

Responses

Editor:

Thank you for your responses to the two referees. After considering both reports and conducting my own assessment, I have decided that the manuscript requires major revisions. Please address the following points in a consolidated revision:

We appreciate the decision and are grateful to you and the two anonymous referees for the insightful and constructive comments. Accordingly, we have conducted a thorough revision of the whole manuscript.

Below please find the point-to-point responses.

1. The physical interpretation of the 0.51 coefficient requires further support. Annual precipitation and streamflow RMSE may both increase during wet or extreme years. Using normalized errors or appropriate controls for hydroclimatic magnitude, please demonstrate that the reported relationship is not primarily attributable to this common dependence.

Thank you for the instructive comment. In the revision, we have clarified the interpretation of the 0.51 coefficient and have added the sensitivity analysis using normalized precipitation and streamflow errors:

“For GloFAS-ERA5 v4.0, the panel regression estimates an average increase of 0.51 mm in streamflow RMSE for each 1 mm increase in precipitation RMSE across the 671 catchments.” (Page 1, Lines 17 to 19)

“The coefficient of precipitation error for the random-effects panel regression is 0.51, indicating that a 1 mm increase in precipitation RMSE is associated with an average increase of 0.51 mm in streamflow RMSE. It is noted that the value 0.51 is the average statistical response estimated from the panel regression for 671 catchments, rather than a direct quantitative estimate of any single hydrological process.” (Page 11, Lines 223 to 227)

“To assess whether the positive relationships are primarily driven by the shared dependence on hydroclimatic magnitude, the sensitivity analysis is performed using normalized RMSE (Text S2 in the Supplement). The regression coefficients remain positive and statistically significant in the panel regressions and in linear regressions for 435 (65% of 671) catchments, as shown in Figs. S3 and S4 in the Supplement, indicating that these relationships persist after reducing the influence of hydroclimatic

magnitude.”(Page 21, Lines 358 to 362)

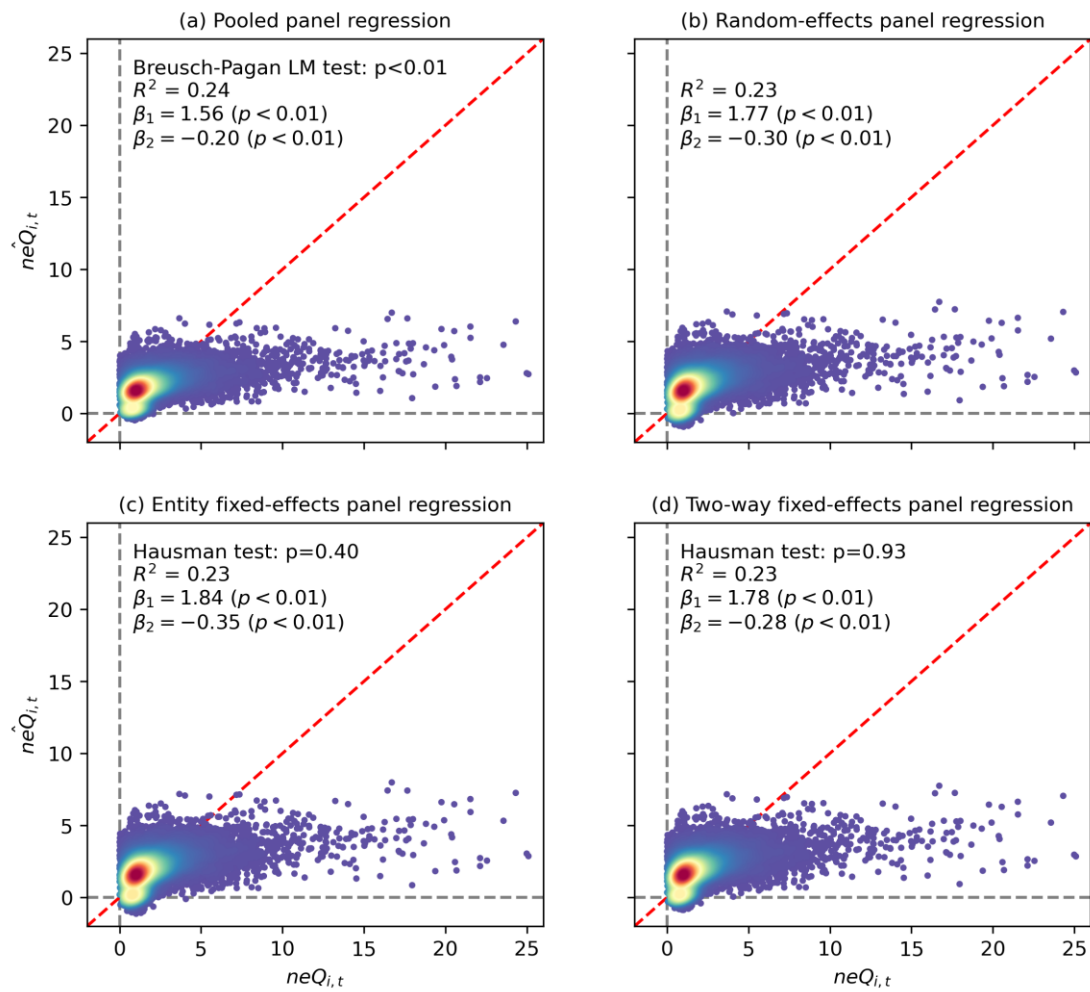


Figure S3. As for Fig. 3, but for normalized root-mean-square error (NRMSE).

