https://doi.org/10.5194/hess-2024-83
Preprint. Discussion started: 20 March 2024

Hydrology and Earth System Sciences
Discussions
Open Access

A decomposition approach to evaluating the local performance of global streamflow reanalysis

Tongtiegang Zhao[1], Zexin Chen[1], Yu Tian[2] and Xiaohong Chen[1]

[1] Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) and Key Laboratory for Water Security in the Guangdong-Hongkong-Macao Greater Bay Area, School of Civil Engineering, Sun Yat-Sen University, Guangzhou, China.
[2] State Key Laboratory of Simulation and Regulation of Water Cycle in River Basin, China Institute of Water Resource and Hydropower Research, Beijing China.

Correspondence to: Tongtiegang Zhao (zhaottg@mail.sysu.edu.cn) and Zexin Chen (chenzx33@mail2.sysu.edu.cn)

Abstract. While global streamflow reanalysis provides valuable information for water resources management, 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 raw 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 raw reanalysis and approximation components are primarily influenced by precipitation seasonality. That is, 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.

https://doi.org/10.5194/hess-2024-83
Preprint. Discussion started: 20 March 2024

Hydrology and Earth System Sciences
Discussions

1 Introduction

Global streamflow reanalysis provides valuable information for water resources management (Beck et al., 2017; Harrigan et al., 2020; Pokhrel et al., 2021). Generated by using climate reanalysis to drive global hydrological models (GHMs, Hersbach et al., 2020; Alfieri et al., 2020; Muñoz-Sabater et al., 2021), there exist multiple streamflow reanalysis datasets, e.g., the Global Flood Awareness System (GloFAS) within the European Centre for Medium-Range Weather Forecasts (ECMWF)'s latest global atmospheric reanalysis (GloFAS-ERA5, Harrigan et al., 2020), the Global Reach-Level A Priori Discharge Estimates for SWOT (GRADES, Lin et al., 2019) and the Global Reach-Level Flood Reanalysis (GRFR, Yang et al., 2021). In practice, streamflow reanalysis can bridge the data gaps for ungauged and poorly gauged catchments and provides estimates on a large spatial scale and with sufficient temporal resolution (Lin et al., 2019; Harrigan et al., 2020; Yang et al., 2021). For example, the recent GloFAS-ERA5 provides streamflow information at the daily time step and with a spatial resolution of 0.1° across the globe (Harrigan et al., 2020).

The local performance plays a critical part in practical applications of global streamflow reanalysis (Veldkamp et al., 2018; Munia et al., 2020; Feng et al., 2021). By comparing global reanalysis to observed streamflow, diagnostic plots and verification metrics are generated to showcase its local performance (Xie et al., 2019; Harrigan et al., 2020; Gao et al., 2020; Cantoni et al., 2022; Han et al., 2023; Liu et al., 2023). In the meantime, hydrological signatures derived from reanalysis are compared to those obtained from observed streamflow to facilitate insights into the effectiveness of hydrological models (Beck et al., 2017; Chen et al., 2021; Zhao et al., 2022). For example, the performances of ten Inter-Sectoral Impact Model Intercomparison Project (ISI-MIP) models are evaluated for low, mean and high flows using five streamflow percentile series (Chen et al., 2021). Considering limited observation data, streamflow reanalysis can serve as reference data to calibrate hydrological models and then the model outputs can be compared to observations to see whether practical applications are available (Senent-Aparicio et al., 2021).

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 (Apaydin et al., 2021). These components can be extracted by using some decomposition approaches (Nalley et al., 2012; 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, of which each is linked to some frequencies (Manikanta and Vema, 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).

While global streamflow reanalysis has been evaluated at different spatial scales (Harrigan et al., 2020; Senent-Aparicio et al., 2021; Chen et al., 2021), the time series characteristics of streamflow reanalysis in the time-frequency domain are yet

https://doi.org/10.5194/hess-2024-83
Preprint. Discussion started: 20 March 2024

Hydrology and Earth System Sciences
Discussions

to be investigated. Meanwhile, it is difficult to interpret the local performance of global streamflow reanalysis across different locations (Sichangi et al., 2016; Ghiggi et al., 2019; Tu et al., 2024), let alone the additional interpretation of the local performance at different timescales. This paper bridges the gap by presenting a novel evaluation of global streamflow reanalysis by combining the discrete wavelet transform (DWT) with machine learning techniques. That is, the DWT is employed to exploit streamflow reanalysis in the time-frequency domain; and then the accumulated local effects (ALEs) are derived by the random forest model to showcase the performance of raw reanalysis and its decomposed components at different scales. As will be demonstrated in the methods and results, streamflow reanalysis does exhibit different local performances at different timescales and the influences of catchment attributes are illustrated.

2 Methods

2.1 Time series decomposition

Both reanalysis and observed streamflow time series are decomposed into detail and approximation components using the wavelet transform (Chalise et al., 2023). Specifically, the utilization of wavelet transform involves the rigorous mathematical deconstruction of a signal into multiple lower resolution levels (Chong et al., 2019). This process is executed by controlling the scaling and shifting factors associated with a mother wavelet (Nalley et al., 2012). 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). Following Wei et al. (2012), the Daubechies wavelet of order 5 is used to decompose the streamflow time series. For a streamflow time series $q(t)$, the discrete wavelet transform is (Talukder et al., 2020):

$$W(a,b)=\sum_{t\in Z} q(t)\psi_{m,n}(t) \tag{1}$$

in which m and n are integers that respectively represent the amount of dilation and translation of the wavelet, t represents the discrete time and ψ represents the wavelet basis function (Nalley et al., 2012):

$$\psi_{m,n}(t)=2^{-\frac{m}{2}}\psi\left(2^{-m}t-n\right) \tag{2}$$

The DWT decomposes a signal into approximation (low-frequency) and detail (high-frequency) coefficients, thereby separating its frequency components based on magnitude (Quilty and Adamowski, 2021). In the initial decomposition that utilizes high-pass and low-pass filters and inverse discrete wavelet transform, the original signal is decomposed into detail component (D1) and approximation component (A1). Subsequently, the approximation component (A1) resulting from this initial stage is furthermore decomposed into D2 and A2, and so on for successive levels. This process is conducted from high-pass and low-pass filters followed by a down-sampling operator:

https://doi.org/10.5194/hess-2024-83
Preprint. Discussion started: 20 March 2024

Hydrology and Earth System Sciences Discussions

$$\left(q \downarrow 2\right)[t] = q[2t] \tag{3}$$

Therefore, streamflow time series is decomposed into the approximation components and detail components (Talukder et al., 2020):

$$\begin{cases} CA_l[t] = \sum_n L[n] q[2t+n] \\ CD_l[t] = \sum_n H[n] q[2t+n] \end{cases} \tag{4}$$

in which $CA_l[t]$ is approximation coefficient, $CD_l[t]$ is detail coefficient, l is decomposition level, L is low-pass filter and H is high-pass filter. The inverse discrete wavelet transform is used to obtain the detail components and approximation components (Guo et al., 2022):

$$\begin{cases} A_l = IDWT\left(CA_l[t]\right) \\ D_l = IDWT\left(CD_l[t]\right) \end{cases} \tag{5}$$

in which $IDWT$ is the inverse discrete wavelet transform, A_l is approximation component and D_l is detail component.

2.2 Verification after decomposition

The Kling-Gupta Efficiency (KGE) stands out as a widely utilized verification metric to evaluate the model performance (Frame et al., 2021; Huang and Zhao, 2022; Zhao et al., 2022). The KGE is utilized to indicate the performance of raw reanalysis, approximation and detail components. When evaluating the performance of raw reanalysis, the KGE is calculated as follows:

$$KGE = 1 - \sqrt{\left(r-1\right)^2 + \left(\beta-1\right)^2 + \left(\gamma-1\right)^2} \tag{6}$$

As can be seen, the KGE is comprised of three components, namely, the Pearson correlation coefficient r, the bias ratio β and the variability ratio γ:

$$r = \frac{\sum_{t=1}^{T}\left(d_t - \mu_d\right)\left(q_t - \mu_q\right)}{\sqrt{\sum_{t=1}^{T}\left(d_t - \mu_d\right)^2}\sqrt{\sum_{t=1}^{T}\left(q_t - \mu_q\right)^2}} \tag{7}$$

$$\beta = \frac{\mu_d}{\mu_q} \tag{8}$$

$$\gamma = \frac{\sigma_d}{\sigma_q} \tag{9}$$

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, respectively. The KGE ranges from $-\infty$ to 1, with a perfect value of 1.

To investigate the relationship between reanalysis and observations, it is necessary to extract the corresponding grid cell for each hydrometric station. The grid cell in which the hydrometric station is located may not overlap with the simulated river network in streamflow reanalysis due to the inaccuracy of the routing module in distributed hydrological model (Chen et al., 2021). There are three steps to identify the target cell: firstly, the initial cell is located according to the latitude and longitude of the hydrometric station; secondly, the KGE between reanalysis and observed streamflow is calculated for the initial cell and its eight surrounding cells; and finally, the cell with the largest KGE is used as the target cell (Zhao et al., 2022).

