought as the percentage of days when  $S_{dyn}$  is the main control of  $Q$ , with  $S_{dyn}$  replenishing when  $P$  exceeds  $Q$  (condition i) and depleting otherwise, with no additional main sources of sustainment for  $Q$  (conditions ii). We found that all catchments met these conditions for most of the days over the study period (Fig. 3 here). The percentage of days with  $S_{dyn}$  as main control of  $Q$  spanned indeed between 54 and 72 % across the catchments. This therefore shows the suitability of the method that we used for the estimation of  $S_{dyn}$  for our case study. We added this analysis to the revised version of the manuscript, by rephrasing L150–157 as follows.

Kirchner (2009) showed that  $S_{dyn}$  can be derived from the  $Q$  time series and recession characteristics of the catchments, under the assumption of  $Q$  mainly controlled by it. To verify this assumption for the study catchments, we followed an approach similar to Kirchner (2009). Specifically, we computed the percentage of winter days (November to April, included) over the study period with (i) rising  $Q$  ( $dQ/dt > 0$ ) when  $P$  exceeds  $Q$  and (ii) decreasing  $Q$  ( $dQ/dt \leq 0$ ) otherwise. These conditions indicate limited influence of additional stores which are fed during  $P$  events and then sustain  $Q$  during dry periods. The percentages of days for which  $S_{dyn}$  can be thought as the main source of  $Q$  spanned between 54 and 72% across the catchments, meaning that the assumption underlying the method proposed by Kirchner (2009) is met for most of the time in all catchments.



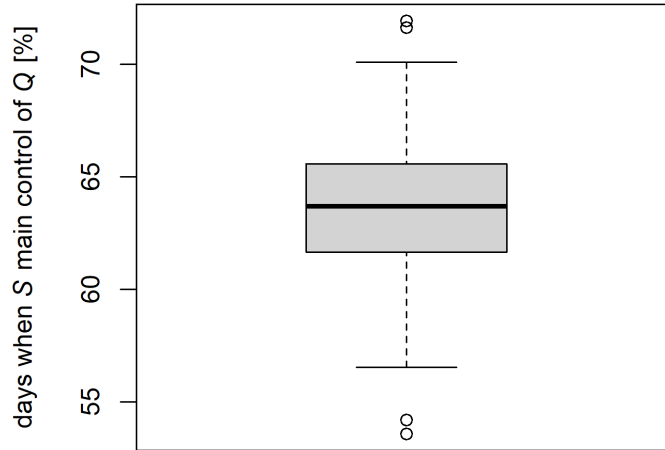


Fig. 3: Distribution of the percentage of days when storage (S) is the main control of streamflow (Q), across all catchments.

3. *So, how do you quantify the uncertainties in  $E$  you derived from water balance as I am not sure the uncertainties in  $S_{dyn}$ .*

Deriving  $E$  through a water balance approach may involve considerable uncertainties from the assumptions around potential changes in  $S$  and data themselves (see reply to comment #1 by Referee #1). Bruno & Duethmann (2024) extensively studied uncertainties in long-term variations in water balance-derived  $E$  for small catchments without substantial water management in Germany, by using the same approach and datasets as here. This work revealed that the main source of uncertainty is  $P$  rather than assumptions regarding  $S$  and long-term variations in  $S_{dyn}$  largely agreed with those from groundwater data. Bruno & Duethmann (2024) furthermore showed that long-term variations in  $E$  were generally robust in the study region with respect to alternative data sources. We realized that we did not discuss potential uncertainties in  $E$  in light of Bruno & Duethmann (2024) in the original version of the manuscript. In the revised version, we referred to this work in the Introduction already, by rephrasing L76–77 as follows, and discussed it in the new Sect. 4.4.

L76–77: Bruno & Duethmann (2024) reported robust increases in water balance-derived  $E$  in small catchments in Germany between the 1970s and 2000s, regardless of uncertainties in  $P$  data and in the consideration of potential  $S$  changes.