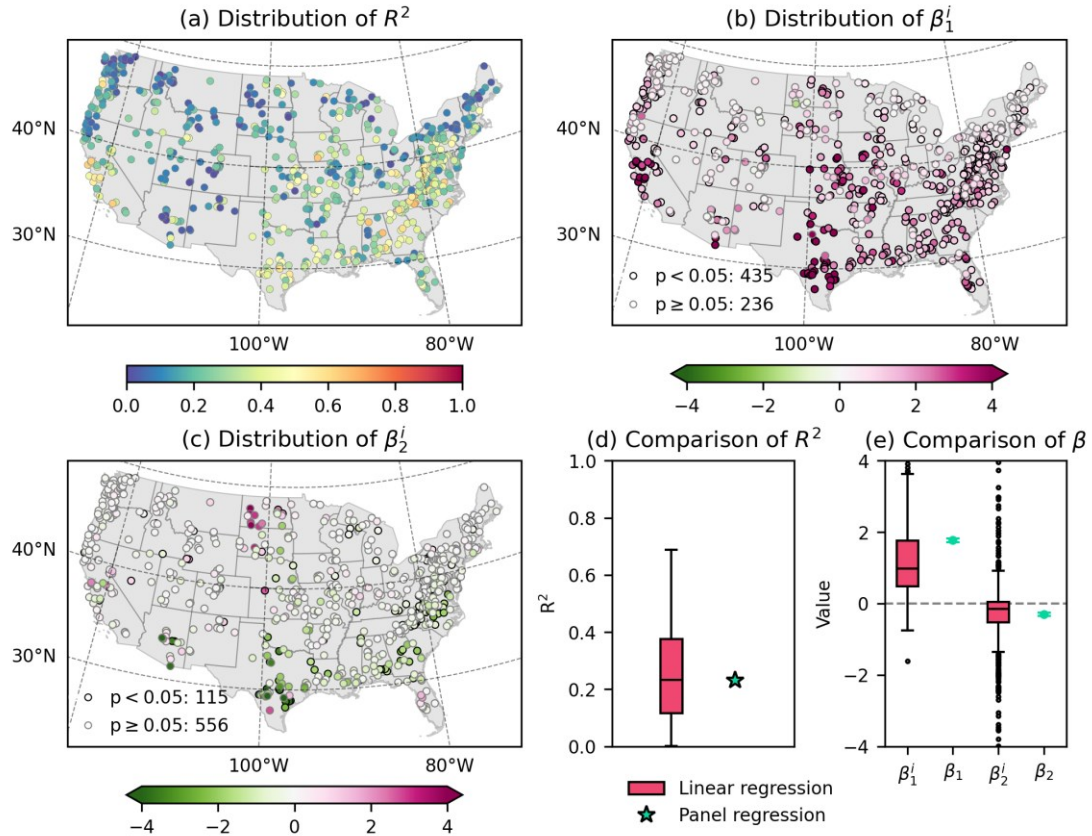


Figure S4. As for Fig. 4, but for normalized root-mean-square error (NRMSE).

2. Please discuss the water-balance implications and unexplained variance more thoroughly, clarify the coupling between precipitation and temperature errors in snow-dominated catchments, acknowledge the limitations of annual RMSE, and avoid causal language not supported by the analysis.

Thank you very much for the insightful comments. Accordingly, we have added additional analysis in the Supplement and have rewritten the Discussion section:

“5.1 Hydrological interpretation of divergent responses

Soil moisture deficits, evapotranspiration and subsurface water storage can limit runoff generation from precipitation in arid catchments (Farmani et al., 2025; Assouline et al., 2024; Cui and Tian, 2026). As shown in Fig. S1 in the Supplement, the correlation between streamflow reanalysis error and soil wetness index (SWI) is statistically significant in 473 catchments, with the median correlation coefficient of 0.52. This result is consistent with the relatively weaker responses of streamflow reanalysis errors to precipitation errors observed in arid catchments and stronger responses in some humid catchments. Dry antecedent soil conditions generally allow a large proportion of precipitation to infiltrate and replenish soil water storage before reaching the thresholds

to generate runoff (Liu et al., 2025; Tian and Xu, 2024). As antecedent soil moisture increases, the remaining storage deficit decreases, thereby enhancing the sensitivity of runoff generation to precipitation inputs (Ye et al., 2023; Ran et al., 2022). Evapotranspiration and groundwater recharge may further modulate the contemporaneous relationship between precipitation and streamflow errors (Berghuijs et al., 2025). Since evapotranspiration and groundwater data are not provided in GloFAS-ERA5 and Daymet datasets (Grimaldi et al., 2023; Thornton et al., 2016), their contributions and the specific distribution of precipitation errors among them cannot be determined.

In high-latitude or mountainous catchments, the seasonal snow storage and subsequent melt can alter the timing and magnitude of the streamflow responses to precipitation (Zhu et al., 2023; Wang et al., 2021; Nan et al., 2026). Precipitation errors in snow season may be retained in snow storage and appear months later as streamflow errors (Wang et al., 2021; Hale et al., 2023). As shown in Fig. S2 in the Supplement, the correlation between precipitation RMSE and temperature RMSE over cold days (Daymet temperature ≤ 0 °C) for each hydrological year is statistically significant in 56 catchments but not significant in 275 catchments. Temperature influences precipitation phase, snow accumulation and melt timing, but these physical linkages do not necessarily imply a strong statistical coupling between precipitation errors and temperature errors. The errors calculated over hydrological year partly integrate intra-annual effects of storage and lag (Berghuijs et al., 2025; de Lavenne et al., 2022), but they obscure seasonal contrasts, snow accumulation–melt transitions and intra-seasonal phase errors (Hale et al., 2023; de Lavenne et al., 2022). The annual errors therefore cannot resolve seasonal or event-scale errors, particularly for extreme precipitation and flood events. In addition, as the Daymet dataset is generated by interpolating observations (Thornton et al., 2016), uncertainties associated with station density and interpolation methods may affect the precipitation errors.” (Page 20, Lines 325 to 349)

“The precipitation errors can explain only a limited proportion of the variance in GloFAS-ERA5 streamflow errors, with the R^2 of the panel regression not exceeding 0.36. The remaining variance may be related to catchment heterogeneity, hydrological models and uncertainties in the reference datasets.” (Page 21, Lines 363 to 365)