The hydrometric stations with outliers in terms of the KGE, correlation, bias ratio and variability ratio are excluded from the investigation. To facilitate the investigation of the influences of catchment attributes on performance, the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is used to remove the outliers of KGE and its three components. The DBSCAN offers a distinctive advantage in detecting outliers by defining clusters as dense regions separated by sparser areas (Smiti, 2020). This characteristic makes the algorithm effective in distinguishing outliers from the main clusters and particularly suited for anomaly detection (Li et al., 2022). There are two key parameters in the DBSCAN, including the maximum cluster radius (ε) and the minimum number of points (MinPts). Points within a distance ε are considered part of a dense region, while those with fewer than MinPts neighbors are treated as outliers. Following the study conducted by Brinkerhoff et al. (2020), the "elbow"-based approach is used to determine the ε and the MinPts is set to 5. By setting these parameters, the DBSCAN effectively identifies and isolates anomalies, promoting accurate anomaly removal while preserving the integrity of the main cluster structures (Hauswirth et al., 2021).

2.3 Influences of catchment attributes

The ALEs are derived by the random forest model to showcase the influences of catchment attributes on the performance of raw reanalysis and its approximation components at different scales. The random forest model is employed to establish a predictive relationship between the performance and multiple catchment attributes. This model is well-suited to capture complex relationships within the dataset through its ensemble of decision trees, which renders it an effective tool for performance prediction (Wei et al., 2023). To implement the model, the data is split into training and testing sets under the ratio of 75:25 (Naghibi et al., 2017). That is, 75% of catchments are randomly allocated for training and the remaining 25% for testing. The random forest model is trained on the training set, with tuning of hyperparameters to optimize its predictive capabilities. Following training, the model is validated on the test set and coefficient of determination (R^2) is calculated to assess its accuracy in predicting performance based on catchment attributes.

https://doi.org/10.5194/hess-2024-83
Preprint. Discussion started: 20 March 2024

Taking the KGE as an example, the prediction of the performance of approximation components for reanalysis using the random forest model is expressed as:

$$KGE = RF\left(X\right) \tag{10}$$

in which KGE is the predicted KGE, $RF\left(\bullet\right)$ is the random forest model and X is the catchment attributes. The R^2 between the KGE and the KGE is denoted by:

$$R^2 = \left(\frac{\sum_{i=1}^{N}\left(KGE_i - \mu_{KGE}\right)\left(KGE_i - \mu_{KGE}\right)}{\sqrt{\sum_{i=1}^{N}\left(KGE_i - \mu_{KGE}\right)^2}\sqrt{\sum_{i=1}^{N}\left(KGE_i - \mu_{KGE}\right)^2}} \right)^2 \tag{11}$$

in which μ is the mean KGE (KGE) and σ is the standard deviation of KGE (KGE).

The ALE is used to describe how catchment attributes influence the performance of approximation components at various scales for reanalysis based on the random forest model. An advantage of the ALE is that it overcomes the confounding effects of correlated catchment attributes (Stein et al., 2021). The ALE curves reveal the relationship between the performance and a specific catchment attribute, indicating whether the association is linear, monotonic or exhibits a more complex pattern (Teng et al., 2022). The uncentered ALE $\hat{\hat{f}}_{j,ALE}\left(x\right)$ is formulated as follows:

$$\hat{\hat{f}}_{j,ALE}\left(x\right) = \sum_{k=1}^{k_j}\frac{1}{n_j\left(k\right)}\sum_{i:x_j^{(i)}\in N_j(k)}\left[f\left(z_{k,j}, x_{-j}^{(i)}\right) - f\left(z_{k-1,j}, x_{-j}^{(i)}\right)\right] \tag{12}$$

in which x is the value of the catchment attribute j, k is one of k_j quantiles. By dividing the range of x, $n_j(k)$ is the number of x that in quantile $Nj(k)$, $z_{k,j}$ is the boundary values of x within that quantile, f is the output of the random forest model and $x_{-j}^{(i)}$ is the values of catchment attribute i except for j.

The ALE $\hat{f}_{j,ALE}\left(x\right)$ is derived from uncentered ALE values by subtracting its mean across all quantiles (Konapala et al., 2020):

$$\hat{f}_{j,ALE}\left(x\right) = \hat{\hat{f}}_{j,ALE}\left(x\right) - \frac{1}{k_j}\sum_{k=1}^{k_j}\hat{\hat{f}}_{j,ALE}\left(x_k\right) \tag{13}$$

Furthermore, the Local Interpretable Model-agnostic Explanations (LIMEs) elucidate individual predictions made by a trained black-box machine learning model (Jiang, 2022). The LIME is used to identify the dominant catchment attribute on performance of approximation components at various scales for each catchment.

A transformation is applied to the bias and variability ratios of raw reanalysis and its approximation components when investigating the influences of catchment attributes. The bias ratio and variability ratio are transformed as follows (Poncelet et al., 2017):

https://doi.org/10.5194/hess-2024-83
Preprint. Discussion started: 20 March 2024

Hydrology and Earth System Sciences
Open Access
Discussions

$$\begin{cases} \beta^* = 1 - |1 - \beta| \\ \gamma^* = 1 - |1 - \gamma| \end{cases} \tag{14}$$

in which β^* represents the bias ratio after transformation, γ^* is the variability ratio after transformation. This operation is owing to that increases of the values of bias and variability ratios do not necessarily indicate improved performance. After the transformation, both β^* and γ^* take the value of 1 to be maximum value that indicates the best performance. Notably, this transformation does not affect the ranking of performance among catchments.

3 Case study

3.1 Streamflow reanalysis

The GloFAS-ERA5 streamflow reanalysis v2.1 provides valuable hydrological time series forced by the latest global atmospheric reanalysis ERA5 (Harrigan et al., 2020). Developed jointly by the Joint Research Centre (JRC) of the European Commission, the University of Reading and the ECMWF (Harrigan et al., 2020), this streamflow reanalysis is generated by coupling the Hydrology Tiled ECMWF Scheme for Surface Exchanges over Land (HTESSEL) land surface model with the LISFLOOD hydrological and channel routing model (Alfieri et al., 2020; Harrigan et al., 2020). Specifically, the daily surface and subsurface runoff generated by the HTESSEL model are routed using the LISFLOOD model (Harrigan et al., 2020). The GloFAS-ERA5 provides a spatial resolution of 0.1º at a daily time step covering the time period from 1 January 1979 to near real time (Harrigan et al., 2020). Harrigan et al. (2020) found the GloFAS-ERA5 streamflow reanalysis to exhibit skill in 86% of tested catchments, but noted considerable variability in skill across locations, including significant positive biases in regions such as the central United States and Africa.

3.2 Observed streamflow

The observed streamflow is sourced from the Catchment Attributes and Meteorology for Large-sample Studies (CAMELS) dataset (Newman et al., 2015; Addor et al., 2017). An advantage of this dataset is its sufficient streamflow time series from 1980 to 2015, which provides valuable reference data for the evaluation of streamflow reanalysis (Addor et al., 2017). This dataset provides streamflow data for 671 catchments across the continental United States (CONUS), which exhibit diverse hydro-meteorological characteristics. Notably, these catchments are primarily located at headwaters, resulting in minimal influence from human activities (Stein et al., 2021). In the meantime, the CAMELS provides information on six categories of catchment attributes, including climate, geology, topography, soil, vegetation and streamflow indices (Addor et al., 2017; Stein et al., 2021). Categorical attributes are not used in the investigation of the influences on model performance

https://doi.org/10.5194/hess-2024-83
Preprint. Discussion started: 20 March 2024

(Stein et al., 2021). The influences of catchment attributes on performance of streamflow time series characteristics are investigated using 38 attributes across five categories: climate, geology, topography, soil and vegetation.

To facilitate the evaluation of streamflow reanalysis, the stations whose data length meets the requirement for decomposition into 10 levels are selected (Nalley et al., 2012). The maximum decomposition level l_m is denoted by:

$$l_m = \frac{\log\left(\frac{N}{2v-1}\right)}{\log(2)} \tag{15}$$

in which v represents the number of vanishing moments of the Daubechies wavelet, set to 5, N is the number of data points. Specifically, 661 stations with a data length exceeding 9216 days are selected for the investigation.

4 Results

4.1 Approximation and detail components

The time series of streamflow reanalysis and observation along with their approximation components are presented in Figure 1. The plots are for the station 6224000 in which raw reanalysis tends to exhibit the highest KGE value of 0.82. For the approximation components between reanalysis and observation, the KGE values are evaluated and illustrated by heatmap. It can be observed that raw reanalysis generally captures the primary features of the streamflow time series. Under the stepwise decomposition of the streamflow time series, the KGE increases from 0.48 for A1 to 0.89 for A6. 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 components becomes smoother. As the decomposition level increases, the reanalysis becomes more able to capture the information in observation. That is, reanalysis can provide more valuable information for seasonal and annual features. The KGE values between approximation components of reanalysis and observation are higher when the scales match, suggesting streamflow reanalysis can be evaluated by the wavelet transform.