Sect. 4.4: Bruno and Duethmann (2024) moreover showed that long-term variations of water balance-derived  $E$  were generally robust to uncertainties in  $P$  and  $S$  for small catchments in Germany over the last five decades, and coherent with those from point-scale  $E$  data.

4. *Also, I have to say, for the catchments with areas ranging from 50-150km<sup>2</sup>, the lateral groundwater flow is significant which has been discussed in Ying Fan's paper 'Are catchments leaky?' and also quantified in our research (not published yet). Therefore, Equation 3 might be problematic.*

We agree that estimating  $E$  from  $P$  and  $Q$  observations implies considerable uncertainties for some catchments (Fan, 2019; Kampf et al., 2020; Safeeq et al., 2020). However, here we are mostly interested in long-term variations in  $E$  at a regional scale. For this, we argue that alternative approaches can be equally challenging, given issues in modelling long-term variations in  $E$  (Duethmann et al., 2020) and in the representativeness of satellite-derived products for small catchments. Thus, we chose this approach, in line with several previous works (e.g., Teuling et al., 2009; Ukkola et al.,

2013; Duethmann and Bloeschl, 2018; Bruno and Duethmann, 2024). Bruno and Duethmann (2024) furthermore compared trends in water balance-derived  $E$  and in  $E$  data from lysimeters and flux towers in Central Europe. Despite the general paucity of point-scale data, these showed similar temporal dynamics to water balance-derived  $E$ . This reinforces the suitability of the water balance-approach to study long-term variations in  $E$  in our study region. We acknowledge, however, that  $E$  estimates may be still uncertain for individual catchments and years, due to potential data issues and intercatchment groundwater flows. To tackle these uncertainties, we therefore performed the multiple linear regressions on cluster-, 5-year averages. To provide more context around our methodological choices, the associated uncertainties, and our strategies to minimize them, we rephrased L180–183 and added more discussion in the new Sect. 4.4 as follows.

L180–183: We adopted 5-year averages to focus on long-term dynamics and reduce potential uncertainties in water balance-derived  $E$ . Moreover, for the cluster-scale analysis we used average time series across the catchments in each cluster to minimize uncertainties in  $E$  for specific catchments, while analysing the main signal at a regional scale.

Sect. 4.4: Due to the generally coarse resolution of satellite-derived  $E$  estimates, we computed  $E$  from observed  $P$  and  $Q$ , and estimates of the  $S_{\text{dyn}}$  of the catchments from  $Q$  data as a first order approximation of  $S$ . This approach may be problematic for specific catchments, such as those with relevant intercatchment groundwater flows (Fan, 2019; Kampf et al., 2020; Safeeq et al., 2021). Yet, previous works often used a water balance-approach to study long-term variations in  $E$  (Teuling et al., 2009; Ukkola and Prentice, 2013; Duethmann and Blöschl, 2018; Bruno and Duethmann 2024). Bruno and Duethmann (2024) moreover showed that long-term variations of water balance-derived  $E$  were generally robust to uncertainties in  $P$  and  $S$  for small catchments in Germany over the last five decades, and coherent with those from point-scale  $E$  data. Here, we further aimed at minimizing uncertainties in  $E$  for specific catchments and years by using cluster-, 5-year averages in the multiple linear regressions for trend attribution (Sect. 2.5).

5. *Line 161,  $P_{\text{DJF}}$  and  $P_{\text{MAM}}$  are used as proxies of storage processes. How and why they are used as proxies?*

Long-term data on soil moisture and groundwater storage are unavailable for the large number of small catchments that we analyze here. Thus, we used  $P_{\text{DJF}}$  and  $P_{\text{MAM}}$  as proxies for storage recharge in the seasons preceding the summer, as frequently done (Duethmann et al., 2015; Saft et al., 2016; Laaha et al., 2017). Following comment #2 by Referee #1 too, we rephrased L174–176 as follows and added more discussion on this point in the new Sect. 4.4.

L174–176: We used  $P_{\text{MAM}}$  and  $P_{\text{DJF}}$  as proxies of storage recharge in the seasons preceding the summer similarly to what done by previous work to overcome the unavailability of long-term data on soil moisture and groundwater storage for the study catchments (Duethmann et al., 2015; Saft et al., 2016; Laaha et al., 2017).

