Responses

Editor:

The editor has decided that Revisions are necessary before the review process can be continued. Please log in using your Copernicus Office user ID 405800 to find the editor report at: https://editor.copernicus.org/HESS/ms_records/hess-2024-83. We kindly ask you to revise your manuscript accordingly and to upload the revised files, a point-by-point reply to the comments, and a marked-up manuscript version showing the changes made no later than 16 Jul 2024 at: https://editor.copernicus.org/HESS/review-file-upload/hess-2024-83

We appreciate the decision and are grateful to the two reviewers for the insightful and constructive review comments. Accordingly, we have conducted a thorough revision of the whole paper.

Below please find the point-to-point responses.

Anonymous Referee #1:

This paper presents a novel approach to investigate the performance of GloFAS streamflow reanalysis from the perspective of the time-frequency domain. The results provide interesting insights into the performance of global streamflow reanalysis datasets and attribution analysis. The paper is well-structured and well-written. Below are some comments for consideration.

We appreciate the positive comments.

1. Abstract - The significance of evaluating global streamflow reanalysis in the time-frequency is in demand, which helps clarify the contribution of this paper.

Thank you for the insightful comment. The abstract has been improved to highlight the significance of evaluating global streamflow reanalysis in the time-frequency domain:

"While global streamflow reanalysis has been evaluated at different spatial scales to facilitate practical applications, its local performance in the time-frequency domain is yet to be investigated. This paper presents a novel decomposition approach to evaluating streamflow reanalysis by combining wavelet transform with machine learning. Specifically, the time series of streamflow reanalysis and observation are respectively decomposed and then the approximation components of reanalysis are compared to those of observed streamflow. Furthermore, the accumulated local effects are derived to showcase the influences of catchment attributes on the performance of streamflow reanalysis at different scales. For streamflow reanalysis generated by the Global Flood Awareness System, a case study is devised based on streamflow observations from the Catchment Attributes and Meteorology for Large-sample Studies. The results highlight that the reanalysis tends to be more effective in characterizing seasonal, annual and multi-annual features than daily, weekly and monthly features. The Kling-Gupta Efficiency (KGE) values of original time series and approximation components are primarily influenced by precipitation seasonality. High values of KGE tend to be observed in catchments where there is more precipitation in winter, which can be due to low evaporation that results in reasonable simulations of soil moisture and baseflow processes. The longitude, mean precipitation and mean slope also influence the local performance of approximation components. On the other hand, attributes on geology, soils and vegetation appear to play a relatively small part in the performance of approximation components. Overall, this paper provides useful information for practical applications of global streamflow reanalysis." (Page 1, Lines 9 to 24)

2. Lines 110 to 120 – Consider adding some diagnostic plots about the clustering results in the supplementary material, as the existence of outliers also indicates the variability of global streamflow reanalysis.

Thank you for the valuable suggestion. We have added four boxplots in Figure S1 in the supplementary material to illustrate the difference between the performance of inliers and outliers using the Density-Based Spatial Clustering of Applications with Noise (DBSCAN). The KGEs of outliers are generally lower than those of inliers:

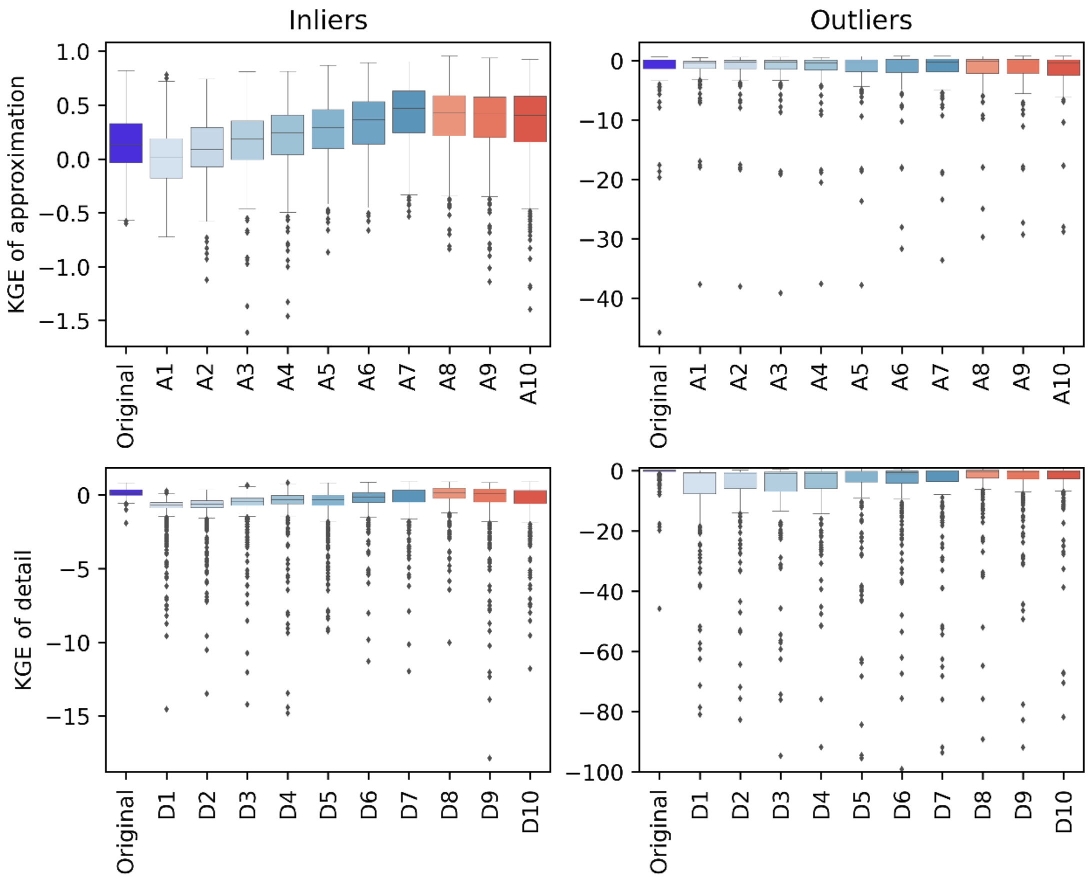

Figure S1. The KGEs of inliers and outliers for approximation and detail components using the Density-Based Spatial Clustering of Applications with Noise (DBSCAN).

3. (11) – Two KGE terms are used in this equation. For clarity, I suggest adding subscripts to differentiate between the observed KGE and the predicted KGE by RF.

Thank you for the constructive suggestion. To clarify Equation (11), p and o are respectively used as subscripts for KGE to represent the predicted KGE by the random forest model and the observed KGE:

"Taking the KGE of original time series as an example, the prediction of the performance of approximation components for reanalysis using the random forest model is denoted as:

$$KGE_p = RF(X) \quad (11)$$

in which KGE_p is the predicted KGE using the random forest model, $RF(\cdot)$ is the random forest model and X is the catchment attributes. The R^2 between the predicted KGE_p and the calculated KGE_o is denoted by:

$$R^2 = \left(\frac{\sum_{i=1}^{N} \left(KGE_{p,i} - \mu_{KGE_p} \right) \left(KGE_{o,i} - \mu_{KGE_o} \right)}{\sqrt{\sum_{i=1}^{N} \left(KGE_{p,i} - \mu_{KGE_p} \right)^2} \sqrt{\sum_{i=1}^{N} \left(KGE_{o,i} - \mu_{KGE_o} \right)^2}} \right)^2 \quad (12)$$

in which μ is the mean KGE. The KGE_p and KGE_o represent the predicted KGE of the random forest model and the calculated KGE between reanalysis and observed streamflow, respectively." (Page 6, Lines 145 to 150)

4. Results – The correspondence between approximation level (e.g., A1, A2) and time scale (e.g., daily, monthly) is mentioned in Lines 255 to 260. To improve readability, consider moving this information to Section 4.1.

