Response

Anonymous Referee #2

This study presents an approach to evaluating streamflow reanalysis from a time-frequency domain using wavelet transforms. By applying the wavelet transform to both reanalysis and observation data, the authors decompose the time series into different scales and conduct a performance evaluation for each component. Additionally, they employ random forests combined with accumulated local effects (ALE) to analyze the influence of catchment attributes on reanalysis performance across various time scales. The results offer valuable insights into the understanding of reanalysis data and provide plausible explanations based on catchment characteristics, which are crucial for the practical application of reanalysis data.

We are grateful to you for the positive comments.

Despite the findings, the manuscript requires improvements to better illustrate both the methods and the results part.

Thank you very much for the constructive comments. We have improved the paper accordingly and provide point-by-point responses.

1. Novelty and Justification: The manuscript does not sufficiently highlight the novelty and importance of the proposed methodology. The authors should provide a clearer rationale for choosing this method over other potential techniques, while the wavelet transform is a powerful tool. Additionally, a more succinct and cohesive summary of the methodology would enhance the manuscript. Specifically, I suggest including a part that clearly outlines the motivation for using the wavelet-based method and the connections among the main steps, including the basic inputs and outputs of each process, with clear symbols and formulas distinguishing between reanalysis and observation data.

Thank you for the constructive suggestion. We have improved the manuscript to better emphasize the novelty and significance of using the wavelet transform to evaluate global streamflow reanalysis. Specifically, we have provided a clearer rationale for choosing the wavelet transform over other techniques, highlighting its ability to perform multiresolution analysis. In the meantime, we have added Section 2.1 to illustrate the novel decomposition approach. This includes detailed descriptions of the basic inputs and outputs of each process, with clear symbols and formulas distinguishing between reanalysis and observation data. Specifically, the subscripts d and q respectively represent reanalysis and observed streamflow:

"Time series analysis is one of the most important approaches to investigating the performance of hydrological models (Lane, 2007; Zuo et al., 2020; Saraiva et al., 2021). From the perspective of time series, hydrological simulations are a combination of the components of periodic motion, trend, seasonality and error, which can be extracted by using decomposition approaches (Zuo et al., 2020; Abebe et al., 2022; Xu et al., 2022). As one of the most important decomposition approaches, wavelet transform decomposes streamflow into time series of wavelet coefficients under certain frequencies (Manikanta and Vema, 2022). Therefore, it allows for multiresolution analysis compared to other decomposition approaches (Montoya et al., 2022). Owing to the time-frequency characterization, wavelet-based features of reanalysis and observed streamflow can be compared in order to zoom into detailed information for multiple time series segments (Manikanta and Vema, 2022). If there are errors in the reanalysis at specific timescales or during specific periods, the sources of these errors can be identified by the technique of time-frequency characterization (Lane, 2007)." (Page 2, Lines 48 to 57)

"A novel decomposition approach that combines the wavelet transform with machine learning techniques is proposed to evaluate global streamflow reanalysis in the time-frequency domain. There are three steps:

(1) Decomposition of time series: the DWT is used to decompose the reanalysis and observed streamflow time series, resulting in approximation and detail components at different scales;

(2) Verification of decomposed series: the Kling-Gupta Efficiency (KGE), correlation, bias ratio and variability ratio are derived to indicate the local performance of original time series, approximation and detail components at various scales. In the meantime, the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm is used to remove outliers from the verification metrics;

(3) Influences of catchment attributes: the ALEs derived from the random forest model is employed to elaborate on the influences of catchment attributes and then identify the driving factors." (Page 3, Lines 71 to 80)

"For reanalysis and observed streamflow time series, the decomposition is denoted as:

$$\begin{cases} d_t = \sum_{l=1}^{l_m} D_{d,l} + A_{d,l_m} \\ q_t = \sum_{l=1}^{l_m} D_{q,l} + A_{q,l_m} \end{cases} \tag{6}$$

in which d_t is the reanalysis, q_t is the observed streamflow and l_m is the maximum decomposition level. The subscripts d and q respectively represent reanalysis and observed streamflow.