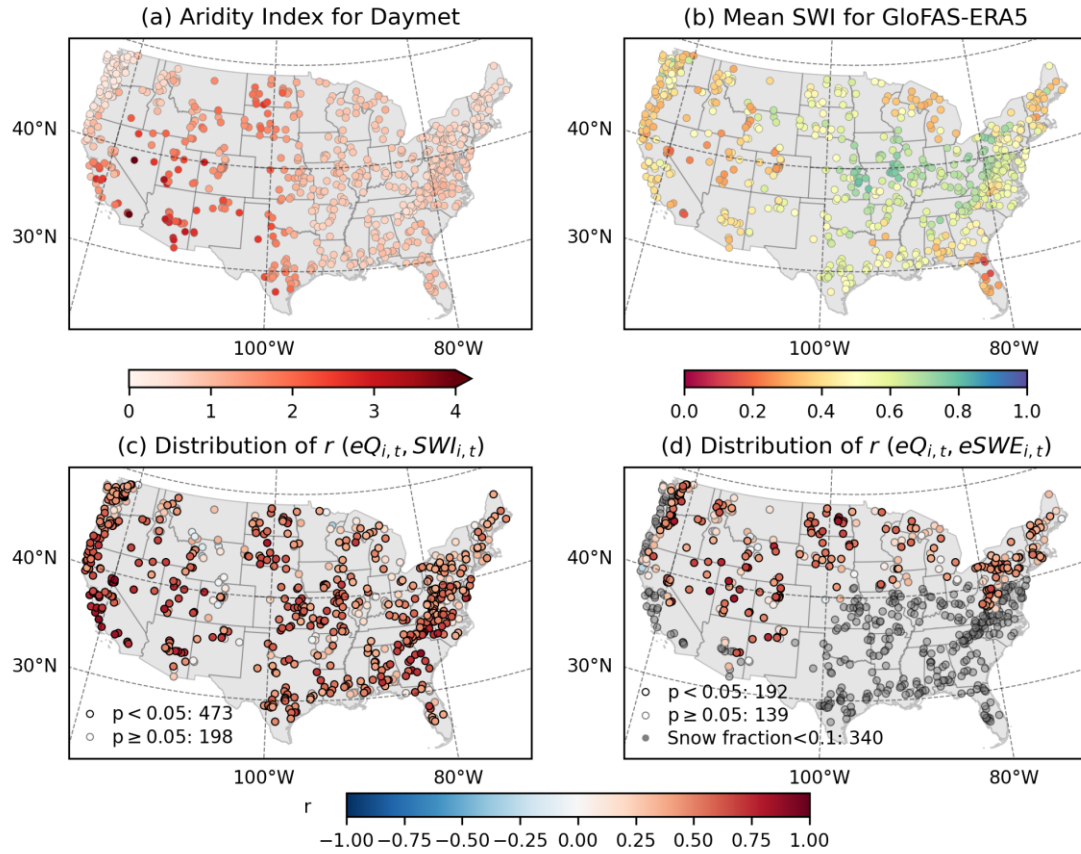


Figure S1. Spatial distribution of (a) aridity index for Daymet, (b) mean soil wetness index (SWI) from the GloFAS-ERA5 dataset and (c-d) Pearson's correlation coefficients of streamflow reanalysis error to SWI and to snow depth water equivalent (SWE) error. The eSWE represents the RMSE of SWE from the GloFAS-ERA5 dataset by hydrological year using the SWE from the Daymet dataset as reference data.

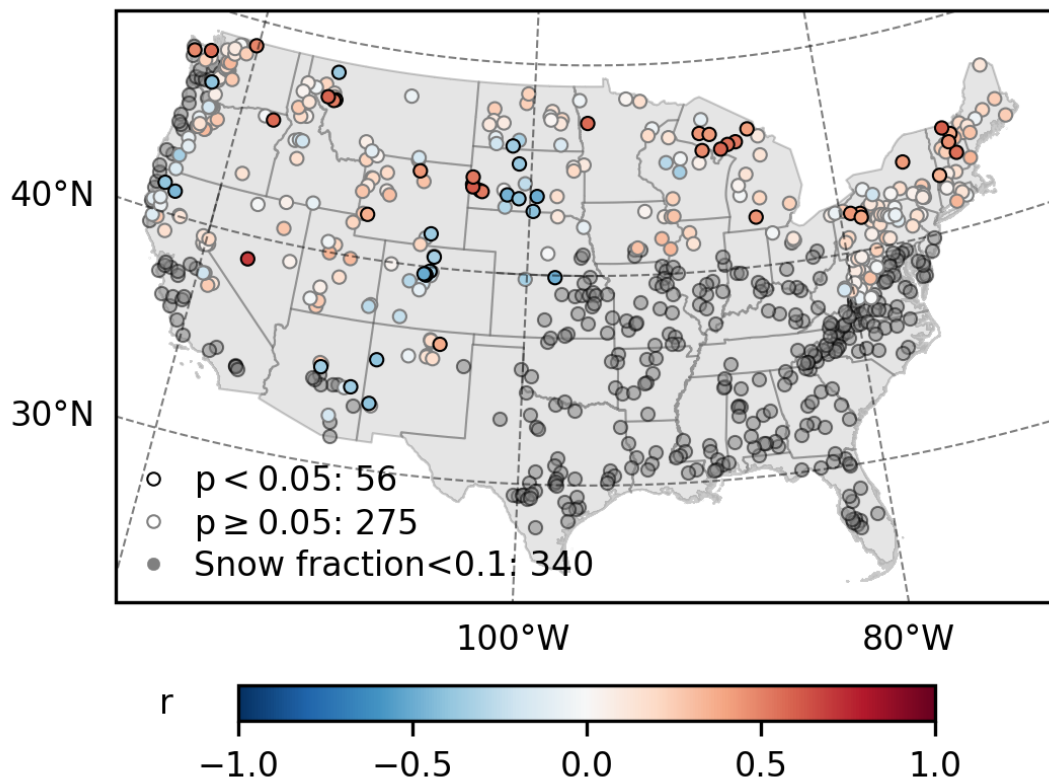


Figure S2. Pearson’s correlation coefficients (r) between precipitation reanalysis errors and temperature reanalysis errors on cold days (Daymet temperature $\leq 0^{\circ}\text{C}$) by hydrological year.

In addition, we have checked throughout this paper and revised the overly strong statements:

“These divergent responses suggest that streamflow errors may be more directly associated with precipitation errors under the saturation-excess conditions while soil moisture deficits can dampen their effects.” (Page 1, Lines 20 to 22)

“The coefficient of precipitation error for the random-effects panel regression is 0.51, indicating that a 1 mm increase in precipitation RMSE is associated with an average increase of 0.51 mm in streamflow RMSE.” (Page 11, Lines 253 to 255)

“This case study indicates that in humid rain-dominated catchments, the performance of streamflow reanalysis is predominantly associated with the accuracy of precipitation reanalysis rather than temperature.” (Page 17, Lines 299 and 300)

3. Please justify the use of the random-effects model given the reported Hausman-test

result ($p=0.03$) and clarify the interpretation of the Breusch–Pagan LM test.

Thank you for the instructive comment. In the revision, the use of the random-effects model and the interpretation of the Breusch-Pagan LM test have been clarified in Section 4.2:

“The Breusch-Pagan Lagrange multiplier test rejects the null hypothesis that residuals of the pooled panel regression are homogenous ($p<0.01$), indicating significant unobserved heterogeneity across catchments. Since the pooled panel regression cannot account for the unobserved heterogeneity, it is inadequate for capturing the average effects here (Steinschneider et al., 2013). Furthermore, the Hausman test cannot reject the null hypothesis that the unobserved catchment-specific effects are uncorrelated with the explanatory variables at the 1% significance level ($p=0.03$ and $p=0.90$), indicating that the random-effects panel regression is appropriate. The random-effects model not only requires fewer parameters but also can accommodate time invariant predictors, such as catchment attributes, whereas these effects are absorbed and cannot be estimated in fixed-effects models (Steinschneider et al., 2013; Bassiouni et al., 2016). Consequently, the random-effects panel regression is selected for subsequent analyses.” (Page 11, Lines 215 to 223)

4. Please explain how the two case-study catchments and the displayed years were selected, and provide access to the derived data and analysis code where possible.