Sect. 4.4: Finally, as potential predictors of changes in summer low flows we approximated storage processes with  $P$  in the season preceding the summer, due to unavailability of long-term  $S$  data for the study catchments. We chose this approach instead of using alternative proxies for  $S$  (e.g., estimates of  $S_{\text{dyn}}$  or baseflow from  $Q$  data) to avoid dependences between predictors and target variable (summer low flows). The satisfactory performances of the multiple linear regressions and the

plausible signs of their coefficients suggest the suitability of the selected predictors to represent the long-term dynamic of summer low flows (Table S1 and Fig. S6).

6. *All conclusions occur in less than 30% of the catchments, so how do you think about the generality of the study?*

The main findings of this study can be summarized as:

- The magnitude of summer low flows consistently decreased across 363 small catchments with no substantial water management in Germany over 1970–2019 (Fig. 4);
- Increases in  $E$  were a relevant driver of these decreases, especially for catchments in the Eastern area (Fig. 6 and 7);
- Changes in the  $P$ - $Q$  relationship occurred in catchments with underlying increases in  $E$  during a multi-year drought between 1989 and 1993 (Fig. 8).

With respect to first finding, most of the catchments experienced a tendency to decreases in summer low flows, with negative trends in 77 % of the catchments and an interquartile range of  $-7.5/-0.6$  % decade<sup>-1</sup> (white boxplot in Fig. 4c). Furthermore, significant negative trends occurred in 31 % of the catchments, while significant positive trends in the 2 % only (Fig. 4b). To better highlight the general decreasing tendency in summer low flows across the catchments, we complemented L20–21 and L225–227 as follows.

L20–21: Summer low flows decreased (increased) significantly in 31 % (2 %) of the catchments, with a median trend of  $-3.7$  % decade<sup>-1</sup> across all catchments.

L225–227: Trends in  $7dQ_{\min, JJA}$  were significantly negative in 31 % of the catchments and significantly positive in 2 % of them (negative in 77 % and positive in 23 %, Fig. 4b), with median (interquartile range, IQR) of  $-3.7$  ( $-7.5/-0.6$ ) % decade<sup>-1</sup> across all catchments (Fig. 4c).

Regarding the relevance of increases in  $E$  for decreases in summer low flows (second finding) and for changes in the  $P$ - $Q$  relationship during multi-year droughts (third finding), we acknowledge that additional processes were also important for decreases in summer low flows in specific clusters (Fig. 6 and 7) and hydrological changes occurred rather sparsely during the multi-year drought under study (see also reply to the comment #5 by Referee #1). However, we believe that our findings are relevant well beyond our study region, given the strong increases in  $E$  in many regions worldwide over recent decades (Duethmann and Blöschl, 2018; Teuling et al., 2019; Yang et al., 2023), and their generally limited consideration in low-flow and drought analyses. We highlighted better the relevance of our findings outside our study region in the revised version of the manuscript, by adding a paragraph in the Discussion, and rephrasing L456–457 and L470–471 as follows.

Discussion, L449–454: However, challenges remain in understanding the causes of decreases in summer low flows (Montanari et al., 2023) and in  $Q$  generation during multi-year droughts (Fowler, Peel, Saft, Peterson, et al., 2022). Here, we revealed that increases in  $E$  contributed (i) to decreases in summer low flows, in particular in catchments tending to arid, and (ii) to changes in the  $P$ - $Q$  relationship during the multi-year drought that occurred in Germany between 1989 and 1993. With strong increases in  $E$  reported for many regions of the world over recent decades (Duethmann and Blöschl, 2018; Teuling et al., 2019; Yang et al., 2023), these findings are relevant beyond our study region.

L456–457: (...) we underline the importance of monitoring changes in  $E$  for the prediction of potential decreases in  $Q$  during dry periods, particularly in arid regions.

L470–471: We illustrated the imprint of long-term increases in  $E$  on decreases in  $Q$  during dry conditions, which can be especially relevant for arid regions and prolonged droughts.