The DWT captures time series information at multiple scales in the time-frequency domain, with each scale corresponding to a specific period (Joo and Kim, 2015; Manikanta and Vema, 2022). Specifically, the approximation and detail components at decomposition level l correspond to the time scale 2^l days (Nalley et al., 2012).

The KGE stands out as a widely used verification metric to evaluate the model performance (Frame et al., 2021; Huang and Zhao, 2022; Zhao et al., 2022). It indicates the performance of original time series, approximation and detail components. When evaluating the performance of original time series, the KGE is calculated as follows:

$$KGE_o = 1 - \sqrt{(r_o - 1)^2 + (\beta_o - 1)^2 + (\gamma_o - 1)^2} \tag{7}$$

As can be seen, the KGE_o is comprised of three components, namely, the Pearson correlation coefficient r_o, the bias ratio β_o and the variability ratio γ_o:

$$r_o = \frac{\sum_{t=1}^{T}(d_t - \mu_d)(q_t - \mu_q)}{\sqrt{\sum_{t=1}^{T}(d_t - \mu_d)^2}\sqrt{\sum_{t=1}^{T}(q_t - \mu_q)^2}} \tag{8}$$

$$\beta_o = \frac{\mu_d}{\mu_q} \tag{9}$$

$$\gamma_o = \frac{\sigma_d}{\sigma_q} \tag{10}$$

in which μ is the mean streamflow and σ is the streamflow standard deviation. The subscripts d and q respectively represent reanalysis and observed streamflow. The KGE ranges from $-\infty$ to 1, with a perfect value of 1." (Pages 4 to 5, Lines 101 to 115)

2. Robustness of Results: Could the authors discuss whether the results hold consistently across different catchment selections for training and testing and provide any relevant sensitivity analyses?

Thank you for the valuable suggestion. We have conducted a randomized allocation of catchments into training and testing sets. Specifically, 75% of catchments are randomly allocated for training and the remaining 25% for testing. This random allocation process was repeated 5000 times to investigate the robustness of results. Subsequently, we calculated the Accumulated Local Effects (ALEs) and presented the results in Figures 2 to 5 in the supplementary material. The results suggest that the influence of catchment attributes on model performance exhibits robustness. It can be observed that the KGE values of original time series and its approximation components are primarily influenced by precipitation seasonality. In the meantime, the correlations of annual and multi-annual features (from A7 to A10) are mainly affected by the precipitation seasonality, while daily, weekly and monthly features are influenced by longitude and mean slope of catchment. Furthermore, the bias ratio is primarily influenced by mean precipitation and the variability ratio is mainly affected by catchment area and depth to bedrock. The geology, soils and vegetation appear to have minor impacts on the local performance of global streamflow reanalysis:

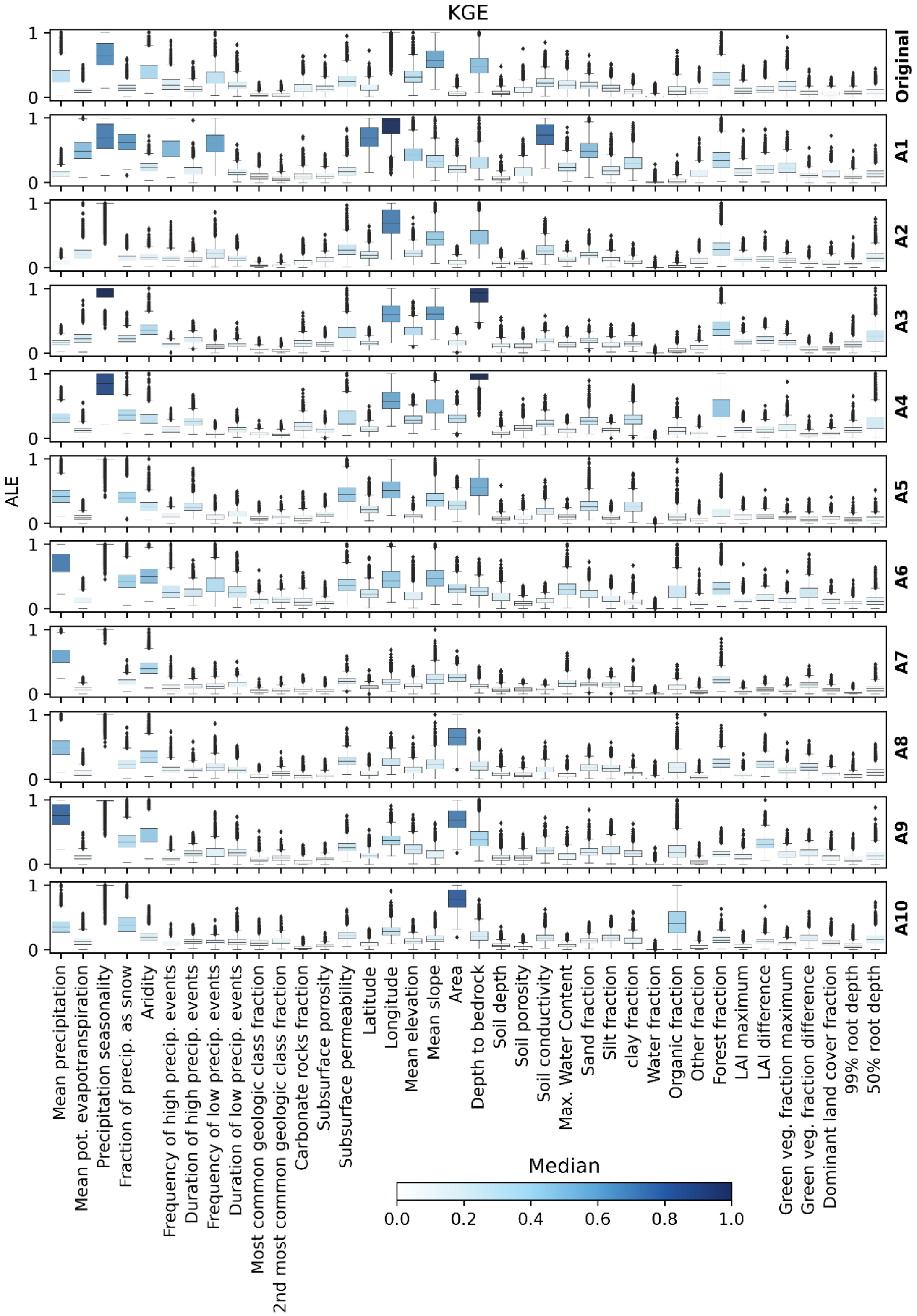

Figure S2. The ALEs of the catchment attributes on the KGE for the original time series and approximation components derived from 5000 selections of catchments for training and testing. The ALE represents the mean absolute values for each ALE curve, normalized for each original time series (or approximation component). The color of the box indicates its median value.