Thank you very much for the insightful comments. We have added the selection criteria of the case-study catchments in the Supplement:

“Text S1: Selection of case-study catchments

The case-study catchments are selected according to criteria that account for reanalysis performance, observational data availability and hydroclimatic diversity. Specifically, candidate catchments are required to have Nash–Sutcliffe efficiency (NSE) greater than 0.5 for the GloFAS-ERA5 streamflow reanalysis relative to observations (Michailovsky et al., 2023; Hwang and Devineni, 2022; Li et al., 2025). This criterion ensures that the reanalysis captures streamflow dynamics with reasonable skill, thereby reducing the likelihood that the diagnosed responses are driven primarily by poor streamflow simulation (Jiang et al., 2022; Mai et al., 2022). To further limit the influence of data gaps on the estimation of annual reanalysis errors and regression coefficients, only catchments with less than 5% missing streamflow observations are retained.

Among catchments satisfying these general criteria, the rain-dominated case study is selected from those with the snow fraction below 0.1 and statistically significant

regression coefficient for precipitation reanalysis error ($p < 0.05$). These criteria are used to identify catchments in which variations in streamflow reanalysis error are primarily associated with precipitation errors, with minimal influence from snow-related processes. By contrast, the snow-dominated case study is selected from catchments with the snow fraction greater than 0.3 and statistically significant regression coefficients for both precipitation and temperature reanalysis errors ($p < 0.05$). These criteria are used to identify catchments in which variations in streamflow reanalysis error reflect the combined influence of precipitation errors and snow processes related to temperature. The two selected catchments therefore provide contrasting examples for presenting the divergent responses under rain-dominated and snow-dominated hydrological regimes.” (Page 1 in the Supplement)

In the revision, how the displayed years are determined has been added in Section 4.4:

“The three years are selected for visualization because they contain contrasting years with relatively high and low precipitation errors, allowing the responses of streamflow errors to be compared clearly.” (Page 17, Lines 291 and 292)

“The displayed years are selected to include years with comparable precipitation RMSE but contrasting temperature RMSE, thereby showing the potential role of snowmelt errors related to temperature.” (Page 18, Lines 309 to 311)

In addition, the code to perform the experiments and analysis has been uploaded to Zenodo and the access link has been added in the Code and data availability:

“The code to perform the experiments and analysis is available from Zenodo (<https://zenodo.org/records/22120405>).” (Page 22, Line 403)

Please also implement the remaining clarifications and corrections requested by the referees. Submit a revised manuscript accompanied by a detailed point-by-point response. The revised manuscript will be reconsidered by the editor and referees. Thank you again for your contribution to HESS.

Thank you very much for the kind reminder. According to the insightful and constructive comments, we have made substantial efforts to improve the whole paper. Below please find the point-by-point responses to the referees’ comments.

Anonymous Referee #1:

This paper precisely addresses a long-standing issue that has been ambiguously treated in the field of global hydrological reanalysis, namely, whether precipitation input errors are amplified or attenuated during the runoff concentration process. It does not merely provide a global average number; rather, for the first time, it systematically quantifies the distinctly different patterns of this error propagation in humid, arid, and snow-covered regions. Using two specific case studies, which include a rain-fed basin and a snow-fed basin, the paper interprets statistical regression coefficients into two physical processes, namely saturation-excess runoff and snowmelt delay, making the conclusions highly convincing.

We are grateful to you for the positive comments.

Nevertheless, the paper still has the following areas that require improvement:

Thank you very much for the insightful comments. Accordingly, we have thoroughly revised this paper. Below please find the point-to-point responses.

(1) Lines 18-19: The physical interpretation of the buffering coefficient of 0.51 is somewhat thin. Although it is mentioned as "the buffering capacity of catchment storage," the storage capacities of the 671 basins (e.g., baseflow index, groundwater recharge rate) vary considerably. The paper does not further cross-validate the average coefficient of 0.51 with basin-specific storage capacity curves or soil moisture memory. Does this 0.51 predominantly reflect soil infiltration, or is it dominated by evapotranspiration (ET) consumption? The current explanation is somewhat vague.

Thank you for the valuable comment. The value 0.51 is a cross-catchment average statistical response estimated from the panel regression for 671 catchments rather than a direct quantitative estimate of any single hydrological process. The catchment-specific coefficients range from near zero to approximately 2.5, showing that the average cannot represent each catchment. In the revision, we have deleted “the buffering capacity of catchment storage” from the Abstract and added the explicit explanation in the Results:

“For GloFAS-ERA5 v4.0, the panel regression estimates an average increase of 0.51 mm in streamflow RMSE for each 1 mm increase in precipitation RMSE across the 671 catchments.” (Page 1, Lines 17 to 19)

“The coefficient of precipitation error for the random-effects panel regression is 0.51,

indicating that a 1 mm increase in precipitation RMSE is associated with an average increase of 0.51 mm in streamflow RMSE. It is noted that the value 0.51 is the average statistical response estimated from the panel regression for 671 catchments, rather than a direct quantitative estimate of any single hydrological process.” (Page 11, Lines 223 to 227)

(2) The paper does not address the issue of water balance closure. The explanation that the error in arid regions is <0.7 is attributed to soil moisture deficit buffering, but this implies a premise—that the runoff error caused by precipitation error is absorbed. Where does the absorbed error go? Does this imply that the evapotranspiration (ET) error in arid regions is amplified?

