Response

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 indicates that for every 1 mm increase in precipitation RMSE, the corresponding streamflow RMSE increases by an average of 0.51 mm." (Page 1, Lines 17 to 19)

"It is noted that 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." (Page 11, Lines 220 to 222)

(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 diagnostic evidence as 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 (Mingjuan et al., 2025; Farmani et al., 2025; Assouline et al., 2024). As shown in Fig. S1 in the Supplement, the correlation between streamflow reanalysis error and soil wetness index (SWI) is 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 great proportion of precipitation to infiltrate and replenish soil water storage before reaching the thresholds to generate runoff (Liu et al., 2025; Cui and Tian, 2026). 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." (Pages 19 and 20, Lines 318 to 329)

"

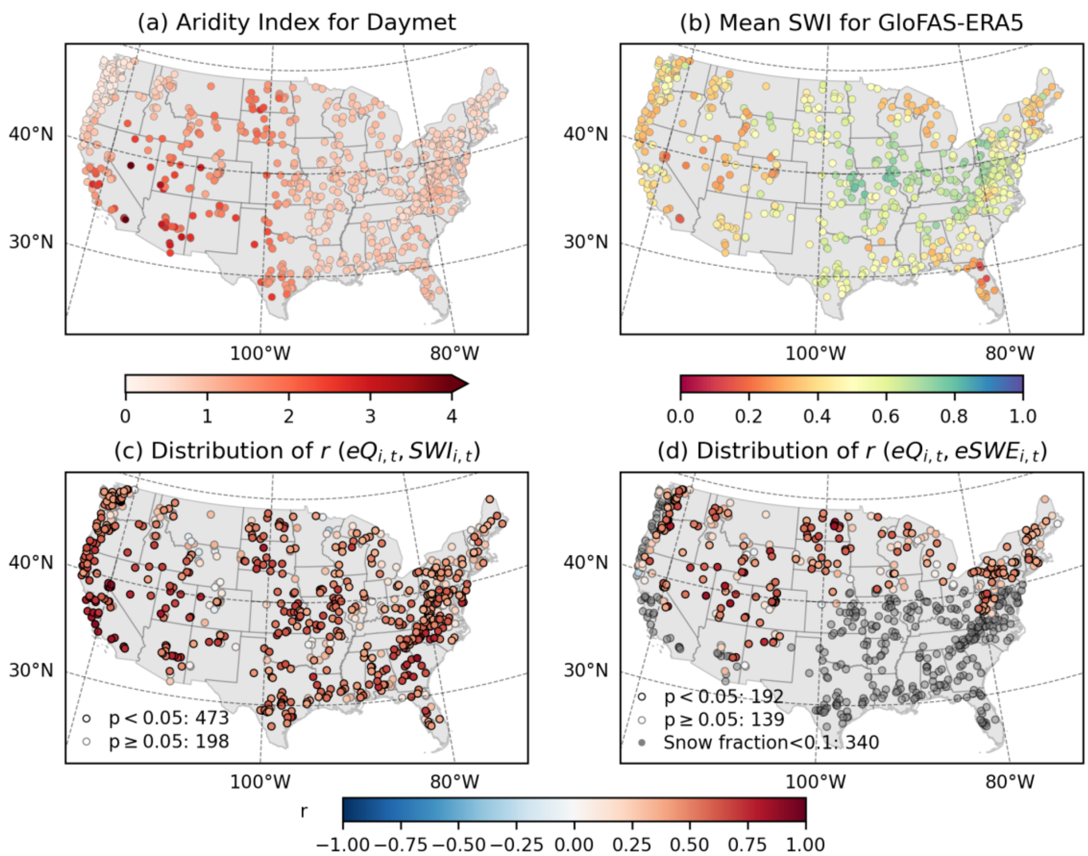

Figure S1. Spatial distribution of (a) aridity index for Daymet, (b) mean soil wetness index (SWI) from the GloFAS-ERA5 dataset and (c-d) Pearson's correlation coefficients of streamflow reanalysis error to SWI and to snow depth water equivalent (SWE) error. The eSWE represents the RMSE of SWE from the GloFAS-ERA5 dataset by hydrological year using the SWE from the Daymet dataset as reference data." (Page 1 in the Supplement)

(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 Results 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 9, Lines 199 to 201)

"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; McMillan, 2020). 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 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 330 to 336)

“

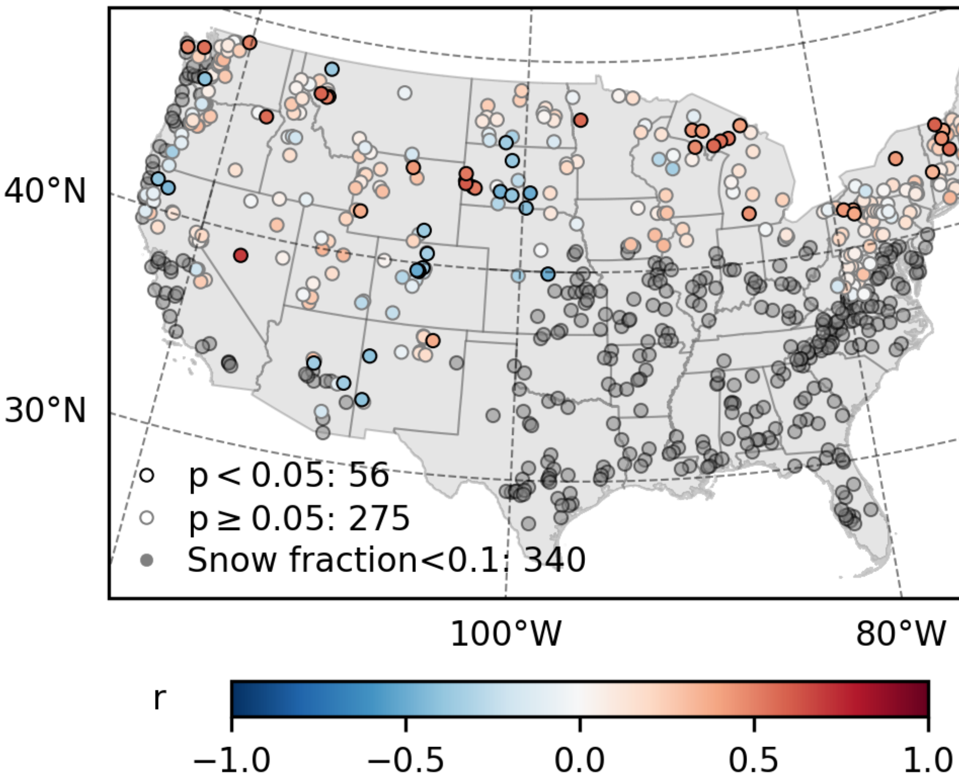

Figure S2. Pearson’s correlation coefficients (r) between precipitation reanalysis errors and temperature reanalysis errors on cold days (Daymet temperature ≤ 0°C) by hydrological year.” (Page 2 in the Supplement)

(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:

catchment-specific linear regressions reveal local heterogeneity, whereas the panel regression estimates an average relationship using spatial and temporal information across the 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 126 and 127)

"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 362 and 363)

(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 336 to 340)