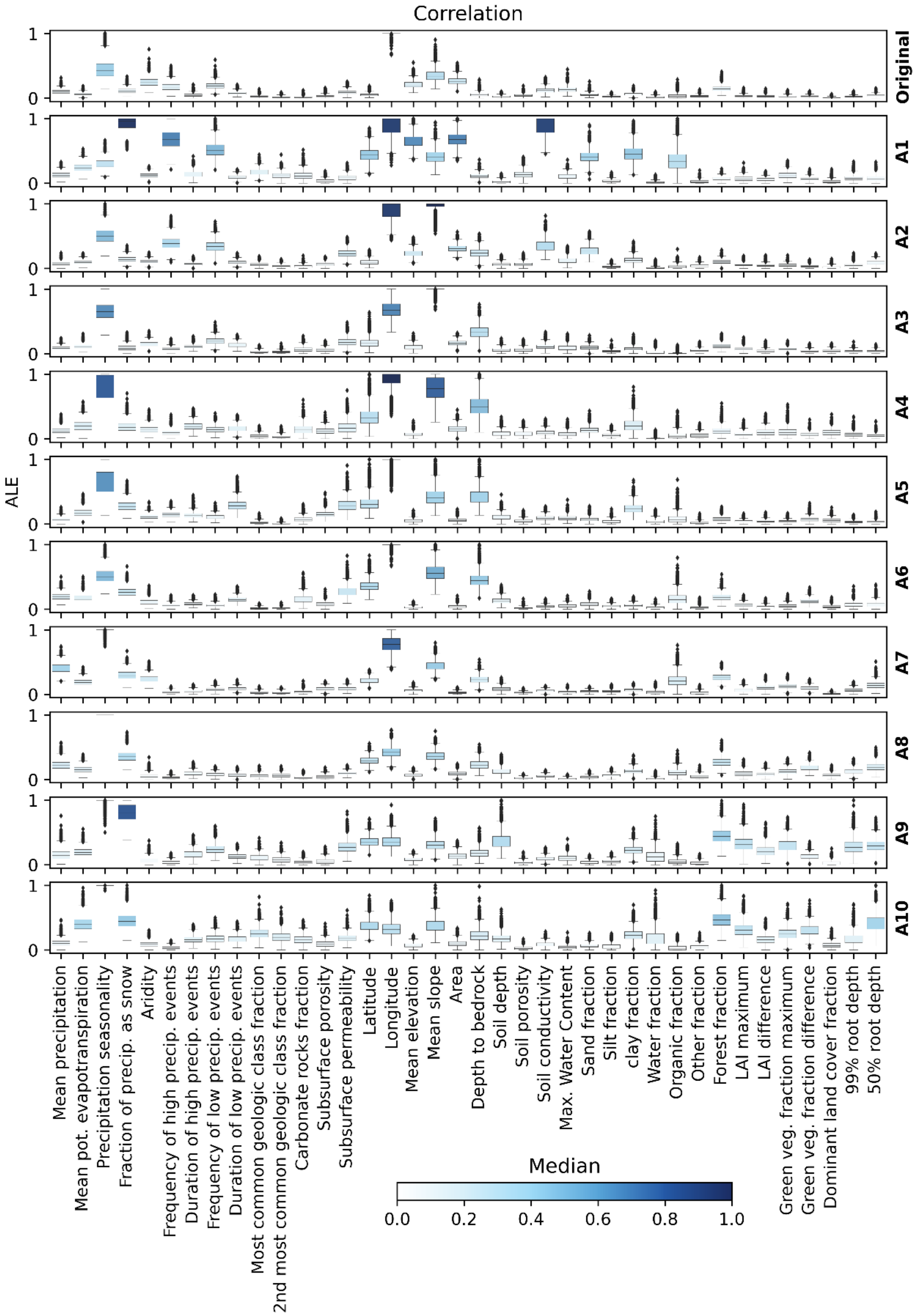

Figure S3. As for Figure S2 but for correlation.