https://doi.org/10.5194/hess-2024-83
Preprint. Discussion started: 20 March 2024

Hydrology and Earth System Sciences
Open Access
Discussions

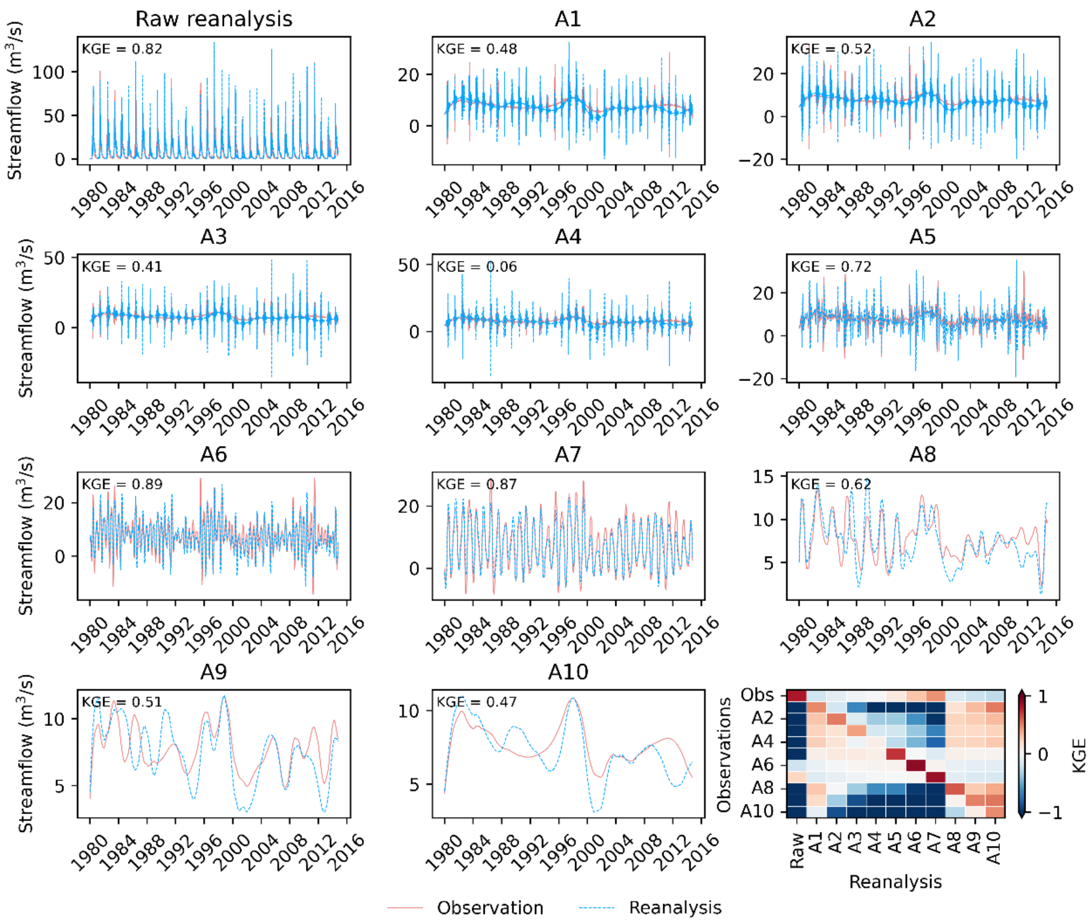
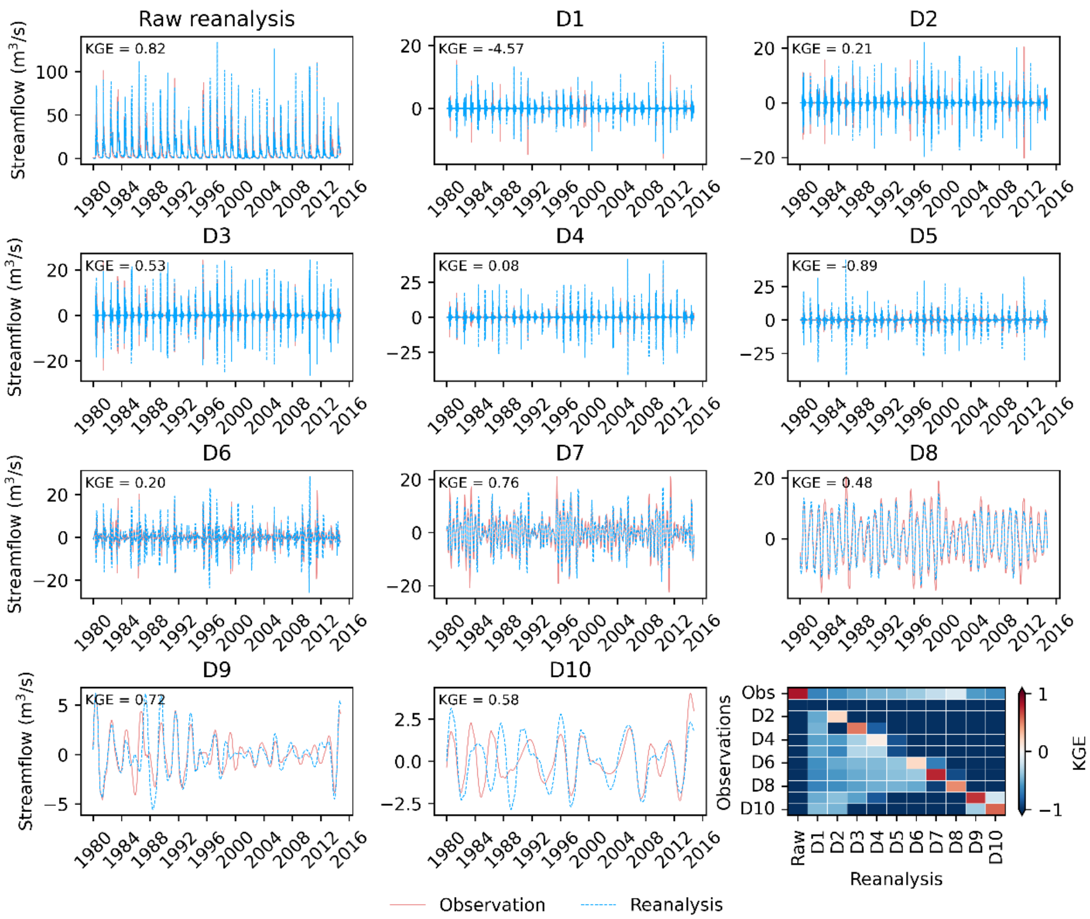


Figure 1: Time series plots of raw reanalysis and its approximation components and heatmap of the KGE between approximation components of reanalysis and observations for the station 6224000.

The performance of raw reanalysis and its detail components for the station 6224000 are illustrated in Figure 2 through eleven time series plots and one heatmap. As the decomposition level increases, it can be observed that the series of detail components becomes smoother. In the meantime, there is an increasing trend in KGE from D1 to D10, indicating improved performance with increasing timescales. The comparison of Figure 2 with Figure 1 suggests that the performance of approximation components is generally better than that of detail components. In other words, the detail components are more difficult to be characterized than the approximation components. Focusing on the heatmap, it can be observed that the KGE along the diagonal is relatively high, suggesting reasonable agreement. That is, the detail components of observation that do not correspond in scale cannot be accurately matched by streamflow reanalysis.



215 **Figure 2:** As for Figure 1 but for the detail components.

4.2 Performance across the CONUS

The KGE values of raw reanalysis and its approximation components after removing the outliers using the DBSCAN are presented in Figure 3. In total, there are 11 spatial plots for raw reanalysis and its components after decomposition. 220 Positive KGE values are marked in red and negative values in blue. It can be observed that raw reanalysis exhibits the highest KGE in the western United States, with comparatively poorer performance in the central United States. These findings are consistent with those of Addor et al. (2017), indicating poor performances in the high plains and desert southwest. Similarly, the approximation components from A1 to A10 exhibit the highest KGE in the western United States and the relatively lower KGE in the central United States. This finding indicates that the KGE values of approximation 225 components are related to the KGE values of raw reanalysis. 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.

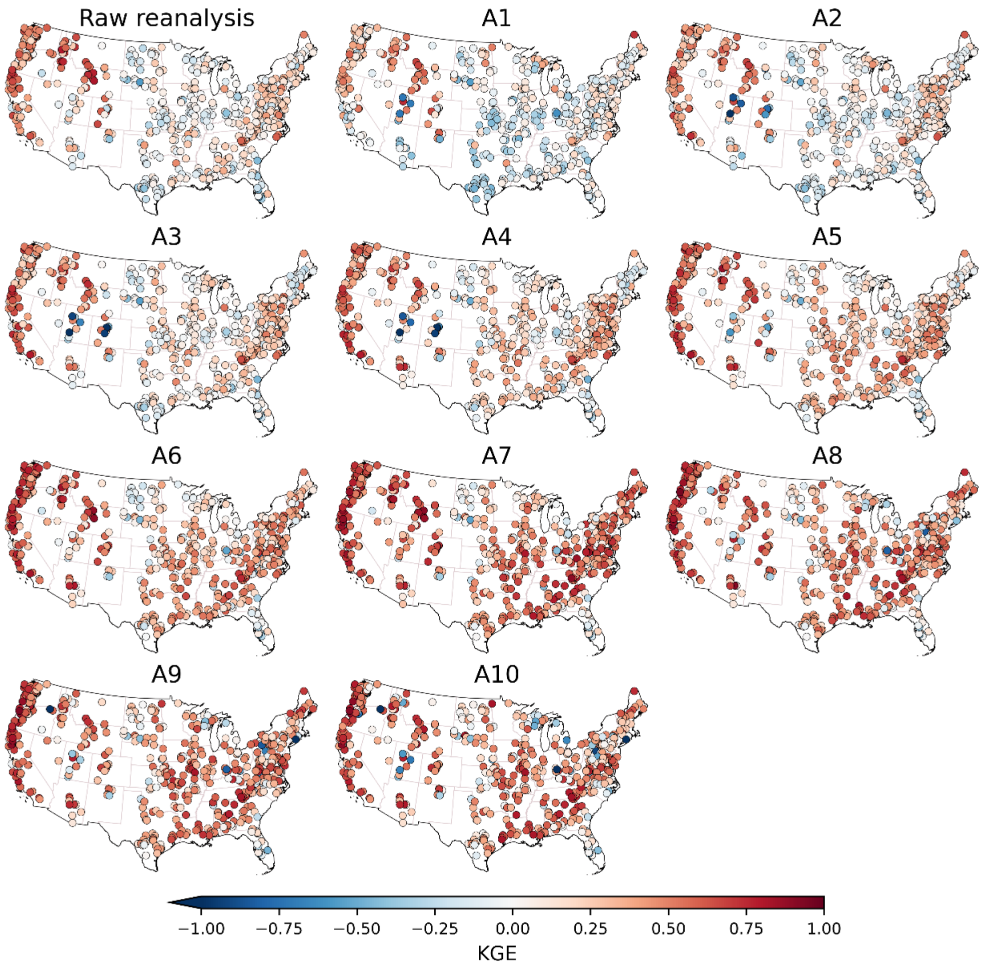


Figure 3: Spatial distribution of the KGE values of raw reanalysis and its approximation components from A1 to A10.