Thank you for the valuable suggestion. We have added information about the correspondence between approximation level and time scale in Sections 2.2 and 4.1:

"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)." (Page 4, Lines 104 to 106)

"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 2^l 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)

5. Results - The varying number of stations under investigation in this paper may raise concerns about the robustness of the results. In Line 185, it is demonstrated that there are 661 stations, while 554 stations are used in Figure 4. If the reduction is due to the use of clustering, please include a description in the paper to clarify this point.

Thank you for the valuable comment. The 554 stations shown in Figure 4 are due to the use of clustering. We have added a description in the paper for clarification:

"The KGE values of original time series and its approximation components for the 554 catchments after removing the outliers are presented in Figure 3. In total, there are 11 spatial plots for original time series and its components after decomposition. It can be observed that the original time series tends to exhibit relatively high KGEs in the western United States and relatively low KGEs in the central United States. This observation is consistent with those of Addor et al. (2017), which found poor performances in the high plains and desert southwest. In the meantime, the approximation components from A1 to A10 tend to exhibit high KGEs in the western United States and low KGEs in the central United States. This finding indicates that the KGE values of approximation components are related to the KGE values of original time series. Moreover, as the scale increases from A1 to A10, the performance of approximation components tends to improve. The KGEs in the central United States change from negative values in A1 to positive values in A10. That is, seasonal, annual and multi-annual features tend to be better represented by streamflow reanalysis than daily, weekly and monthly features." (Page 11, Lines 230 to 239)

6. Results – While the Results section is well-written, it would benefit from further interpretations. Overall, similar analyses are conducted for both raw reanalysis and the decomposition. Consider adding further illustrations to highlight the added value or new findings that cannot be directly found based on raw data but are derived from the novel approach.

Thank you for the constructive comment. We have replaced Figures 1 and 2. The new figures demonstrate that the approximation and detail components reveal features of the time series that are not directly found in the original time series. The results indicate that streamflow reanalysis captures seasonal and annual information more effectively than daily, weekly, and monthly information:

"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)