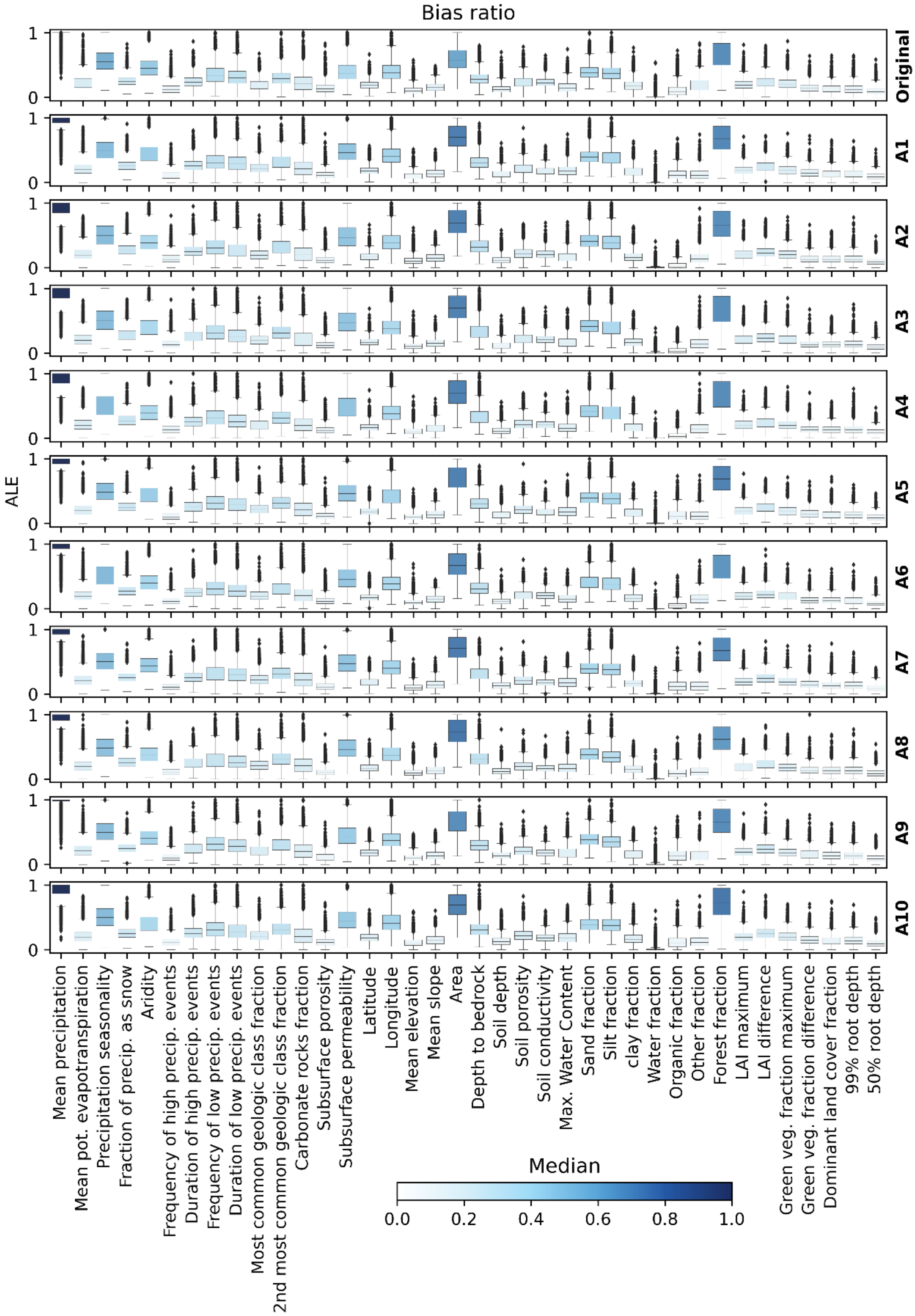

Figure S4. As for Figure S2 but for bias ratio.