Thank you for the instructive comment. The weaker responses of streamflow reanalysis errors in arid catchments do not mean that precipitation errors disappear or are fully absorbed. Such errors may appear in evapotranspiration, soil water, groundwater, snow storage or combinations of these components. Since the available variables in the GloFAS-ERA5 and Daymet datasets only support analysis of soil water and snow storage, we have added additional analysis in the Supplement and have stated this limitation explicitly in the Discussion:

“Soil moisture deficits, evapotranspiration and subsurface water storage can limit runoff generation from precipitation in arid catchments (Farmani et al., 2025; Assouline et al., 2024; Cui and Tian, 2026). As shown in Fig. S1 in the Supplement, the correlation between streamflow reanalysis error and soil wetness index (SWI) is statistically significant in 473 catchments, with the median correlation coefficient of 0.52. This result is consistent with the relatively weaker responses of streamflow reanalysis errors to precipitation errors observed in arid catchments and stronger responses in some humid catchments. Dry antecedent soil conditions generally allow a large proportion of precipitation to infiltrate and replenish soil water storage before reaching the thresholds to generate runoff (Liu et al., 2025; Tian and Xu, 2024). As antecedent soil moisture increases, the remaining storage deficit decreases, thereby enhancing the sensitivity of runoff generation to precipitation inputs (Ye et al., 2023; Ran et al., 2022). Evapotranspiration and groundwater recharge may further modulate the contemporaneous relationship between precipitation and streamflow errors (Berghuijs et al., 2025). Since evapotranspiration and groundwater data are not provided in GloFAS-ERA5 and Daymet datasets (Grimaldi et al., 2023; Thornton et al., 2016), their contributions and the specific distribution of precipitation errors among them cannot be determined.” (Page 20, Lines 326 to 337)

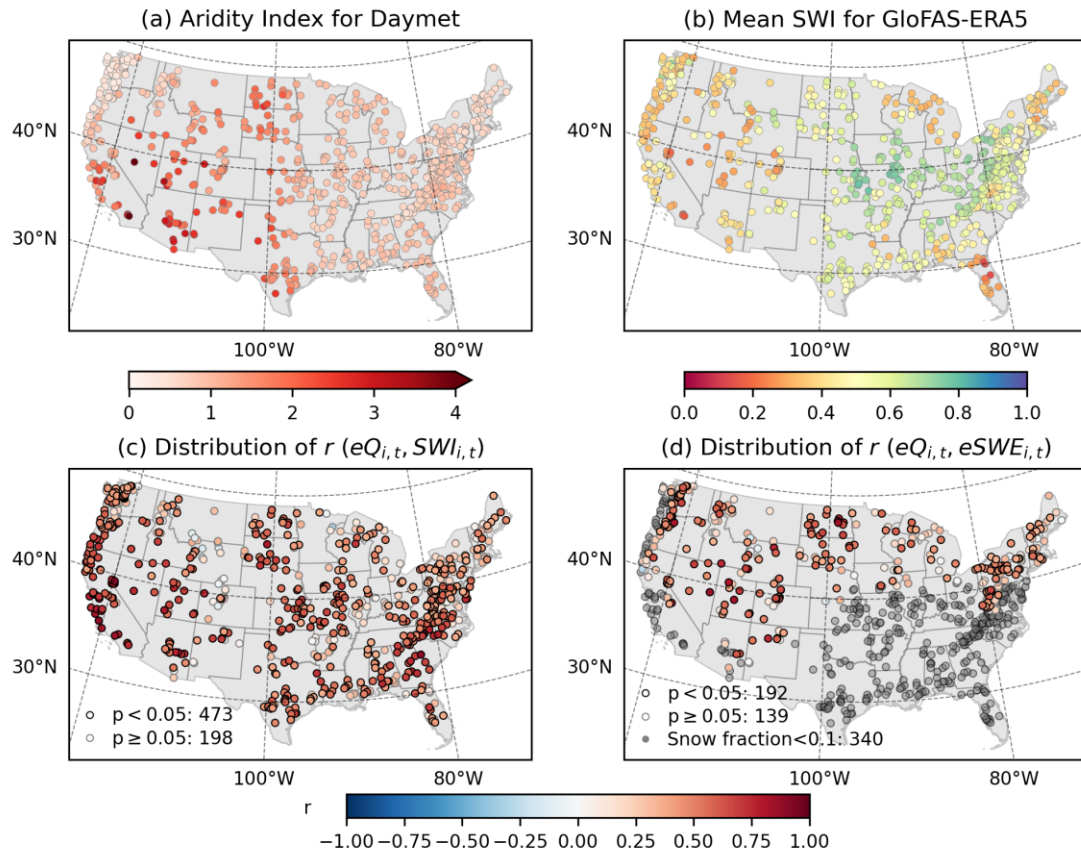


Figure S1. Spatial distribution of (a) aridity index for Daymet, (b) mean soil wetness index (SWI) from the GloFAS-ERA5 dataset and (c-d) Pearson's correlation coefficients of streamflow reanalysis error to SWI and to snow depth water equivalent (SWE) error. The eSWE represents the RMSE of SWE from the GloFAS-ERA5 dataset by hydrological year using the SWE from the Daymet dataset as reference data.

(3) Although the paper states that $VIF < 5$, indicating weak collinearity between precipitation and temperature, in snow seasons, higher precipitation is often accompanied by lower temperatures, and the actual impacts of the two are highly temporally coupled. While the panel regression passes statistical tests, the physical simultaneity (i.e., winter low temperatures cause snowfall, and the quality of both low-temperature and precipitation data often deteriorates simultaneously) has not been sufficiently disentangled and discussed.

Thank you for the insightful comment. The VIF values below 5 indicate only that the annual RMSE predictors do not exhibit severe linear multicollinearity. They do not demonstrate that the physical effects of precipitation and temperature are separated. In

the revision, we have revised the interpretation in the Section 4.1 and have added analysis of relationships between precipitation errors and temperature errors on the cold days in the Discussion:

“Their Variance Inflation Factor (VIF) values are consistently below 5 across all catchments, indicating that precipitation and temperature errors by hydrological year do not exhibit severe collinearity. Consequently, both variables can be utilized as predictors in the subsequent regression models.” (Page 10, Lines 201 to 203)

“In high-latitude or mountainous catchments, the seasonal snow storage and subsequent melt can alter the timing and magnitude of the streamflow responses to precipitation (Zhu et al., 2023; Wang et al., 2021; Nan et al., 2026). Precipitation errors in snow season may be retained in snow storage and appear months later as streamflow errors (Wang et al., 2021; Hale et al., 2023). As shown in Fig. S2 in the Supplement, the correlation between precipitation RMSE and temperature RMSE over cold days (Daymet temperature $\leq 0^{\circ}\text{C}$) for each hydrological year is statistically significant in 56 catchments but not significant in 275 catchments. Temperature influences precipitation phase, snow accumulation and melt timing, but these physical linkages do not necessarily imply a strong statistical coupling between precipitation errors and temperature errors.” (Page 20, Lines 338 to 344)