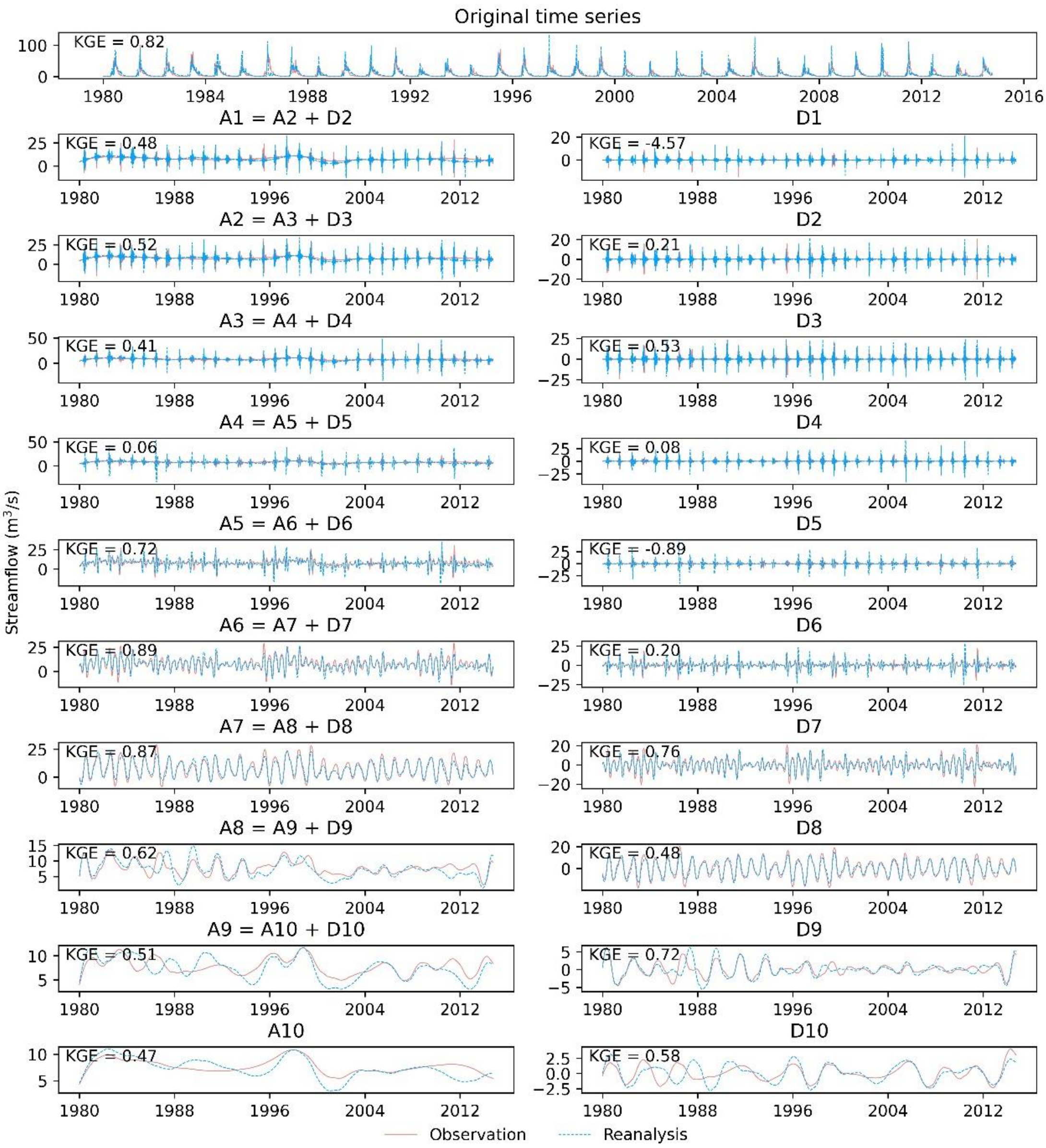

Figure 1: Time series plots of original time series and its approximation and detail components for the station 6224000.

"The KGEs of approximation and detail components across the CONUS are illustrated in Figure 2. There are respectively 554 and 417 catchments for the approximation and detail components by removing outliers. It can be observed that the KGEs of the approximation components tend to increase from A1 to A10 and that by contrast, the KGEs of the detail components exhibit considerable fluctuations from D1 to D10. The comparison between the left and right parts of Figure 2 indicates that the detail components are more difficult to be characterized than the approximation components. This outcome is attributable to the presence of environmental noises in the original time series (Freire et al., 2019). Given that the KGEs of the detail components can drop below -2.5 in some catchments, the attention is paid to the approximation components in the subsequent analysis." (Page 11, Lines 217 to 224)

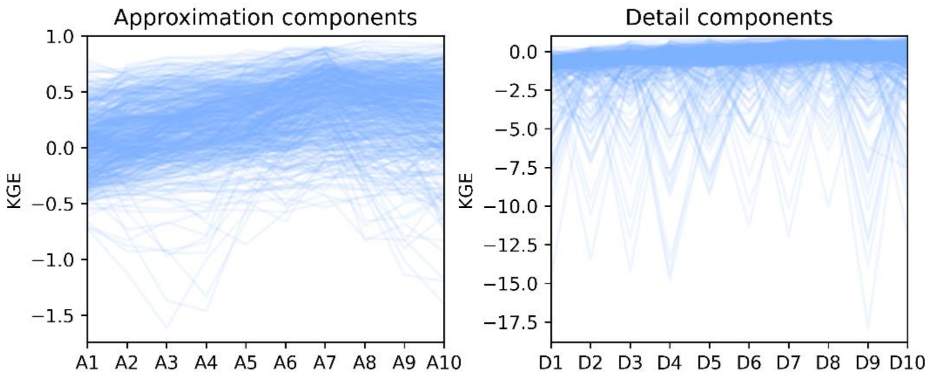

Figure 2: The KGEs of approximation and detail components across the CONUS.