The performances of raw reanalysis and its approximation components across 554 catchments in the CONUS are shown in Figure 4. For the KGE between streamflow reanalysis and observations, it can be observed that the local performance of streamflow reanalysis generally improves from A1 to A7 and then remains promising from A8 to A10. Specifically, the median value of KGE is 0.02 for A1, 0.09 for A2, 0.19 for A3, 0.24 for A4, 0.29 for A5, 0.36 for A6, 0.47 for A7, 0.43 for A8, 0.42 for A9 and 0.40 for A10. This trend is due to the correlation and variability ratio tend towards 1 from A1 to A7. Meanwhile, the performance of A7 is better than that of raw reanalysis, suggesting that errors in raw reanalysis primarily stem from daily, weekly and monthly components. Focusing on the correlation, the medians of correlation for approximation components exceed 0.2, implying valuable information in multiple timescales approximations. Furthermore, the bias ratio remains nearly constant at each scale for approximation components. That is, the mean values of approximation components are generally similar to the mean values of the raw reanalysis.

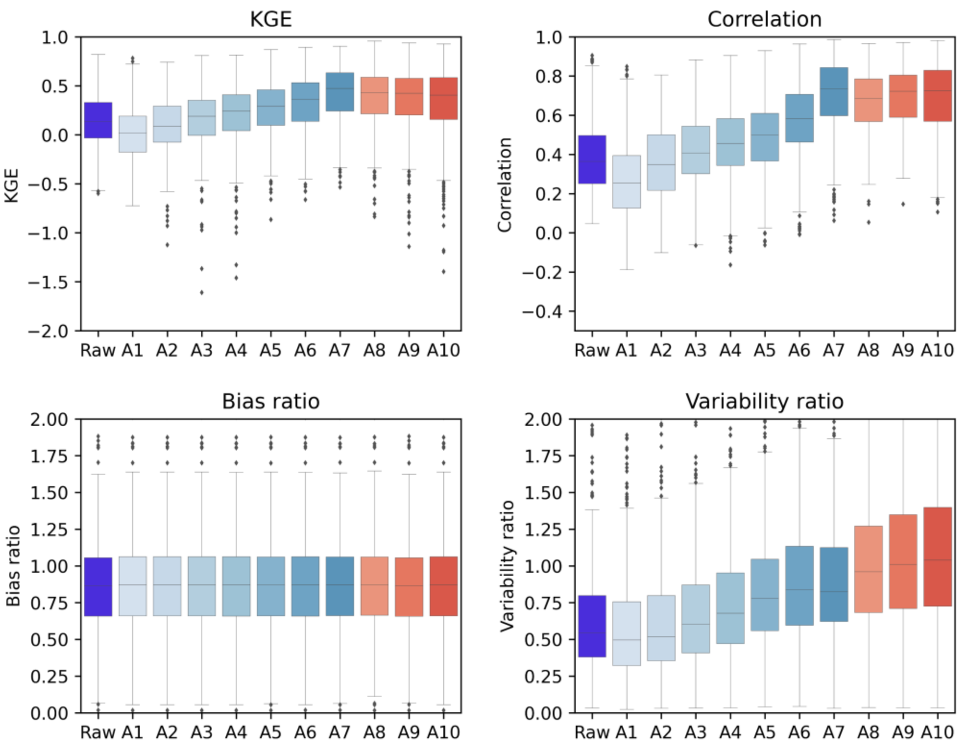


Figure 4: Boxplots of the KGE and its three components for raw reanalysis and its approximation components across 554 catchments in the CONUS. The lines within the boxes mark the median values. The boxes illustrate the interquartile range (IQR), where the lower and upper boundaries of the boxes respectively indicate the lower quartile (Q1) and upper quartile (Q3). The lower and upper whiskers show the smallest and largest values within the range of Q1-1.5IQR to Q3+1.5IQR. Dark grey diamonds represent outliers that lie beyond the whiskers.

4.3 Influences of catchment attributes on performance

The influences of catchment attributes on the KGE and its three components are measured by the mean absolute ALE and illustrated in Figure 5. From the first row, it can be observed that the KGE values of raw reanalysis and its approximation components are primarily influenced by precipitation seasonality. Positive (negative) values of precipitation seasonality indicate that precipitation peaks in summer (winter). That is, the season with more precipitation has a significant impact on the KGE. Longitude and mean slope also have a significant impact on the KGE across raw reanalysis and daily,

https://doi.org/10.5194/hess-2024-83
Preprint. Discussion started: 20 March 2024

weekly and monthly features (from A1 to A5). 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. This result suggests that the influences of catchment attributes on correlation of annual and multi-annual features are different from daily, weekly and monthly features. 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 forecasts.

https://doi.org/10.5194/hess-2024-83
Preprint. Discussion started: 20 March 2024

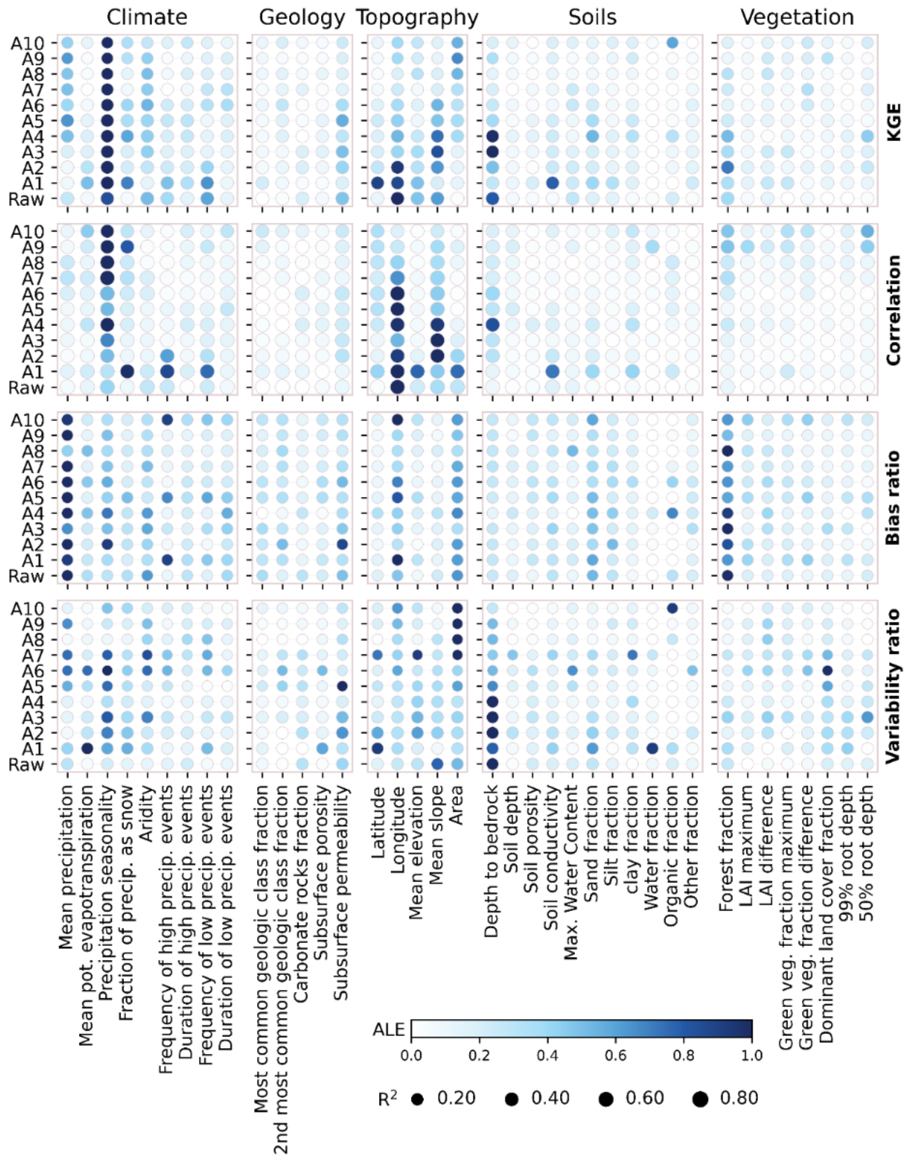


Figure 5: The ALE of the catchment attributes on the KGE, correlation, bias ratio and variability ratio. The color denotes the mean absolute values for each ALE curve, which is normalized for each raw reanalysis (approximation component). The sizes of point represent prediction accuracy indicated by R^2 for the random forest model using testing set. The "Raw" represents "raw reanalysis".