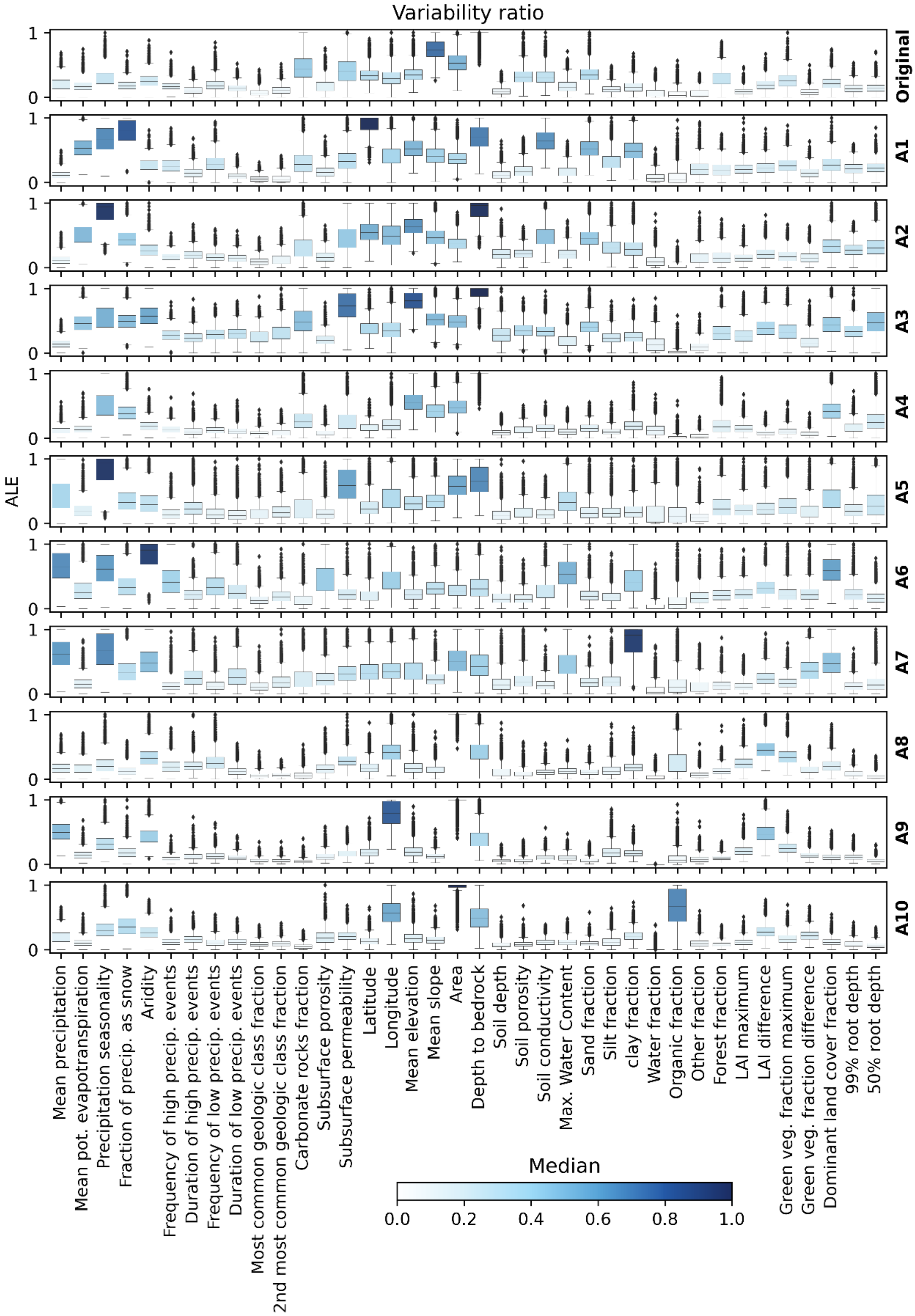

Figure S5. As for Figure S2 but for variability ratio.

3. Terminology Clarification: In Line 124, the authors introduce ALEs and use an abbreviation. Given that ALEs might not be familiar to all readers within the hydrology community, could the authors provide a full name and a clear explanation of when the term was first introduced? Although ALEs are mentioned in the introduction, they should be explicitly defined and elaborated upon to ensure clarity throughout the manuscript.

Thank you for the insightful comment. We have provided a clear explanation of the ALE in the methods:

"The ALEs are used to describe how catchment attributes influence the performance of approximation components at various scales for reanalysis based on the random forest model. They illustrate how changes in one input variable impact model predictions by analysing the differences within small quantile-based intervals (Stein et al., 2021). An advantage of the ALEs is the overcome of the confounding effects of correlated catchment attributes (Stein et al., 2021). The ALE curves reveal whether the association is linear or exhibits more complex patterns (Teng et al., 2022)." (Pages 6 to 7, Lines 151 to 155)

4. Figure 1. The Kling-Gupta Efficiency (KGE) is upper bounded by 1, with higher values indicating better performance. It would be more effective to use a linear color scale to represent this metric, enhancing the interpretability of the figure.

Thank you for the insightful comment. The Figure 1 has been replaced by the time series plots of original time series and its approximation and detail components:

"The time series of streamflow reanalysis and observation along with their approximation and detail components are presented in Figure 1. The plots are for the station 6224000 in which streamflow reanalysis tends to exhibit the highest KGE value of 0.82. The approximation and detail components at the level l correspond to the time scale of 2l days. For example, A1 and A8 correspond to the periods of 2 and 256 days, respectively. It can be observed that the original time series of reanalysis generally captures the primary features of the observed streamflow. Under the stepwise decomposition of the streamflow time series, the KGE tends to increase from 0.48 for A1 to 0.62 for A8 and increase from -4.57 for D1 to 0.48 for D8. This result indicates that streamflow reanalysis tends to capture seasonal and annual information more effectively than daily, weekly and monthly information. At higher decomposition levels, the series of approximation and detail components becomes smoother. As the decomposition level increases, the reanalysis becomes more able to capture the information in the observation." (Page 9, Lines 203 to 212)