<u>Anonymous Referee #2</u>

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 (Saraiva et al., 2021; Manikanta and Vema, 2022;

Guo et al., 2022). 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 (Abebe et al., 2022; Manikanta and Vema, 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:

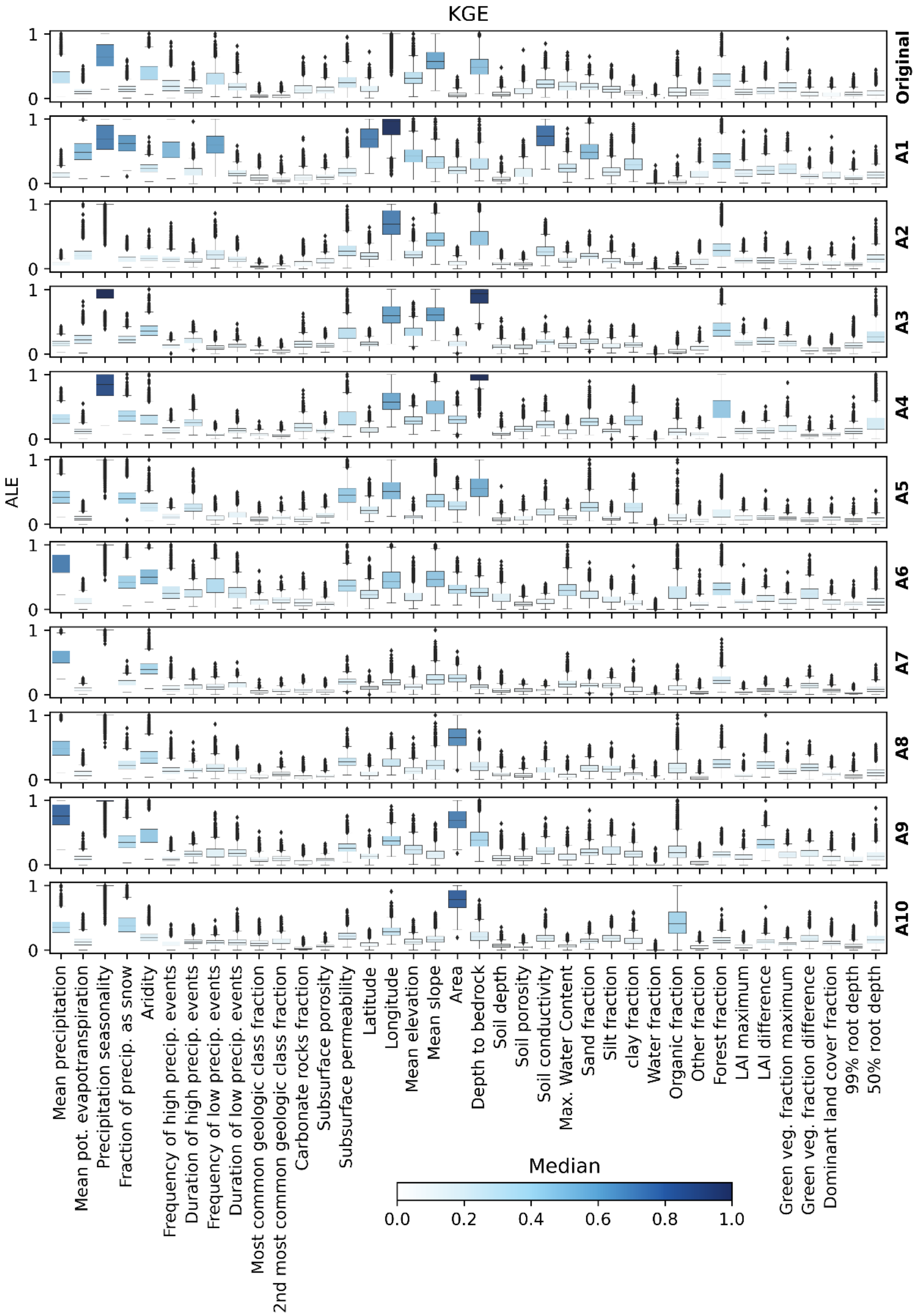

Figure S2. The ALEs of the catchment attributes on the KGE for the original time series and approximation components derived from 5000 selections of catchments for training and testing. The ALE represents the mean absolute values for each ALE curve, normalized for each original time series (or approximation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