To further illustrate how catchment attributes affect the performances of raw reanalysis and its approximation components, the ALE curves are presented for three influential attributes, namely, precipitation seasonality, mean precipitation and mean slope of catchment. The influences of precipitation seasonality on the KGE and its three components are presented in Figure 6. It can be observed that the relationships between the KGE and precipitation seasonality are generally nonlinear. The KGE gradually decreases with the increasing precipitation seasonality. That is, the KGE values are notably low when precipitation tends to concentrate in summer and turn out to be high when precipitation tends to concentrate in winter. The ALE curves of the daily, weekly and monthly features (from A1 to A5) are similar to raw reanalysis, sharply decreasing around -0.5. The seasonal, annual and multi-annual features (from A6 to A10) sharply decrease around 0. In the meantime, the influences of precipitation seasonality on the correlation, bias and variability ratios are similar to that on the KGE. These results can be due to low evaporation in winter that results in reasonable simulations of soil moisture and baseflow processes (Poncelet et al., 2017).

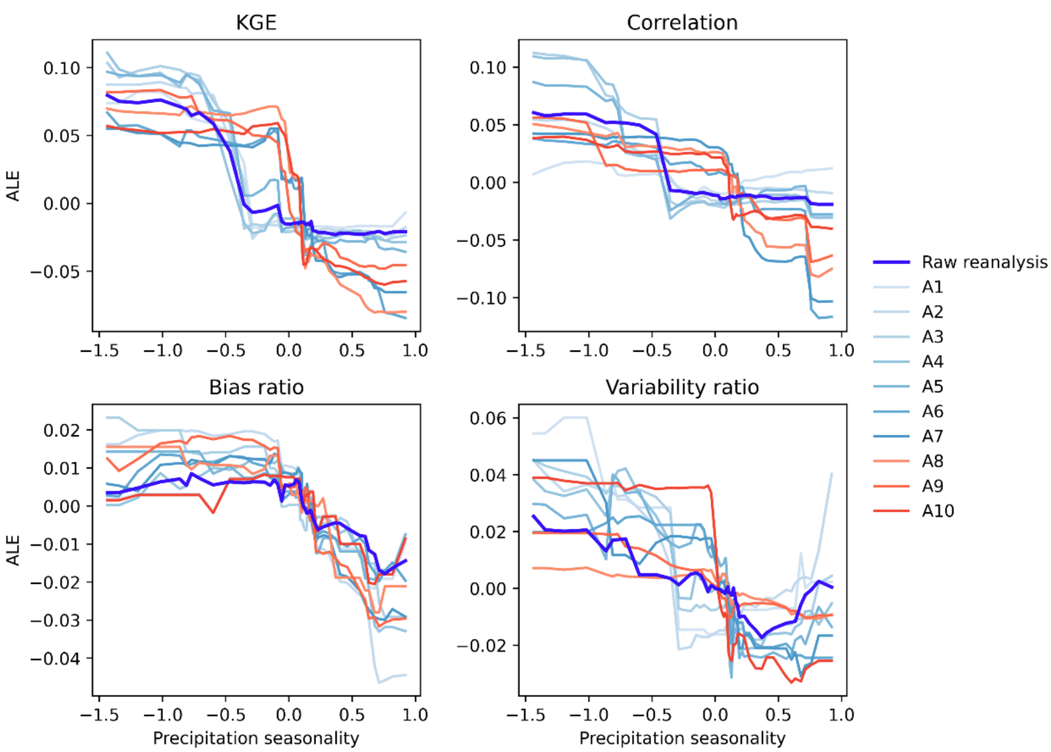


Figure 6: The ALE curves of precipitation seasonality on the KGE, correlation, bias ratio and variability ratio for raw reanalysis and its approximation components.

The influences of mean precipitation on the KGE, correlation, bias ratio and variability ratio across different scales are illustrated in Figure 7. The mean precipitation has a positive effect on the KGE of raw reanalysis and its approximation components, with a nonlinear increase of the KGE with rising mean precipitation, particularly for the annual and multi-annual features. In the meantime, it affects the correlation, bias ratio and variability ratio of raw reanalysis positively. This result suggests that mean precipitation has a consistent influence on the KGE, correlation, bias and variability ratios for

https://doi.org/10.5194/hess-2024-83
Preprint. Discussion started: 20 March 2024

approximation components, leading to differences in the KGE values across various catchments. These results can be due to the fact that rainfall-runoff processes are more linear in humid catchments than in arid catchments, leading to less variability in hydrologic states and facilitating more accurate simulations (Parajka et al., 2013).

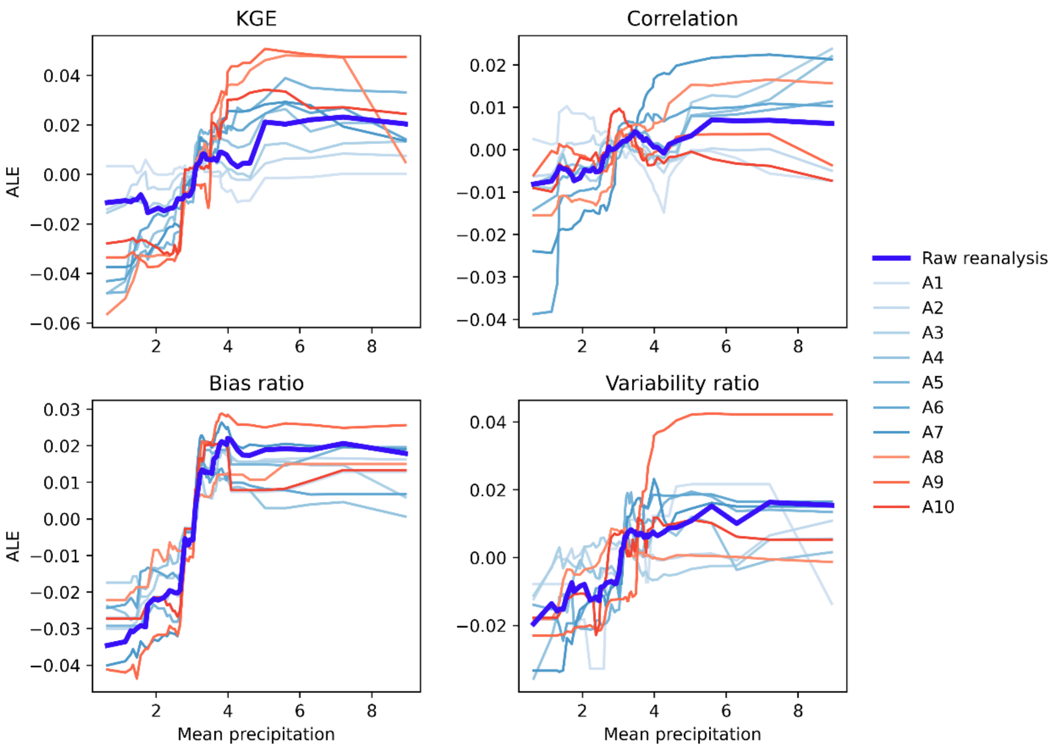
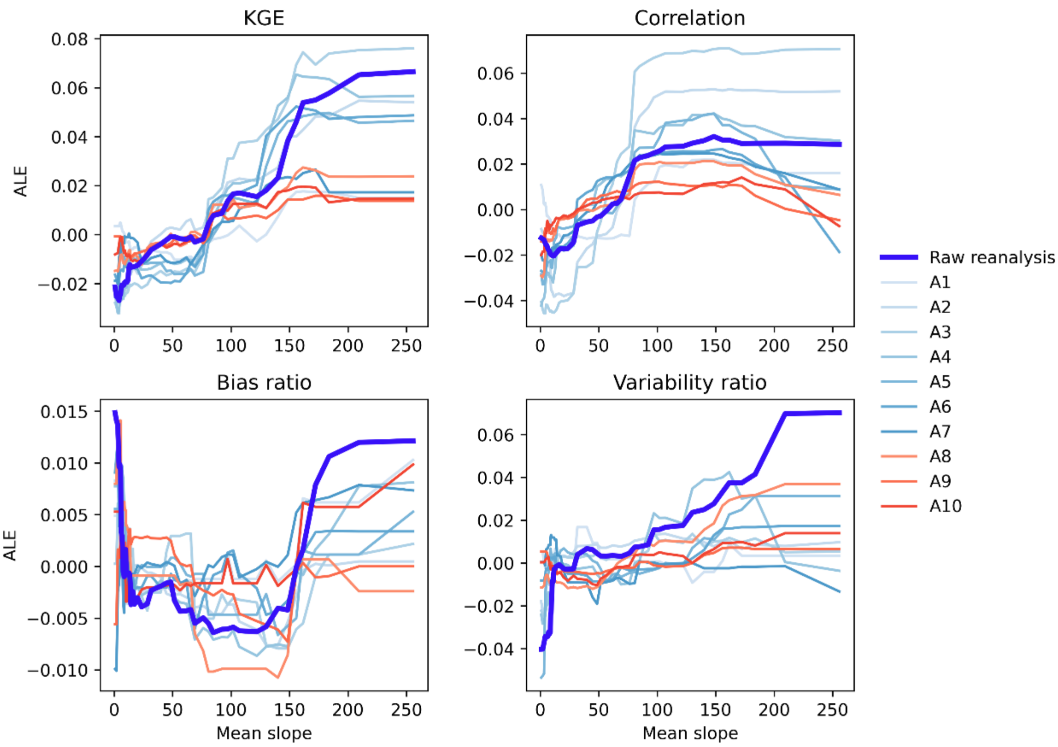


Figure 7: As for Figure 6 but for mean precipitation.