7. *I am wondering why  $ET$  is increasing? The land cover change or the temperature increasing? Authors listed a lot of attributes of the catchments in table 1 but are limited used in the analysis.*

In Bruno & Duethmann (2024), we provided a first assessment of potential causes of past changes in  $E$  in small catchments in Germany, by showing that these changes were consistent with those in  $P$  and solar radiation. However, an in-depth investigation on the causes of past increases in  $E$  in the study region, including changes in land cover, would be noteworthy. Yet, we see this analysis as out of the scope of the current study. The catchment attributes in Table 1 are static attributes, which we used for a characterization of the study catchments. Additional, dynamic attributes would be needed to relate changes in  $E$  with changes in land cover for instance. To provide some background on possible causes of increases in  $E$  to the readers, we added this point to the Introduction as follows.

L54–55: Over recent decades,  $E$  has increased in many regions following changes in climate and land cover (Duethmann and Blöschl, 2018; Teuling et al., 2019; Yang et al., 2023; Bruno & Duethmann, 2024).

8. *Have you cited the paper talking about the similar thing? Tran et al. (2023), Frontiers in Water.*

Thank you for pointing to this piece of literature which we added to the Introduction and Discussion.

L56–57: These increases in  $E$  contributed to decreases in annual  $Q$  on the long-term (Teuling et al., 2009; Fischer et al., 2023) and during dry years, especially in mountain catchments (Tran et al., 2023).

L359–360: Our findings on the role of increases in  $E$  on decreases in summer low flows expand results at annual time scale (Fischer et al., 2023; Tran et al., 2023)

## References:

- Alvarez-Garreton, C., Boisier, J. P., Garreaud, R., Seibert, J., & Vis, M. (2021). Progressive water deficits during multiyear droughts in basins with long hydrological memory in Chile. *Hydrology and Earth System Sciences*, 25(1), 429–446. <https://doi.org/10.5194/hess-25-429-2021>
- 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(9), 4317–4337. <https://doi.org/10.5194/hess-24-4317-2020>
- Berghuijs, W. R., Hartmann, A., & Woods, R. A. (2016). Streamflow sensitivity to water storage changes across Europe. *Geophysical Research Letters*, 43(5), 1980–1987. <https://doi.org/10.1002/2016GL067927>
- Boeing, F., Wagener, T., Marx, A., Rakovec, O., Kumar, R., Samaniego, L., & Attinger, S. (2024). Increasing influence of evapotranspiration on prolonged water storage recovery in Germany. *Environmental Research Letters*, 19(2), 024047. <https://doi.org/10.1088/1748-9326/ad24ce>
- Bruno, G., & Duethmann, D. (2024). Increases in Water Balance-Derived Catchment Evapotranspiration in Germany During 1970s–2000s Turning Into Decreases Over the Last Two Decades, Despite Uncertainties. *Geophysical Research Letters*, 51(6), e2023GL107753. <https://doi.org/10.1029/2023GL107753>