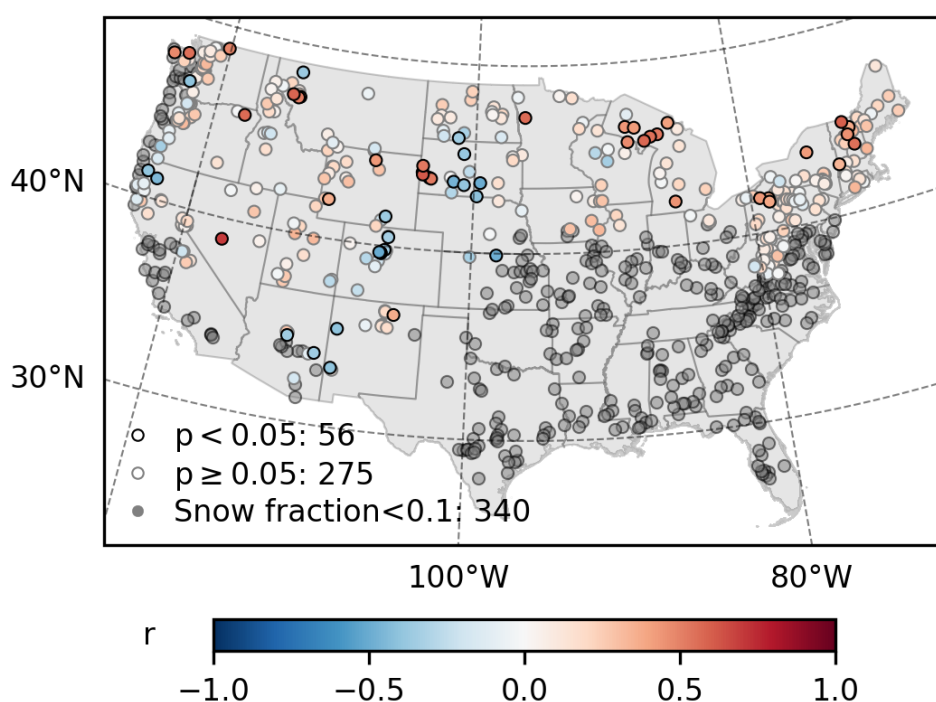


Figure S2. Pearson's correlation coefficients (r) between precipitation reanalysis errors and temperature reanalysis errors on cold days (Daymet temperature $\leq 0^{\circ}\text{C}$) by hydrological year.

(4) Logical confusion: Section 4.2 indicates that basin-specific regressions perform better than panel regression. Why then continue to stubbornly focus on improving the panel regression? Although the authors explain that panel regression provides a stable average, they do not explicitly answer: given the substantial individual differences (ranging from 0 to 2.5), does the forced use of a global panel regression (even with interaction terms) statistically obscure the extreme physical mechanisms of the extremes? The logic would be clearer if a sentence were added, such as: "Panel regression is primarily used to reveal cross-basin universal laws, rather than to precisely predict specific values for a given basin."

Thank you for the instructive comment. The two approaches address different questions: the catchment-specific linear regressions reveal local heterogeneity, whereas the panel regression estimates an average relationship using spatial and temporal information from all 671 catchments. We have added the purpose of panel regression in the Methods and have added its limitation in the Discussion:

“It is intended to identify cross-catchment average relationships rather than to provide precise predictions for individual catchments (Wang et al., 2026; Over et al., 2025).” (Page 6, Lines 127 and 128)

“The panel regression is designed to estimate the average relationship across catchments, but it can obscure catchment-specific responses or the behavior of extreme events (Wang et al., 2026; Steinschneider et al., 2013).” (Page 21, Lines 372 and 373)

(5) The limitation of the temporal scale is not discussed: all error calculations are based on annual RMSE. The annual-scale RMSE smooths out intra-seasonal phase errors.

Thank you for the insightful comment. In the revision, the limitation of annual RMSE has been added in the Discussion:

“The errors calculated over hydrological year partly integrate intra-annual effects of storage and lag (Berghuijs et al., 2025; de Lavenne et al., 2022), but they obscure seasonal contrasts, snow accumulation–melt transitions and intra-seasonal phase errors (Hale et al., 2023; de Lavenne et al., 2022). The annual errors therefore cannot resolve seasonal or event-scale errors, particularly for extreme precipitation and flood events.” (Page 20, Lines 344 to 348)

Anonymous Referee #2:

This manuscript investigates the responses of GloFAS-ERA5 streamflow reanalysis errors to ERA5 precipitation reanalysis errors and further explores the moderating effects of catchment climatic and physiographic heterogeneity. By combining catchment-specific linear regression, panel regression and case studies of representative catchments, the study analyzes 671 CAMELS catchments and provides insights into both overall relationships and catchment heterogeneity. The research topic is novel to some extent and the results have practical implications for understanding and applying global streamflow reanalysis products.

We appreciate the positive comments.

Several issues related to the physical interpretation of the results, regression modeling, and study limitations should be addressed.

Thank you very much for the insightful comments. Accordingly, we have thoroughly revised this paper. Below please find the point-to-point responses.

Major comments

1. The attenuation of precipitation errors by catchments may involve multiple hydrological processes, including soil infiltration, soil water storage, groundwater recharge, evapotranspiration consumption, snow storage, and river channel routing. Since the 671 catchments exhibit substantial differences in baseflow index, soil water storage capacity, and hydrological memory, the average response coefficient of 0.51 should not be interpreted as a direct quantification of a single physical process. The authors are encouraged to provide further discussion on the specific hydrological mechanisms underlying this coefficient.

Thank you for the valuable comment. The value 0.51 is a cross-catchment average statistical response estimated from the panel regression for 671 catchments rather than a direct quantitative estimate of any single hydrological process. In the revision, we have deleted “the buffering capacity of catchment storage” from the Abstract and added the explicit explanation in the Results:

“For GloFAS-ERA5 v4.0, the panel regression estimates an average increase of 0.51 mm in streamflow RMSE for each 1 mm increase in precipitation RMSE across the 671 catchments.” (Page 1, Lines 17 to 19)

“The coefficient of precipitation error for the random-effects panel regression is 0.51,

indicating that a 1 mm increase in precipitation RMSE is associated with an average increase of 0.51 mm in streamflow RMSE. It is noted that the value 0.51 is the average statistical response estimated from the panel regression for 671 catchments, rather than a direct quantitative estimate of any single hydrological process.” (Page 11, Lines 223 to 227)

Also, the hydrological interpretation of divergent responses has been analyzed in the Discussion:

