We thank all reviewers for their constructive feedback. We have carefully collated all comments, addressed each point thoroughly, and documented the corresponding revisions. Our detailed responses to each comment are provided below.

Comments and responses regarding Dr. Giacomo Medici

Lines 68-71. Insert new and recent literature for fissures and pipes in karstic dolostones and limestones:

Medici, G., Munn, J.D., Parker, B.L. 2024. Delineating aquitard characteristics within a Silurian dolostone aquifer using high-density hydraulic head and fracture datasets. Hydrogeology Journal 32(6), 1663-1691.

Giese, M., Reimann, T., Bailly-Comte, V., Maréchal, J. C., Sauter, M., Geyer, T. 2018. Turbulent and laminar flow in karst conduits under unsteady flow conditions: Interpretation of pumping tests by discrete conduit-continuum modeling. Water Resources Research, 54(3), 1918-1933.

Reply: These two references illustrate the uniqueness of karst aquifers, which aligns with the discussion in lines 68–71 of the manuscript regarding the distinct hydrological characteristics of karst areas compared to other regions. Therefore, we have included them as references in the main text.

Lines 112-113. More detail (e.g., possible ranges and maximum value etc etc) here or in the methodology on KGE and NSE coefficients.

Line 193. Here you can insert the necessary methodological details for KGE and NSE coefficients.

Reply: We have combined the responses to these two comments as they both address the need for a more detailed description of the performance metrics (KGE and NSE). We agree that providing explicit definitions, ranges, and interpretations for these coefficients is crucial for clarity. Accordingly, we have substantially expanded the description in Section 2.2.2 (Methodology). We have specified the theoretical ranges ($-\infty$-1), interpreted the values (closer to 1 indicates higher accuracy), and clarified the method of using "strict baseflow points" for validation. The revised text in Section 2.2.2 is as follows:

"In order to validate the accuracy of different baseflow separation methods in karstic regions, we chose two performance metrics: the KGE and NSE coefficients. Theoretically, these coefficients vary from ∞ to 1; values closer to the maximum of 1 indicate higher accuracy, while values below 0 typically suggest unacceptable performance. The assessment methodology followed Xie et al. (2020), centering on

the screening of strict baseflow points to benchmark the separation methods.

The core of screening strict baseflow points lies in excluding periods with precipitation recharge and anomalies in the runoff curve, retaining only the stages where baseflow acts as the primary runoff source. In such stages, there should be a distinct consistency between baseflow and runoff. Therefore, using NSE and KGE coefficients as evaluation criteria can reasonably assess the effectiveness of baseflow separation methods."

Line 114. The aim is clear, but you need to specify the 3 to 4 specific objectives of your research by using numbers (e.g., i, ii, iii).

Lines 495. Four points in the discussion. Maybe, four objectives to disclose in the introduction?

Reply: We appreciate this constructive suggestion. We agree that explicitly aligning the objectives in the Introduction with the four points in the Discussion significantly improves the logical flow of the manuscript.

In conjunction with the comment on Line 114 (which suggested using numbered objectives), we have rewritten the final paragraph of the Introduction. The revised objectives now correspond directly to the structure of our Discussion section (Specificity, Heterogeneity across types, Climate response, and Methodological applicability). The revised text is as follows:

"The overarching aim of this study is to explore the baseflow characteristics and their internal differences across global karst regions, and to evaluate the influence of different environmental factors on these characteristics. To achieve this, the following four specific objectives are defined:(i) To evaluate the applicability of twelve baseflow separation methods (including graphical and digital filter methods) in karst watersheds and identify the most suitable approaches.(ii) To reveal the overall specificity of karst baseflow and analyze its spatiotemporal evolution characteristics under changing climate conditions.(iii) To quantify the heterogeneity of baseflow across different karst landform types and clarify the internal differences.(iv) To identify the dominant climatic, topographic, and geological drivers of baseflow variability using the XGBoost machine learning model."

Figure and tables

Figure 1. Increase graphic resolution or make the figures larger.