- Court, A. (1962). Measures of streamflow timing. *Journal of Geophysical Research*, 67(11), 4335–4339. <https://doi.org/10.1029/JZ067i011p04335>
- Duethmann, D., & Blöschl, G. (2018). Why has catchment evaporation increased in the past 40 years? A data-based study in Austria. *Hydrology and Earth System Sciences*, 22(10), 5143–5158. <https://doi.org/10.5194/hess-22-5143-2018>
- Duethmann, D., Blöschl, G., & Parajka, J. (2020). Why does a conceptual hydrological model fail to correctly predict discharge changes in response to climate change? *Hydrology and Earth System Sciences*, 24(7), Article 7. <https://doi.org/10.5194/hess-24-3493-2020>
- Duethmann, D., Bolch, T., Farinotti, D., Kriegel, D., Vorogushyn, S., Merz, B., Pieczonka, T., Jiang, T., Su, B., & Güntner, A. (2015). Attribution of streamflow trends in snow and glacier melt-dominated catchments of the T arim R iver, Central A sia. *Water Resources Research*, 51(6), 4727–4750. <https://doi.org/10.1002/2014WR016716>
- Fan, Y. (2019). Are catchments leaky? *WIREs Water*, 6(6), e1386. <https://doi.org/10.1002/wat2.1386>
- Fischer, M., Pavlík, P., Vizina, A., Bernsteinová, J., Parajka, J., Anderson, M., Řehoř, J., Ivančicová, J., Štěpánek, P., Balek, J., Hain, C., Tachecí, P., Hanel, M., Lukeš, P., Bláhová, M., Dlabal, J., Zahradníček, P., Máca, P., Komma, J., ... Trnka, M. (2023). Attributing the drivers of runoff decline in the Thaya river basin. *Journal of Hydrology: Regional Studies*, 48, 101436. <https://doi.org/10.1016/j.ejrh.2023.101436>
- Fowler, K., Peel, M., Saft, M., Nathan, R., Horne, A., Wilby, R., McCutcheon, C., & Peterson, T. (2022). Hydrological Shifts Threaten Water Resources. *Water Resources Research*, 58(8), Article 8. <https://doi.org/10.1029/2021WR031210>
- Fowler, K., Peel, M., Saft, M., Peterson, T. J., Western, A., Band, L., Petheram, C., Dharmadi, S., Tan, K. S., Zhang, L., Lane, P., Kiem, A., Marshall, L., Griebel, A., Medlyn, B. E., Ryu, D., Bonotto, G., Wasko, C., Ukkola, A., ... Nathan, R. (2022). Explaining changes in rainfall–runoff relationships during and after Australia’s Millennium Drought: A community perspective. *Hydrology and Earth System Sciences*, 26(23), 6073–6120. <https://doi.org/10.5194/hess-26-6073-2022>
- Han, J., Liu, Z., Woods, R., McVicar, T. R., Yang, D., Wang, T., Hou, Y., Guo, Y., Li, C., & Yang, Y. (2024). Streamflow seasonality in a snow-dwindling world. *Nature*, 629(8014), 1075–1081. <https://doi.org/10.1038/s41586-024-07299-y>
- Kampf, S. K., Burges, S. J., Hammond, J. C., Bhaskar, A., Covino, T. P., Eurich, A., Harrison, H., Lefsky, M., Martin, C., McGrath, D., Puntenney-Desmond, K., & Willi, K. (2020). The Case for an Open Water Balance: Re-envisioning Network Design and Data Analysis for a Complex, Uncertain World. *Water Resources Research*, 56(6), e2019WR026699. <https://doi.org/10.1029/2019WR026699>
- Kirchner, J. W. (2009). Catchments as simple dynamical systems: Catchment characterization, rainfall–runoff modeling, and doing hydrology backward: CATCHMENTS AS SIMPLE DYNAMICAL SYSTEMS. *Water Resources Research*, 45(2). <https://doi.org/10.1029/2008WR006912>
- Laaha, G., Gauster, T., Tallaksen, L. M., Vidal, J.-P., Stahl, K., Prudhomme, C., Heudorfer, B., Vlnas, R., Ionita, M., Van Lanen, H. A. J., Adler, M.-J., Caillouet, L., Delus, C., Fendekova, M., Gailliez, S., Hannaford, J., Kingston, D., Van Loon, A. F., Mediero, L., ... Wong, W. K. (2017). The European 2015 drought from a hydrological perspective. *Hydrology and Earth System Sciences*, 21(6), 3001–3024. <https://doi.org/10.5194/hess-21-3001-2017>
- Massari, C., Avanzi, F., Bruno, G., Gabellani, S., Penna, D., & Camici, S. (2022). Evaporation enhancement drives the European water-budget deficit during multi-year droughts. *Hydrology and Earth System Sciences*, 26(6), 1527–1543. <https://doi.org/10.5194/hess-26-1527-2022>
- Montanari, A., Nguyen, H., Rubinetti, S., Ceola, S., Galelli, S., Rubino, A., & Zanchettin, D. (2023). Why the 2022 Po River drought is the worst in the past two centuries. *Science Advances*, 9(32), eadg8304. <https://doi.org/10.1126/sciadv.adg8304>

