Dear Referee,

we thank you for your time in reviewing our manuscript and your constructive feedback. We see feasible implementing all your suggestions in a revised version of the manuscript. In particular, we have now performed the multiple linear regression analysis also at the catchment scale, which we plan to introduce in a revised version of the manuscript. We also intend discussing more the reasons for our methodological choices and potential uncertainties in the study.

Please find below our point-by-point reply to your comments (italic, numbered) and the changes we propose to do in the manuscript to address them (underlined).

Best regards,

Giulia Bruno and co-authors

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 in the manuscript). 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 do not discuss in the manuscript at the moment. We propose to modify L134–137 to make clearer our assumption regarding the correction procedure and to add some discussion concerning the uncertainty in *P* data in a new section (4.4 Sources of uncertainty).

L134–137: We corrected the dataset for gauge undercatch following the method proposed for Germany by Richter (1995):

$$P_{corr} = P_{uncorr} + aP_{uncorr}^{\ b} \tag{2}$$

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).

4.4 Sources of uncertainty: 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_{DJF}) and spring precipitation (P_{MAM}) as wetness conditions in the seasons preceding the dry period in the study region 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 propose to rephrase L161–162 as follows and to discuss this point in the new section 4.4.

L161–162: Since long-term data on soil moisture and groundwater storage are not available for the study catchments, we used P_{MAM} and P_{DJF} as proxies of storage recharge in the seasons preceding the dry one (Duethmann et al., 2015; Saft et al., 2016; Laaha et al., 2017).

4.4 Sources of uncertainty**:** Finally, as potential predictors of changes in summer low flows we approximated storage processes with P in the season preceding the dry one, 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_{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).

3. *For multiple linear regression in predicting summer low flows, the authors showed the R2 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 have now 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_{JJA}) was the most recurrent one across all clusters (Fig. 1c). P_{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_{JJA} in the Pre-Alpine cluster, P_{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_{JJA} dominant predictor in the Pre-Alpine cluster, P_{MAM} significant predictor in the South-Central one, and E in the Eastern and Northern clusters, Fig. 6 in the manuscript). To reinforce our trend attribution, we intend to add this analysis in a revised version of the manuscript, by introducing Fig. 1 as a new Fig. S6 in the Supplement, and adapting the description of methods and results as follows.

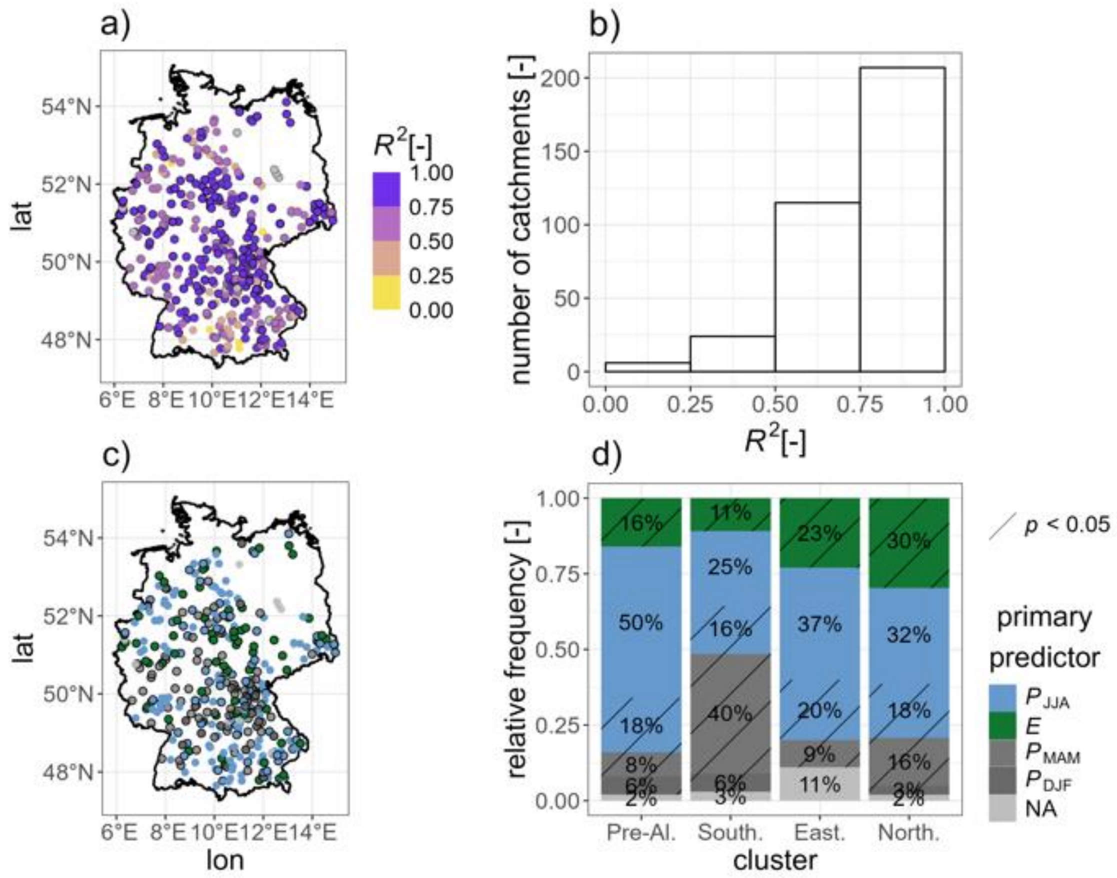

Fig. 1: Model results from the multiple linear regressions on catchment-scale data. (a) Map of coefficients of determination of the models (R^2). (b) Histogram of R^2. (c) Map of primary predictors (black edges if significant). (d) Relative frequency of primary predictors by cluster. Light grey in (a, 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: Firstly, we modelled the temporal dynamics of summer low flows ($7dQ_{min,\ JJA}$ in Eq.(4) both at a catchment- and cluster-scale from the dynamics of the predictors through multiple linear regression:

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

with α_i (i = 1…4) the regression coefficient for each predictor and ε 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: 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).

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 during the wet season. 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 in the manuscript). Thus, increases in P_{DJF} may have indirectly contributed to decreases in summer low flows in the Eastern cluster. We propose to expand the mechanistic explanation on this point as follows.

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 *E*-storage feedbacks (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 propose to rephrase L360–365 as follows.

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-Alphine 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 propose to add the following summary table in Section 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 ± 2 % decade^{-1} 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}

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

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

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., 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 meltdominated catchments of the Tarim River, Central Asia. *Water Resources Research*, *51*(6), 4727–4750. https://doi.org/10.1002/2014WR016716

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

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

Richter, D. (1995). *Ergebnisse methodischer Untersuchungen zur Korrektur des systematischen Messfehlers des Hellmann- Niederschlagsmessers*. *Offenbach*.

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