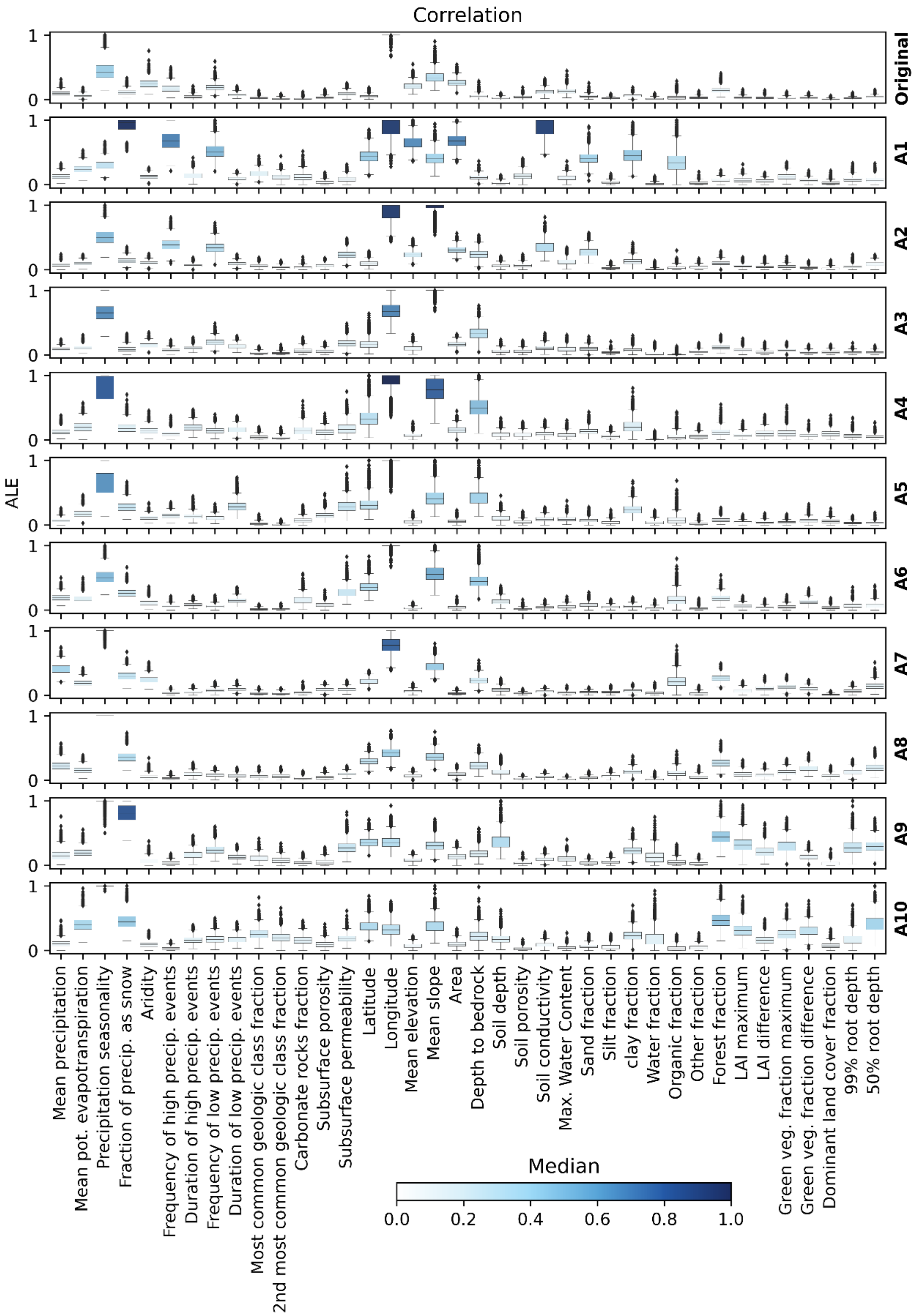

Figure S3. As for Figure S2 but for correlation.