- Renner, M., & Hauffe, C. (2024). Impacts of climate and land surface change on catchment evapotranspiration and runoff from 1951 to 2020 in Saxony, Germany. *Hydrology and Earth System Sciences*, 28(13), 2849–2869. <https://doi.org/10.5194/hess-28-2849-2024>
- Richter, D. (1995). *Ergebnisse methodischer Untersuchungen zur Korrektur des systematischen Messfehlers des Hellmann- Niederschlagsmessers*. Offenbach.
- Safeeq, M., Bart, R. R., Pelak, N. F., Singh, C. K., Dralle, D. N., Hartsough, P., & Wagenbrenner, J. W. (2021). How realistic are water-balance closure assumptions? A demonstration from the southern sierra critical zone observatory and kings river experimental watersheds. *Hydrological Processes*, 35(5), e14199. <https://doi.org/10.1002/hyp.14199>
- 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. *Water Resources Research*, 52(12), 9290–9305. <https://doi.org/10.1002/2016WR019525>
- Saft, M., Western, A. W., Zhang, L., Peel, M. C., & Potter, N. J. (2015). The influence of multiyear drought on the annual rainfall-runoff relationship: An Australian perspective. *Water Resources Research*, 51(4), 2444–2463. <https://doi.org/10.1002/2014WR015348>
- Staudinger, M., Stoelzle, M., Seeger, S., Seibert, J., Weiler, M., & Stahl, K. (2017). Catchment water storage variation with elevation. *Hydrological Processes*, 31(11), 2000–2015. <https://doi.org/10.1002/hyp.11158>
- Stoelzle, M., Stahl, K., & Weiler, M. (2013). Are streamflow recession characteristics really characteristic? *Hydrology and Earth System Sciences*, 17(2), 817–828. <https://doi.org/10.5194/hess-17-817-2013>
- Teuling, A. J., De Badts, E. A. G., Jansen, F. A., Fuchs, R., Buitink, J., Hoek Van Dijke, A. J., & Sterling, S. M. (2019). Climate change, reforestation/afforestation, and urbanization impacts on evapotranspiration and streamflow in Europe. *Hydrology and Earth System Sciences*, 23(9), 3631–3652. <https://doi.org/10.5194/hess-23-3631-2019>
- Teuling, A. J., Hirschi, M., Ohmura, A., Wild, M., Reichstein, M., Ciais, P., Buchmann, N., Ammann, C., Montagnani, L., Richardson, A. D., Wohlfahrt, G., & Seneviratne, S. I. (2009). A regional perspective on trends in continental evaporation: EVAPORATION TRENDS. *Geophysical Research Letters*, 36(2), n/a-n/a. <https://doi.org/10.1029/2008GL036584>
- Tran, H., Yang, C., Condon, L. E., & Maxwell, R. M. (2023). The Budyko shape parameter as a descriptive index for streamflow loss. *Frontiers in Water*, 5, 1258367. <https://doi.org/10.3389/frwa.2023.1258367>
- Trotter, L., Saft, M., Peel, M. C., & Fowler, K. J. A. (2024). Recession constants are non-stationary: Impacts of multi-annual drought on catchment recession behaviour and storage dynamics. *Journal of Hydrology*, 630, 130707. <https://doi.org/10.1016/j.jhydrol.2024.130707>
- Ukkola, A. M., & Prentice, I. C. (2013). A worldwide analysis of trends in water-balance evapotranspiration. *Hydrology and Earth System Sciences*, 17(10), 4177–4187. <https://doi.org/10.5194/hess-17-4177-2013>
- Yang, Y., Roderick, M. L., Guo, H., Miralles, D. G., Zhang, L., Fatichi, S., Luo, X., Zhang, Y., McVicar, T. R., Tu, Z., Keenan, T. F., Fisher, J. B., Gan, R., Zhang, X., Piao, S., Zhang, B., & Yang, D. (2023). Evapotranspiration on a greening Earth. *Nature Reviews Earth & Environment*, 4(9), 626–641. <https://doi.org/10.1038/s43017-023-00464-3>