“5.1 Hydrological interpretation of divergent responses

Soil moisture deficits, evapotranspiration and subsurface water storage can limit runoff generation from precipitation in arid catchments (Farmani et al., 2025; Assouline et al., 2024; Cui and Tian, 2026). As shown in Fig. S1 in the Supplement, the correlation between streamflow reanalysis error and soil wetness index (SWI) is statistically significant in 473 catchments, with the median correlation coefficient of 0.52. This result is consistent with the relatively weaker responses of streamflow reanalysis errors to precipitation errors observed in arid catchments and stronger responses in some humid catchments. Dry antecedent soil conditions generally allow a large proportion of precipitation to infiltrate and replenish soil water storage before reaching the thresholds to generate runoff (Liu et al., 2025; Tian and Xu, 2024). As antecedent soil moisture increases, the remaining storage deficit decreases, thereby enhancing the sensitivity of runoff generation to precipitation inputs (Ye et al., 2023; Ran et al., 2022). Evapotranspiration and groundwater recharge may further modulate the contemporaneous relationship between precipitation and streamflow errors (Berghuijs et al., 2025). Since evapotranspiration and groundwater data are not provided in GloFAS-ERA5 and Daymet datasets (Grimaldi et al., 2023; Thornton et al., 2016), their contributions and the specific distribution of precipitation errors among them cannot be determined.

In high-latitude or mountainous catchments, the seasonal snow storage and subsequent melt can alter the timing and magnitude of the streamflow responses to precipitation (Zhu et al., 2023; Wang et al., 2021; Nan et al., 2026). Precipitation errors in snow season may be retained in snow storage and appear months later as streamflow errors (Wang et al., 2021; Hale et al., 2023). As shown in Fig. S2 in the Supplement, the correlation between precipitation RMSE and temperature RMSE over cold days (Daymet temperature ≤ 0 °C) for each hydrological year is statistically significant in 56 catchments but not significant in 275 catchments. Temperature influences precipitation phase, snow accumulation and melt timing, but these physical linkages do not

necessarily imply a strong statistical coupling between precipitation errors and temperature errors. The errors calculated over hydrological year partly integrate intra-annual effects of storage and lag (Berghuijs et al., 2025; de Lavenne et al., 2022), but they obscure seasonal contrasts, snow accumulation–melt transitions and intra-seasonal phase errors (Hale et al., 2023; de Lavenne et al., 2022). The annual errors therefore cannot resolve seasonal or event-scale errors, particularly for extreme precipitation and flood events. In addition, as the Daymet dataset is generated by interpolating observations (Thornton et al., 2016), uncertainties associated with station density and interpolation methods may affect the precipitation errors.” (Page 20, Lines 325 to 349)

2. The sources of unexplained errors should be discussed more explicitly. Although interaction terms involving precipitation seasonality and snow fraction are incorporated, the coefficient of determination of the panel regression remains 0.36, indicating that a considerable proportion of streamflow reanalysis errors is still not explained by the current predictors. The authors should discuss other potential factors contributing to the remaining uncertainty, such as soil properties, groundwater processes, model structure, or other catchment characteristics.

Thank you for the insightful comment. In the revision, the potential factors contributing to the remaining uncertainty have been added in the Discussion:

“The performance of streamflow reanalysis is primarily constrained by climate forcings, initial conditions, representation of physical processes and model parameterization (Liu et al., 2023; Nourani et al., 2026; Nanding et al., 2021b).” (Page 20, Lines 352 and 353)

“The precipitation errors can explain only a limited proportion of the variance in GloFAS-ERA5 streamflow errors, with the R^2 of the panel regression not exceeding 0.36. The remaining variance may be related to catchment heterogeneity, hydrological models and uncertainties in the reference datasets.” (Page 21, Lines 363 to 365)

3. The manuscript attributes the weaker influence of precipitation errors on streamflow errors in arid catchments to soil moisture deficits absorbing part of the precipitation errors. However, further explanation is needed regarding where the “absorbed” errors are ultimately reflected in the water balance. For example, do these errors mainly contribute to changes in evapotranspiration, soil water storage or other hydrological components? A more cautious discussion of this mechanism is recommended.

Thank you for the instructive comment. Since the available variables in the GloFAS-ERA5 and Daymet datasets only support analysis of soil water and snow storage, we

have added additional analysis in the Supplement and have stated this limitation explicitly in the Discussion:

“Soil moisture deficits, evapotranspiration and subsurface water storage can limit runoff generation from precipitation in arid catchments (Farmani et al., 2025; Assouline et al., 2024; Cui and Tian, 2026). As shown in Fig. S1 in the Supplement, the correlation between streamflow reanalysis error and soil wetness index (SWI) is statistically significant in 473 catchments, with the median correlation coefficient of 0.52. This result is consistent with the relatively weaker responses of streamflow reanalysis errors to precipitation errors observed in arid catchments and stronger responses in some humid catchments. Dry antecedent soil conditions generally allow a large proportion of precipitation to infiltrate and replenish soil water storage before reaching the thresholds to generate runoff (Liu et al., 2025; Tian and Xu, 2024). As antecedent soil moisture increases, the remaining storage deficit decreases, thereby enhancing the sensitivity of runoff generation to precipitation inputs (Ye et al., 2023; Ran et al., 2022). Evapotranspiration and groundwater recharge may further modulate the contemporaneous relationship between precipitation and streamflow errors (Berghuijs et al., 2025). Since evapotranspiration and groundwater data are not provided in GloFAS-ERA5 and Daymet datasets (Grimaldi et al., 2023; Thornton et al., 2016), their contributions and the specific distribution of precipitation errors among them cannot be determined.” (Page 20, Lines 326 to 337)