The influences of mean slope on the KGE and its three components across different scales are shown in Figure 8. It can be observed that there is a nonlinear relationship between the KGE and mean slope of catchment. As the mean slope increases, the KGE of raw reanalysis and its approximation components tend to increases. This result may be due to the mean slope of catchment affecting the simulation of runoff generation and infiltration (Massmann, 2020; Stein et al., 2021). It is noted that the KGE values of approximation components gradually increase when the mean slope of catchment surpasses 150. In particular, the correlation and variability ratio of raw reanalysis generally increase with the increase in the KGE. That is, the mean slope of catchment has a similar effect on the KGE, correlation and variability ratio. On the other hand, bias ratio decreases initially and then increases with the increase in mean slope of catchment. In other words, the relationship between bias ratio and mean slope of catchment is nonmonotonic.

https://doi.org/10.5194/hess-2024-83
Preprint. Discussion started: 20 March 2024

Hydrology and Earth System Sciences
Open Access
EGU
Discussions



310 **Figure 8:** As for Figure 6 but for mean slope.

4.4 Driving factors of each catchment

The most important attribute that influences the KGE is identified for each catchment by the LIME method and then illustrated by spatial plots in Figure 9. It can be observed that the most important attributes influencing the KGE exhibit regional clustering. The KGE of raw reanalysis is primarily influenced by precipitation seasonality in the western and central United States while by depth to bedrock in the eastern United States (Pfister et al., 2017; Addor et al., 2017). That is, the substantial differences in precipitation seasonality between the western and central United States result in significant differences in the KGE (Figure 3). On the other hand, the most important attribute controlling the KGE of approximation components is different from that of raw reanalysis. It can be observed that the KGE values of approximation components from A6 to A8 are primarily controlled by precipitation seasonality in the eastern United States, while raw reanalysis is controlled by depth to bedrock. The higher depth to bedrock may exhibit larger storage values, consequently leading to higher baseflow (Pfister et al., 2017). In the meantime, the number of catchments controlled by precipitation seasonality tends to increase from A1 to A8, with a high proportion observed in A6, A7 and A8. That is, the performance of the annual variability of streamflow reanalysis is influenced by precipitation seasonality.

https://doi.org/10.5194/hess-2024-83
Preprint. Discussion started: 20 March 2024

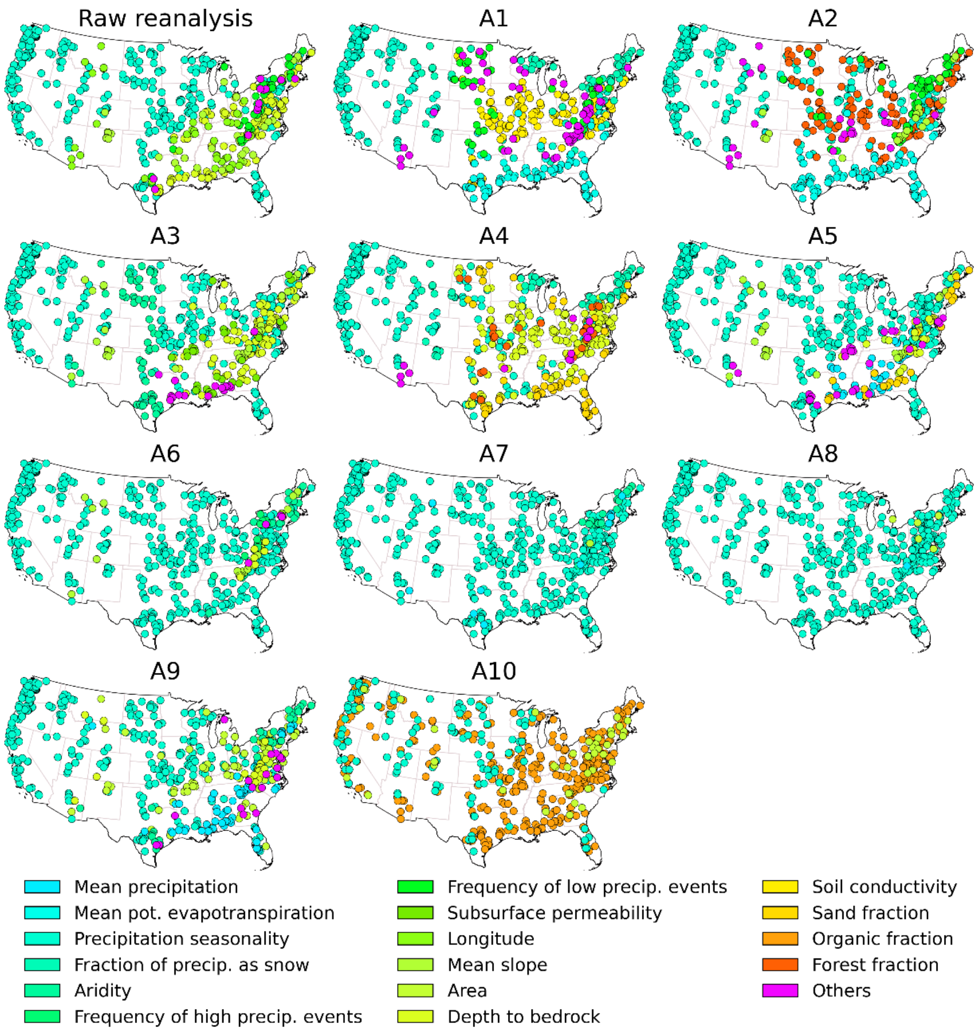


Figure 9: Spatial patterns of the controlling catchment attribute on the KGE of raw reanalysis and approximation components for each catchment. For each spatial distribution map, if there are more than five catchment attributes, only the top five attributes are presented, while the rest are labelled as others.

5 Discussion

Global streamflow reanalysis provides valuable information for water resources management (Alfieri et al., 2020; Harrigan et al., 2020; Yang et al., 2021). Building upon previous studies evaluating the performance of hydrological signatures derived from reanalysis and observed streamflow (Beck et al., 2017; Chen et al., 2021; Tu et al., 2024), this paper presents a novel evaluation by combining the wavelet transform with machine learning. Specifically, streamflow reanalysis

https://doi.org/10.5194/hess-2024-83
Preprint. Discussion started: 20 March 2024

and observation are respectively decomposed by the DWT into detail and approximation components at different scales. As a result, streamflow characteristics in the time-frequency domain are unravelled by extracting features and removing noise from the original signal (Manikanta and Vema, 2022). This approach provides a new perspective by paying attention to the difference between global streamflow reanalysis and observed streamflow in the time-frequency domain. The KGE generally indicates that streamflow reanalysis exhibits a robust capability to capture the information of seasonal, annual and multi-annual variability, particularly the annual fluctuations. This result suggests that hydrological simulations at daily or even hourly timescales are more challenging.

Hydrological models generally exhibit different performance across different catchments (Newman et al., 2015; O'Neill et al., 2021; Tu et al., 2024). The differences can be related to heterogeneous streamflow patterns under unique combinations of climate and catchment attributes (Jehn et al., 2020; Stein et al., 2021). Previous studies have found that model performance is related to aridity index, with generally better performance in wetter catchments compared to drier ones (Poncelet et al., 2017). In addition to aridity index, other factors are also linked to the model performance, such as impact of snow (Newman et al., 2015), catchment area (Harrigan et al., 2020), precipitation intermittency (Newman et al., 2015) and human activities (Veldkamp et al., 2018). In this paper, it is found that the KGE values of raw reanalysis and approximation components are primarily influenced by precipitation seasonality. This outcome can be due to lower evaporation in winter, when the soil moisture is higher and baseflow can be better simulated (Poncelet et al., 2017). On the other hand, the relationships between KGE and catchment attributes are nonlinear. The results highlight that the wavelet transform can facilitate the evaluation of the local performance of global streamflow reanalysis to provide more effective information.

6 Conclusions

This paper has presented a novel evaluation of global streamflow reanalysis by combining the widely used wavelet transform and machine learning. Specifically, the raw reanalysis and observed streamflow are decomposed by the DWT and then they are used to indicate the local performance of the time series characteristics in the time-frequency domain. Furthermore, the influences of catchment attributes on the performance of raw reanalysis and its approximation components at various scales are investigated using the ALE. A large-sample test is conducted for the CAMELS dataset so as to evaluate the effectiveness of GloFAS streamflow reanalysis. The results show that the streamflow reanalysis tends to characterize seasonal, annual and multi-annual variabilities better than daily, weekly and monthly variabilities. Precipitation seasonality is identified to be the most important attribute influencing the KGE of raw reanalysis and its approximation components using the ALE. The longitude, mean precipitation and mean slope also influence the performance of approximation components. On the other hand, the attributes on geology, soils and vegetation seem to have a relatively minor influence on the performance of approximation components. Overall, global streamflow reanalysis can be evaluated at different timescales using decomposition approaches to facilitate its practical applications.