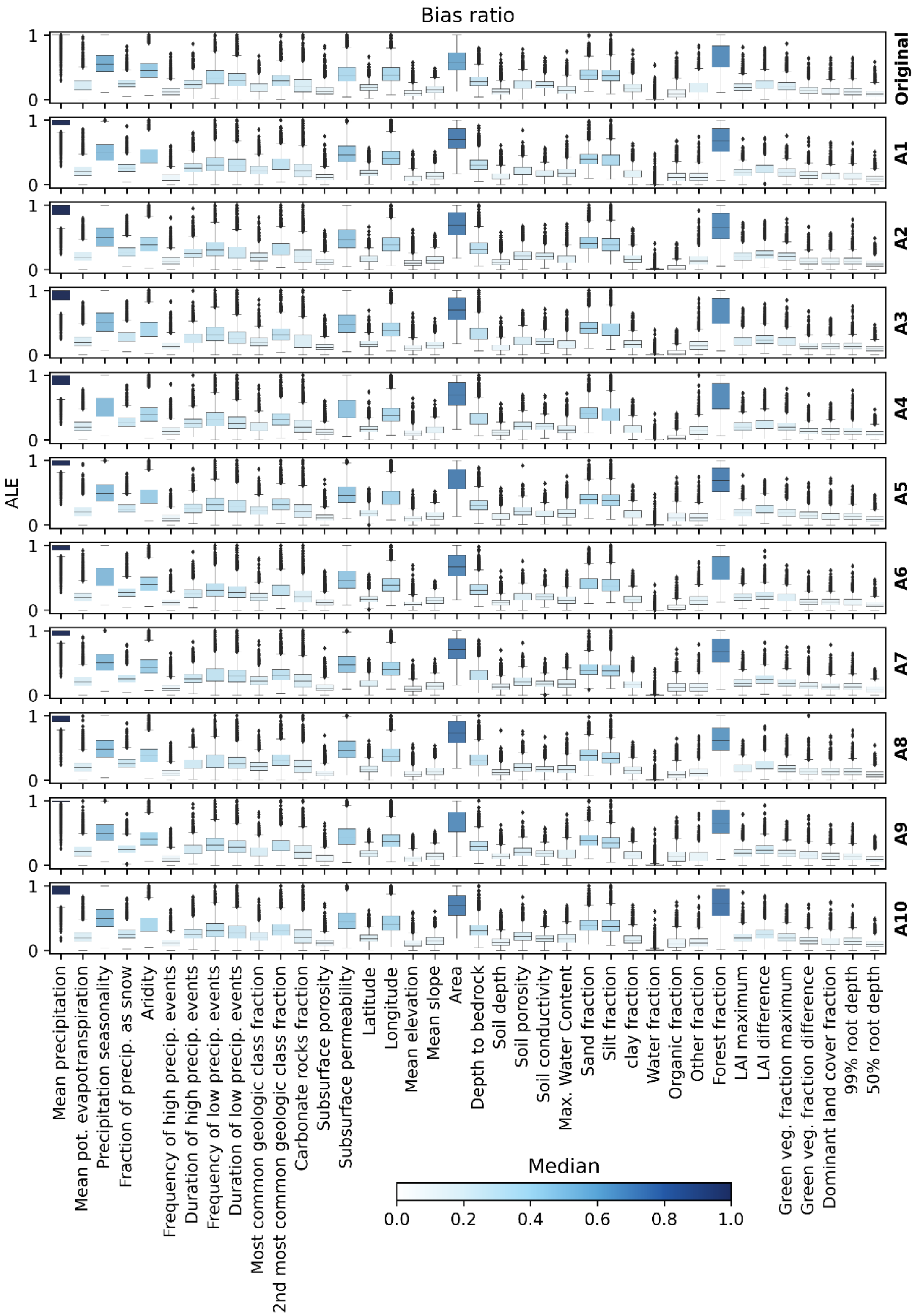

Figure S4. As for Figure S2 but for bias ratio.