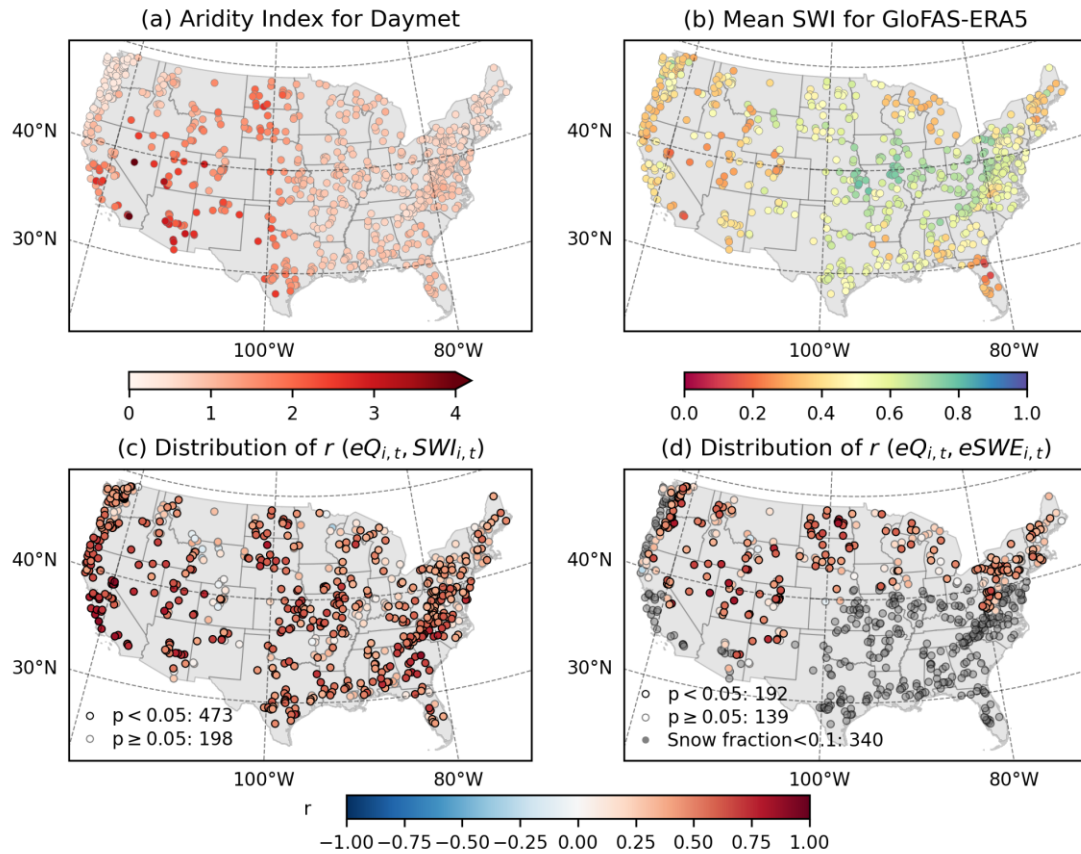


Figure S1. Spatial distribution of (a) aridity index for Daymet, (b) mean soil wetness index (SWI) from the GloFAS-ERA5 dataset and (c-d) Pearson's correlation coefficients of streamflow reanalysis error to SWI and to snow depth water equivalent (SWE) error. The eSWE represents the RMSE of SWE from the GloFAS-ERA5 dataset by hydrological year using the SWE from the Daymet dataset as reference data.”

4. This study mainly investigates error responses through regression relationships between annual RMSE values. Therefore, the results provide evidence of statistical associations rather than strict causal effects. The authors should avoid using overly strong causal language and consider reducing expressions such as “control”, “confirm”, and “entirely dependent on”. For example, in Line 287, the statement “almost entirely dependent on precipitation accuracy” appears too strong and could be revised to “predominantly associated with precipitation accuracy during the examined period”.

Thank you for the insightful comment. In the revision, we have checked throughout this paper and revised the overly strong statements:

“These divergent responses suggest that streamflow errors may be more directly associated with precipitation errors under the saturation-excess conditions while soil

moisture deficits can dampen their effects.” (Page 1, Lines 20 to 22)

“The coefficient of precipitation error for the random-effects panel regression is 0.51, indicating that a 1 mm increase in precipitation RMSE is associated with an average increase of 0.51 mm in streamflow RMSE.” (Page 11, Lines 253 to 255)

“This case study indicates that in humid rain-dominated catchments, the performance of streamflow reanalysis is predominantly associated with the accuracy of precipitation reanalysis rather than temperature.” (Page 17, Lines 299 and 300)

Minor comments

1. The GloFAS streamflow data are generally provided in discharge units, whereas streamflow errors in this study are expressed in millimeters. The authors should clarify in the Methods section whether streamflow values were converted from discharge to runoff depth.

Thank you for the instructive comment. The preprocessing of the GloFAS-ERA5 streamflow reanalysis has been added in the Methods:

“The streamflow data of the GloFAS-ERA5 dataset are converted from discharge (m³/s) into runoff depth (mm/d) based on the catchment area.” (Page 4, Lines 101 and 102)

2. Daymet is a high-resolution gridded dataset generated through interpolation based on station observations. The authors are encouraged to briefly acknowledge the uncertainties associated with Daymet in the Discussion section.

Thank you for the insightful comment. In the revision, we have added this limitation in the Discussion:

“In addition, as the Daymet dataset is generated by interpolating observations (Thornton et al., 2016), uncertainties associated with station density and interpolation methods may affect the precipitation errors.” (Page 20, Lines 348 and 349)

3. Please check the formulation of the R² equation. In Equation (4), the numerator and denominator appear to lack squared terms.

We are sorry for the typo. It has been corrected:

“For linear regression of catchment i , the R^2 is calculated:

$$R^2 = 1 - \frac{\sum(eQ_{i,t} - \widehat{eQ}_{i,t})^2}{\sum(eQ_{i,t} - \overline{eQ}_i)^2} \quad (4)$$

where $\widehat{eQ}_{i,t}$ and $\overline{eQ}_i$ denote the fitted value and mean value of $eQ_{i,t}$, respectively.”

(Page 6, Lines 121 to 123)

4. Please provide a clearer definition of precipitation seasonality. The calculation method and the physical meaning of positive and negative values should be explicitly described.

Thank you for the insightful comment. The calculation and physical meaning have been added in the Methods:

“The precipitation seasonality is estimated by fitting sinusoidal functions to the mean annual cycles of temperature and precipitation. Positive (negative) values indicate that precipitation peaks in summer (winter), whereas values near zero represent a relatively uniform distribution of precipitation throughout the year (Addor et al., 2017).” (Page 7, Lines 144 to 146)

5. Line 163: The word “diversifies” should be used instead of “diversify” in the phrase “how catchment heterogeneity diversify the error responses”.

Thank you for spotting the typo. We have corrected it:

“Furthermore, by incorporating interaction terms into the random-effects panel regression, this analysis isolates and quantifies the marginal effects of precipitation and temperature errors, revealing how catchment heterogeneity diversifies the error responses.” (Page 7, Lines 166 to 168)

6. Line 246: There is an extra word “the” in the phrase “the the saturation-excess mechanism”. Please remove the redundant word.

Thank you for spotting the typo. We have corrected it:

“This result suggests that streamflow errors may be more directly associated with precipitation errors under the saturation-excess conditions while soil moisture deficits can dampen their effects.” (Page 14, Lines 257 and 258)