Data availability

The GloFAS-ERA5 streamflow reanalysis v2.1 is downloaded from the Climate Data Store – Copernicus (https://cds.climate.copernicus.eu/). The CAMELS dataset is sourced from The National Center for Atmospheric Research (https://gdex.ucar.edu/dataset/camels.html).

Author contribution

TZ and ZC designed the experiments. ZC and YT carried them out. TZ and ZC developed the model code and performed the experiments. ZC, TZ and XC prepared the manuscript.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

This research is supported by the Ministry of Science and Technology of China (2023YFF0804900), the National Natural Science Foundation of China (52379033) and the Guangdong Provincial Department of Science and Technology (2019ZT08G090).

References

Abebe, S.A., Qin, T., Zhang, X., Yan, D., 2022. Wavelet transform-based trend analysis of streamflow and precipitation in Upper Blue Nile River basin. Journal of Hydrology: Regional Studies 44, 101251. https://doi.org/10.1016/j.ejrh.2022.101251

Addor, N., Newman, A.J., Mizukami, N., Clark, M.P., 2017. The CAMELS data set: catchment attributes and meteorology for large-sample studies. Hydrology and Earth System Sciences 21, 5293–5313. https://doi.org/10.5194/hess-21-5293-2017

Alfieri, L., Lorini, V., Hirpa, F.A., Harrigan, S., Zsoter, E., Prudhomme, C., Salamon, P., 2020. A global streamflow reanalysis for 1980–2018. Journal of Hydrology X 6, 100049. https://doi.org/10.1016/j.hydroa.2019.100049

Apaydin, H., Taghi Sattari, M., Falsafian, K., Prasad, R., 2021. Artificial intelligence modelling integrated with Singular Spectral analysis and Seasonal-Trend decomposition using Loess approaches for streamflow predictions. Journal of Hydrology 600, 126506. https://doi.org/10.1016/j.jhydrol.2021.126506

Beck, H.E., van Dijk, A.I.J.M., de Roo, A., Dutra, E., Fink, G., Orth, R., Schellekens, J., 2017. Global evaluation of runoff from 10 state-of-the-art hydrological models. Hydrology and Earth System Sciences 21, 2881–2903. https://doi.org/10.5194/hess-21-2881-2017

Brinkerhoff, C.B., Gleason, C.J., Feng, D., Lin, P., 2020. Constraining Remote River Discharge Estimation Using Reach-Scale Geomorphology. Water Resources Research 56, e2020WR027949. https://doi.org/10.1029/2020WR027949

Cantoni, E., Tramblay, Y., Grimaldi, S., Salamon, P., Dakhlaoui, H., Dezetter, A., Thiemig, V., 2022. Hydrological performance of the ERA5 reanalysis for flood modeling in Tunisia with the LISFLOOD and GR4J models. Journal of Hydrology: Regional Studies 42, 101169. https://doi.org/10.1016/j.ejrh.2022.101169

Chalise, D.R., Sankarasubramanian, A., Olden, J.D., Ruhi, A., 2023. Spectral Signatures of Flow Regime Alteration by Dams Across the United States. Earth's Future 11, e2022EF003078. https://doi.org/10.1029/2022EF003078

Chen, H., Liu, J., Mao, G., Wang, Z., Zeng, Z., Chen, A., Wang, K., Chen, D., 2021. Intercomparison of ten ISI-MIP models in simulating discharges along the Lancang-Mekong River basin. Science of The Total Environment 765, 144494. https://doi.org/10.1016/j.scitotenv.2020.144494

Chong, K.L., Lai, S.H., El-Shafie, A., 2019. Wavelet Transform Based Method for River Stream Flow Time Series Frequency Analysis and Assessment in Tropical Environment. Water Resour Manage 33, 2015–2032. https://doi.org/10.1007/s11269-019-02226-7

Feng, D., Gleason, C.J., Lin, P., Yang, X., Pan, M., Ishitsuka, Y., 2021. Recent changes to Arctic river discharge. Nat Commun 12, 6917. https://doi.org/10.1038/s41467-021-27228-1

Frame, J.M., Kratzert, F., Raney II, A., Rahman, M., Salas, F.R., Nearing, G.S., 2021. Post-Processing the National Water Model with Long Short-Term Memory Networks for Streamflow Predictions and Model Diagnostics. JAWRA Journal of the American Water Resources Association 57, 885–905. https://doi.org/10.1111/1752-1688.12964

Gao, H., Dong, J., Chen, X., Cai, H., Liu, Z., Jin, Z., Mao, D., Yang, Z., Duan, Z., 2020. Stepwise modeling and the importance of internal variables validation to test model realism in a data scarce glacier basin. Journal of Hydrology 591, 125457. https://doi.org/10.1016/j.jhydrol.2020.125457

https://doi.org/10.5194/hess-2024-83
Preprint. Discussion started: 20 March 2024

Hydrology and Earth System Sciences
Discussions

Ghiggi, G., Humphrey, V., Seneviratne, S.I., Gudmundsson, L., 2019. GRUN: an observation-based global gridded runoff dataset from 1902 to 2014. Earth System Science Data 11, 1655–1674. https://doi.org/10.5194/essd-11-1655-2019

Guo, J., Sun, H., Du, B., 2022. Multivariable Time Series Forecasting for Urban Water Demand Based on Temporal Convolutional Network Combining Random Forest Feature Selection and Discrete Wavelet Transform. Water Resour Manage 36, 3385–3400. https://doi.org/10.1007/s11269-022-03207-z

Han, J., Miao, C., Gou, J., Zheng, H., Zhang, Q., Guo, X., 2023. A new daily gridded precipitation dataset for the Chinese mainland based on gauge observations. Earth System Science Data 15, 3147–3161. https://doi.org/10.5194/essd-15-3147-2023

Harrigan, S., Zsoter, E., Alfieri, L., Prudhomme, C., Salamon, P., Wetterhall, F., Barnard, C., Cloke, H., Pappenberger, F., 2020. GloFAS-ERA5 operational global river discharge reanalysis 1979–present. Earth System Science Data 12, 2043–2060. https://doi.org/10.5194/essd-12-2043-2020

Hauswirth, S.M., Bierkens, M.F.P., Beijk, V., Wanders, N., 2021. The potential of data driven approaches for quantifying hydrological extremes. Advances in Water Resources 155, 104017. https://doi.org/10.1016/j.advwatres.2021.104017

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., Thépaut, J.-N., 2020. The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society 146, 1999–2049. https://doi.org/10.1002/qj.3803

Huang, Z., Zhao, T., 2022. Predictive performance of ensemble hydroclimatic forecasts: Verification metrics, diagnostic plots and forecast attributes. WIREs Water 9, e1580. https://doi.org/10.1002/wat2.1580

Jehn, F.U., Bestian, K., Breuer, L., Kraft, P., Houska, T., 2020. Using hydrological and climatic catchment clusters to explore drivers of catchment behavior. Hydrology and Earth System Sciences 24, 1081–1100. https://doi.org/10.5194/hess-24-1081-2020

Jiang, E., 2022. UniformLIME: A Uniformly Perturbed Local Interpretable Model-Agnostic Explanations Approach for Aerodynamics. J. Phys.: Conf. Ser. 2171, 012025. https://doi.org/10.1088/1742-6596/2171/1/012025

Joo, T.W., Kim, S.B., 2015. Time series forecasting based on wavelet filtering. Expert Systems with Applications 42, 3868–3874. https://doi.org/10.1016/j.eswa.2015.01.026

Konapala, G., Kao, S.-C., Painter, S.L., Lu, D., 2020. Machine learning assisted hybrid models can improve streamflow simulation in diverse catchments across the conterminous US. Environ. Res. Lett. 15, 104022. https://doi.org/10.1088/1748-9326/aba927

Lane, S.N., 2007. Assessment of rainfall-runoff models based upon wavelet analysis. Hydrological Processes 21, 586–607. https://doi.org/10.1002/hyp.6249

Li, Z., Gao, S., Chen, M., Gourley, J.J., Hong, Y., 2022. Spatiotemporal Characteristics of US Floods: Current Status and Forecast Under a Future Warmer Climate. Earth's Future 10, e2022EF002700. https://doi.org/10.1029/2022EF002700

Lin, P., Pan, M., Beck, H.E., Yang, Y., Yamazaki, D., Frasson, R., David, C.H., Durand, M., Pavelsky, T.M., Allen, G.H., Gleason, C.J., Wood, E.F., 2019. Global Reconstruction of Naturalized River Flows at 2.94 Million Reaches. Water Resources Research 55, 6499–6516. https://doi.org/10.1029/2019WR025287

Liu, L., Zhou, L., Gusyev, M., Ren, Y., 2023. Unravelling and improving the potential of global discharge reanalysis dataset in streamflow estimation in ungauged basins. Journal of Cleaner Production 419, 138282. https://doi.org/10.1016/j.jclepro.2023.138282

Manikanta, V., Vema, V.K., 2022. Formulation of Wavelet Based Multi-Scale Multi-Objective Performance Evaluation (WMMPE) Metric for Improved Calibration of Hydrological Models. Water Resources Research 58, e2020WR029355. https://doi.org/10.1029/2020WR029355