Reply: We have adjusted the color differentiation between categories and removed
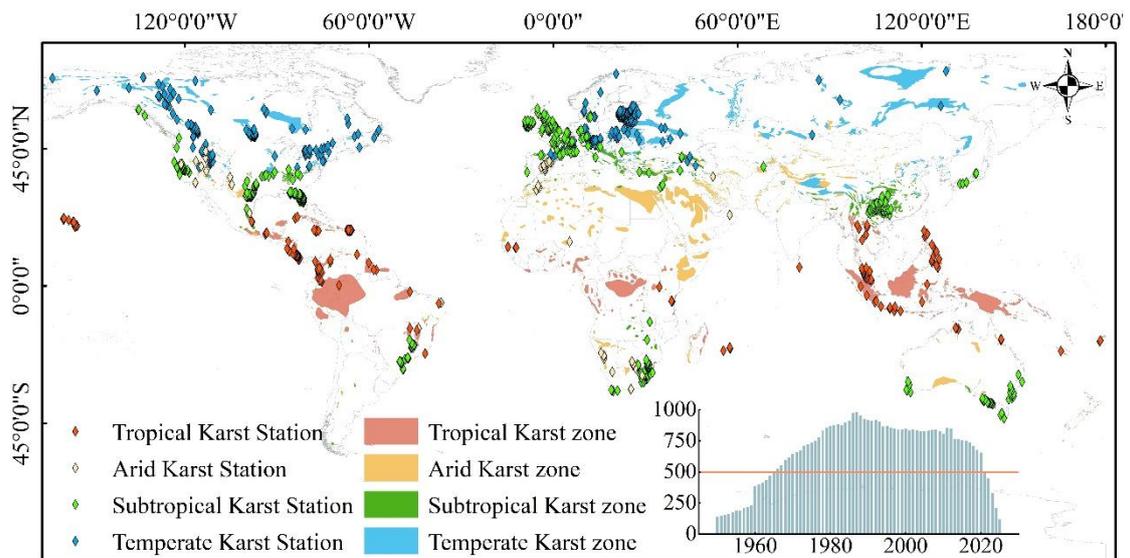
extraneous ranges.