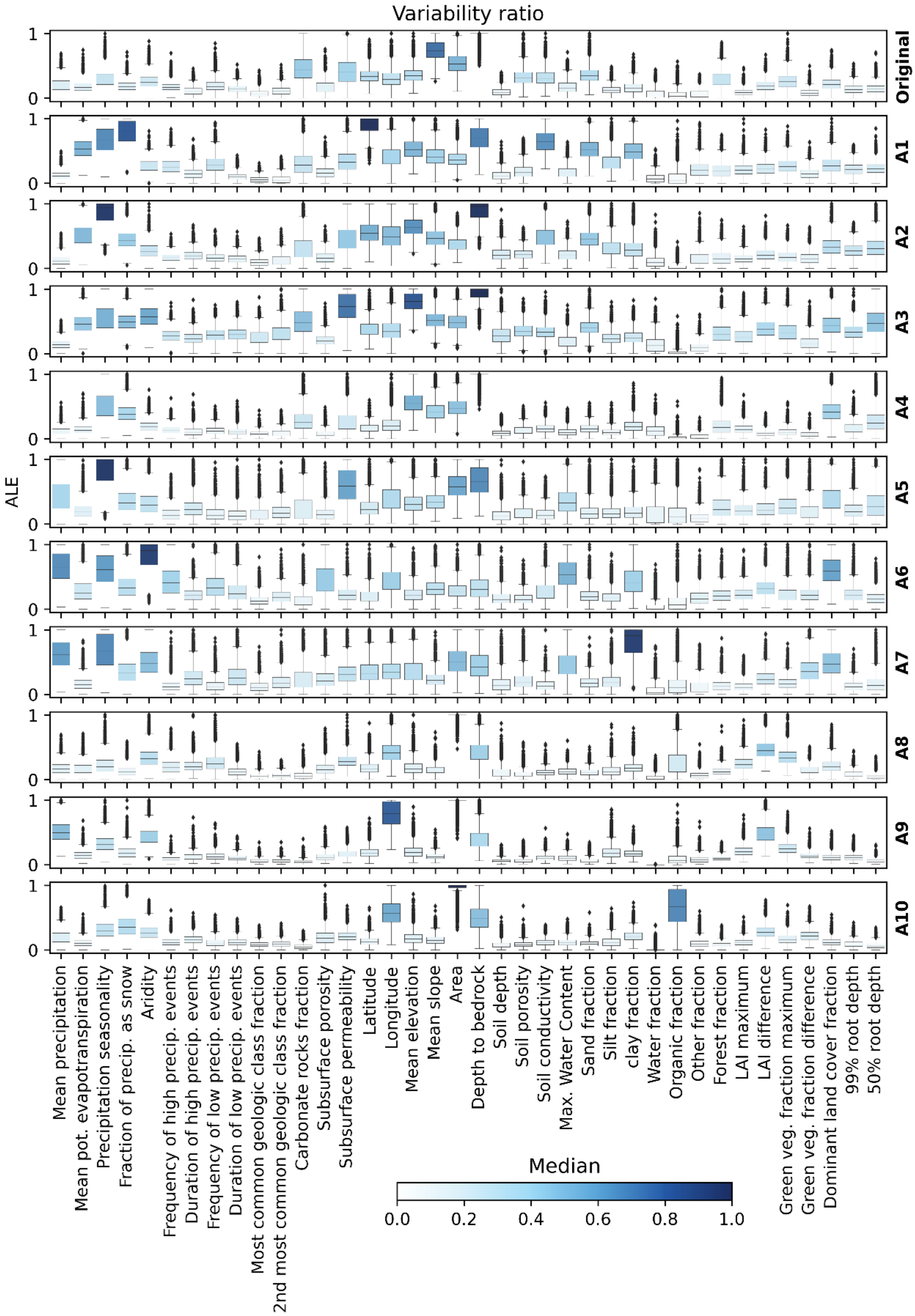

Figure S5. As for Figure S2 but for variability ratio.

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.