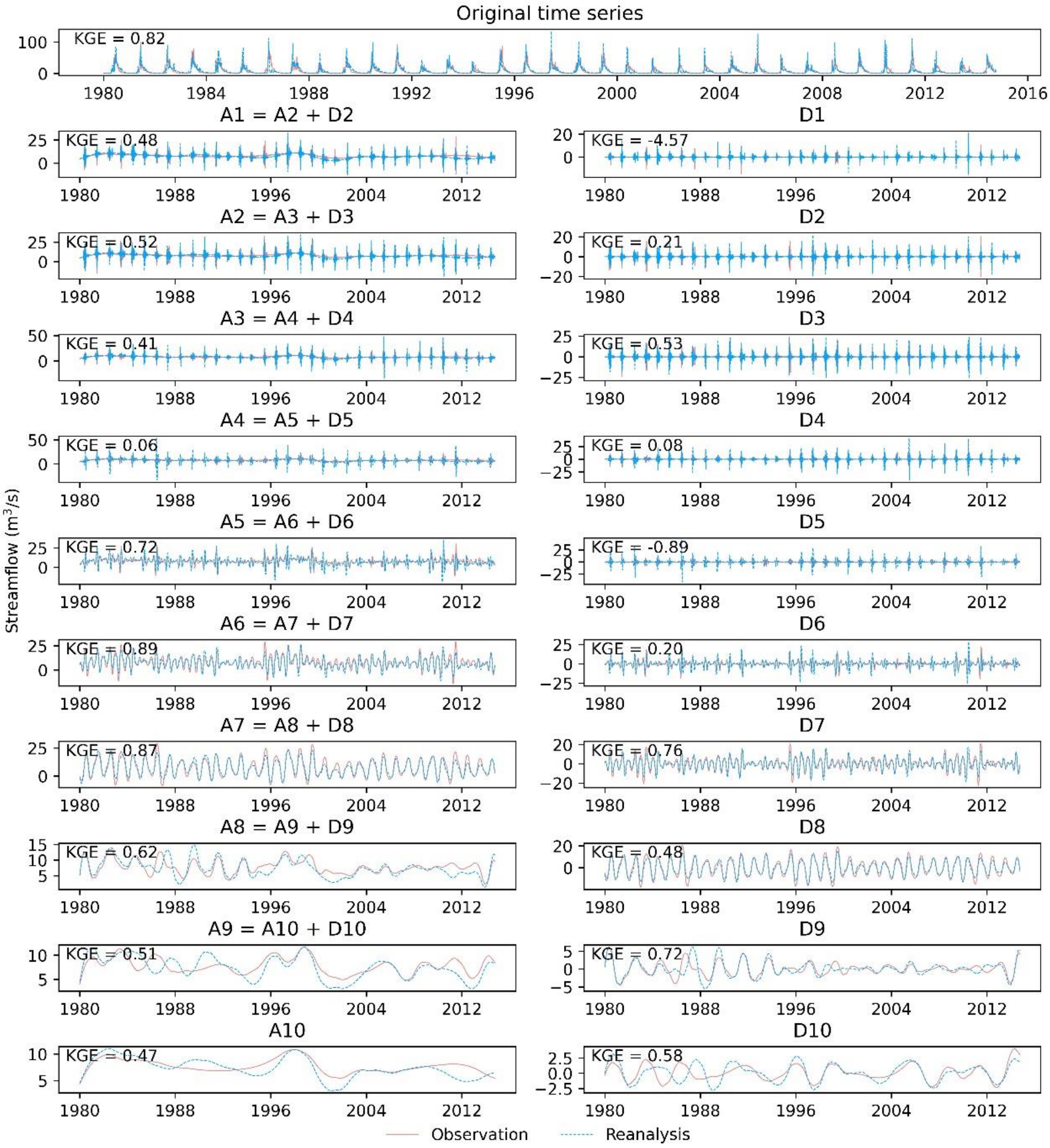

Figure 1: Time series plots of original time series and its approximation and detail components for the station 6224000.