Massmann, C., 2020. Identification of factors influencing hydrologic model performance using a top-down approach in a large number of U.S. catchments. Hydrological Processes 34, 4–20. https://doi.org/10.1002/hyp.13566

Munia, H.A., Guillaume, J.H.A., Wada, Y., Veldkamp, T., Virkki, V., Kummu, M., 2020. Future Transboundary Water Stress and Its Drivers Under Climate Change: A Global Study. Earth's Future 8, e2019EF001321. https://doi.org/10.1029/2019EF001321

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D.G., Piles, M., Rodríguez-Fernández, N.J., Zsoter, E., Buontempo, C., Thépaut, J.-N., 2021. ERA5-Land: a state-of-the-art global reanalysis dataset for land applications. Earth System Science Data 13, 4349–4383. https://doi.org/10.5194/essd-13-4349-2021

Naghibi, S.A., Ahmadi, K., Daneshi, A., 2017. Application of Support Vector Machine, Random Forest, and Genetic Algorithm Optimized Random Forest Models in Groundwater Potential Mapping. Water Resour Manage 31, 2761–2775. https://doi.org/10.1007/s11269-017-1660-3

Nalley, D., Adamowski, J., Khalil, B., 2012. Using discrete wavelet transforms to analyze trends in streamflow and precipitation in Quebec and Ontario (1954–2008). Journal of Hydrology 475, 204–228. https://doi.org/10.1016/j.jhydrol.2012.09.049

Newman, A.J., Clark, M.P., Sampson, K., Wood, A., Hay, L.E., Bock, A., Viger, R.J., Blodgett, D., Brekke, L., Arnold, J.R., Hopson, T., Duan, Q., 2015. Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: data set characteristics and assessment of regional variability in hydrologic model performance. Hydrology and Earth System Sciences 19, 209–223. https://doi.org/10.5194/hess-19-209-2015

O'Neill, M.M.F., Tijerina, D.T., Condon, L.E., Maxwell, R.M., 2021. Assessment of the ParFlow–CLM CONUS 1.0 integrated hydrologic model: evaluation of hyper-resolution water balance components across the contiguous United States. Geoscientific Model Development 14, 7223–7254. https://doi.org/10.5194/gmd-14-7223-2021

Parajka, J., Viglione, A., Rogger, M., Salinas, J.L., Sivapalan, M., Blöschl, G., 2013. Comparative assessment of predictions in ungauged basins – Part 1: Runoff-hydrograph studies. Hydrology and Earth System Sciences 17, 1783–1795. https://doi.org/10.5194/hess-17-1783-2013

Pfister, L., Martínez-Carreras, N., Hissler, C., Klaus, J., Carrer, G.E., Stewart, M.K., McDonnell, J.J., 2017. Bedrock geology controls on catchment storage, mixing, and release: A comparative analysis of 16 nested catchments. Hydrological Processes 31, 1828–1845. https://doi.org/10.1002/hyp.11134

Pokhrel, Y., Felfelani, F., Satoh, Y., Boulange, J., Burek, P., Gädeke, A., Gerten, D., Gosling, S.N., Grillakis, M., Gudmundsson, L., Hanasaki, N., Kim, H., Koutroulis, A., Liu, J., Papadimitriou, L., Schewe, J., Müller Schmied, H., Stacke, T., Telteu, C.-E., Thiery, W., Veldkamp, T., Zhao, F., Wada, Y., 2021. Global terrestrial water storage and drought severity under climate change. Nat. Clim. Chang. 11, 226–233. https://doi.org/10.1038/s41558-020-00972-w

Poncelet, C., Merz, R., Merz, B., Parajka, J., Oudin, L., Andréassian, V., Perrin, C., 2017. Process-based interpretation of conceptual hydrological model performance using a multinational catchment set. Water Resources Research 53, 7247–7268. https://doi.org/10.1002/2016WR019991

Quilty, J., Adamowski, J., 2021. A maximal overlap discrete wavelet packet transform integrated approach for rainfall forecasting – A case study in the Awash River Basin (Ethiopia). Environmental Modelling & Software 144, 105119. https://doi.org/10.1016/j.envsoft.2021.105119

Saraiva, S.V., Carvalho, F. de O., Santos, C.A.G., Barreto, L.C., Freire, P.K. de M.M., 2021. Daily streamflow forecasting in Sobradinho Reservoir using machine learning models coupled with wavelet transform and bootstrapping. Applied Soft Computing 102, 107081. https://doi.org/10.1016/j.asoc.2021.107081

Senent-Aparicio, J., Blanco-Gómez, P., López-Ballesteros, A., Jimeno-Sáez, P., Pérez-Sánchez, J., 2021. Evaluating the Potential of GloFAS-ERA5 River Discharge Reanalysis Data for Calibrating the SWAT Model in the Grande San Miguel River Basin (El Salvador). Remote Sensing 13, 3299. https://doi.org/10.3390/rs13163299

https://doi.org/10.5194/hess-2024-83
Preprint. Discussion started: 20 March 2024

Sichangi, A.W., Wang, L., Yang, K., Chen, D., Wang, Z., Li, X., Zhou, J., Liu, W., Kuria, D., 2016. Estimating continental river basin discharges using multiple remote sensing data sets. Remote Sensing of Environment 179, 36–53. https://doi.org/10.1016/j.rse.2016.03.019

Smiti, A., 2020. A critical overview of outlier detection methods. Computer Science Review 38, 100306. https://doi.org/10.1016/j.cosrev.2020.100306

Stein, L., Clark, M.P., Knoben, W.J.M., Pianosi, F., Woods, R.A., 2021. How Do Climate and Catchment Attributes Influence Flood Generating Processes? A Large-Sample Study for 671 Catchments Across the Contiguous USA. Water Resources Research 57, e2020WR028300. https://doi.org/10.1029/2020WR028300

Talukder, S., Singh, R., Bora, S., Paily, R., 2020. An Efficient Architecture for QRS Detection in FPGA Using Integer Haar Wavelet Transform. Circuits Syst Signal Process 39, 3610–3625. https://doi.org/10.1007/s00034-019-01328-2

Teng, L.Y., Mattar, C.N.Z., Biswas, A., Hoo, W.L., Saw, S.N., 2022. Interpreting the role of nuchal fold for fetal growth restriction prediction using machine learning. Sci Rep 12, 3907. https://doi.org/10.1038/s41598-022-07883-0

Tu, T., Wang, J., Zhao, G., Zhao, T., Dong, X., 2024. Scaling from global to regional river flow with global hydrological models: Choice matters. Journal of Hydrology 633, 130960. https://doi.org/10.1016/j.jhydrol.2024.130960

Veldkamp, T.I.E., Zhao, F., Ward, P.J., Moel, H. de, Aerts, J.C.J.H., Schmied, H.M., Portmann, F.T., Masaki, Y., Pokhrel, Y., Liu, X., Satoh, Y., Gerten, D., Gosling, S.N., Zaherpour, J., Wada, Y., 2018. Human impact parameterizations in global hydrological models improve estimates of monthly discharges and hydrological extremes: a multi-model validation study. Environ. Res. Lett. 13, 055008. https://doi.org/10.1088/1748-9326/aab96f

Wei, D., Gephart, J.A., Iizumi, T., Ramankutty, N., Davis, K.F., 2023. Key role of planted and harvested area fluctuations in US crop production shocks. Nat Sustain 6, 1177–1185. https://doi.org/10.1038/s41893-023-01152-2

Wei, S., Song, J., Khan, N.I., 2012. Simulating and predicting river discharge time series using a wavelet-neural network hybrid modelling approach. Hydrological Processes 26, 281–296. https://doi.org/10.1002/hyp.8227

Xie, J., Xu, Y.-P., Gao, C., Xuan, W., Bai, Z., 2019. Total Basin Discharge From GRACE and Water Balance Method for the Yarlung Tsangpo River Basin, Southwestern China. Journal of Geophysical Research: Atmospheres 124, 7617–7632. https://doi.org/10.1029/2018JD030025

https://doi.org/10.5194/hess-2024-83
Preprint. Discussion started: 20 March 2024

Xu, Z., Mo, L., Zhou, J., Fang, W., Qin, H., 2022. Stepwise decomposition-integration-prediction framework for runoff forecasting considering boundary correction. Science of The Total Environment 851, 158342. https://doi.org/10.1016/j.scitotenv.2022.158342

Yang, Y., Pan, M., Lin, P., Beck, H.E., Zeng, Z., Yamazaki, D., David, C.H., Lu, H., Yang, K., Hong, Y., Wood, E.F., 2021. Global Reach-Level 3-Hourly River Flood Reanalysis (1980–2019). Bulletin of the American Meteorological Society 102, E2086–E2105. https://doi.org/10.1175/BAMS-D-20-0057.1

Zhao, T., Chen, Z., Tu, T., Yan, D., Chen, X., 2022. Unravelling the potential of global streamflow reanalysis in characterizing local flow regime. Science of The Total Environment 838, 156125. https://doi.org/10.1016/j.scitotenv.2022.156125

Zuo, G., Luo, J., Wang, N., Lian, Y., He, X., 2020. Decomposition ensemble model based on variational mode decomposition and long short-term memory for streamflow forecasting. Journal of Hydrology 585, 124776. https://doi.org/10.1016/j.jhydrol.2020.124776