Reply: We have adjusted the font size in the figure to improve its clarity and readability. In addition, in response to other feedback, we have incorporated a significance difference analysis to illustrate the distinctions and correlations between different methods. In the figure, the letters above each box plot indicate differences among the various methods, while the horizontal lines below denote significant differences between the graphical method and the digital filtering method.
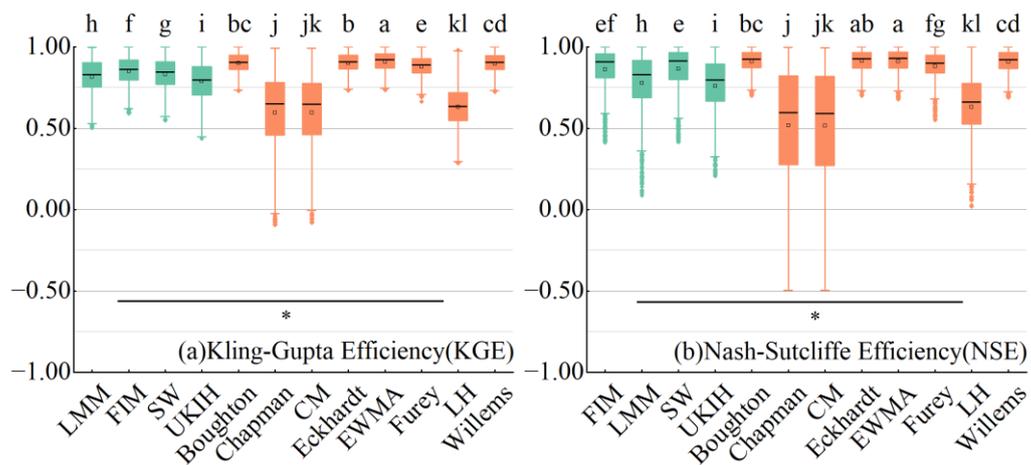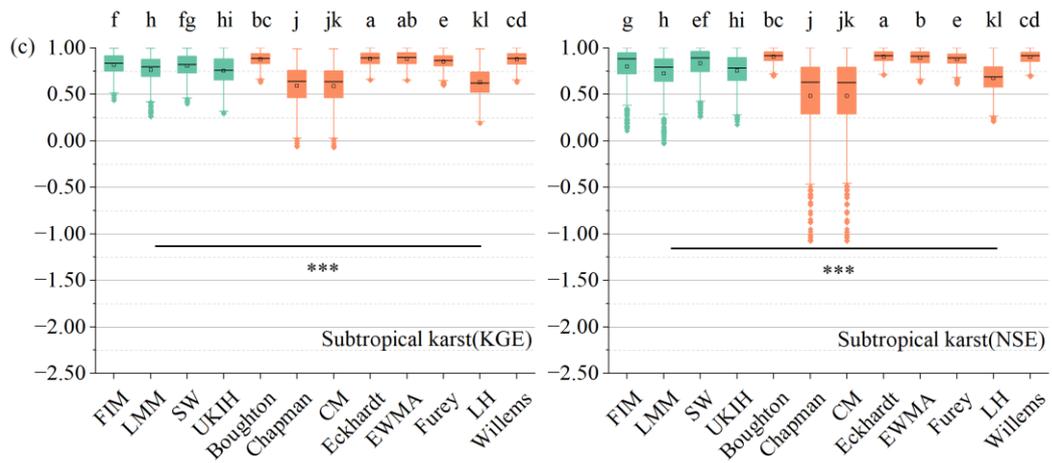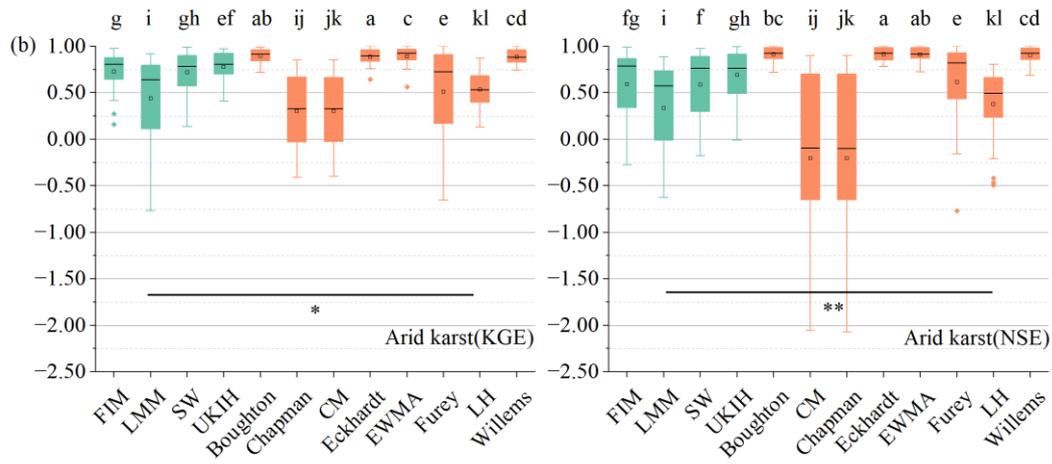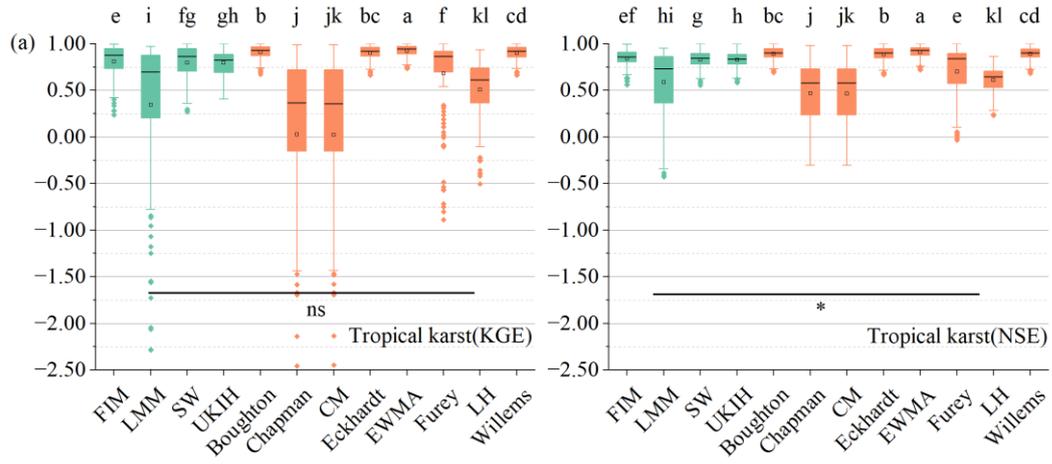
The specific modified graphic is as follows:

Figure 3

(a) Tropical karst(KGE) / Tropical karst(NSE)

(b) Arid karst(KGE) / Arid karst(NSE)
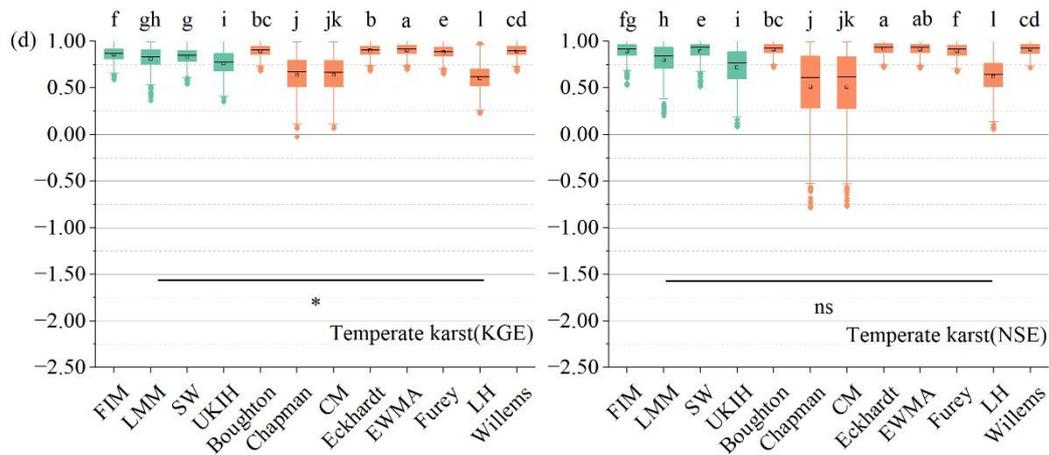
(c) Subtropical karst(KGE) / Subtropical karst(NSE)

Figure 4.

I would provide more detail in the introduction on the definitions of the climate. What does it mean "temperate karst". Semi-arid, humid?

Reply: We appreciate this constructive suggestion. We realize that using terms like "temperate" or "tropical" karst without context in the Introduction might cause confusion. Action taken:

In the Introduction (Paragraph 3): We have integrated a brief clarification of these climatic terms directly into the flow of the text. We specified that these classifications are generally based on hydrothermal conditions (e.g., Köppen classification). For instance, we now explicitly associate "tropical" regions with high humidity and temperature and "arid" regions with water limitations to contextualize the observed Baseflow Index (BFI) differences. This ensures the reader is oriented before we discuss specific regional characteristics.

In the Methodology (Section 2.1.1): To provide a rigorous definition as per scientific standards, we have also added a detailed explanation of the specific Köppen classification criteria used to categorise the 1,375 watersheds in this study.

Figure 8. Please, provide more detail on the description of the results for the (i) R2, (ii) NSE and (iii) KGE values which look good. R2 maybe a bit low, but this is ok.

Reply: We have added a detailed description of the goodness-of-fit metrics(L369-378). The results show a high NSE (0.825) and KGE (0.817), indicating that the linear model captures the long-term trajectory of the baseflow very well. Regarding the R^2 (0.62), as the reviewer correctly noted, it is slightly lower due to the inherent inter-annual fluctuations of hydrological data. However, in conjunction with the high NSE and KGE, it confirms the reliability of the increasing trend.

Reply: We have adjusted the layout of Figure (c) by enlarging it and modifying the overall color scheme to better distinguish karst types across different climatic zones. Below is the updated figure.
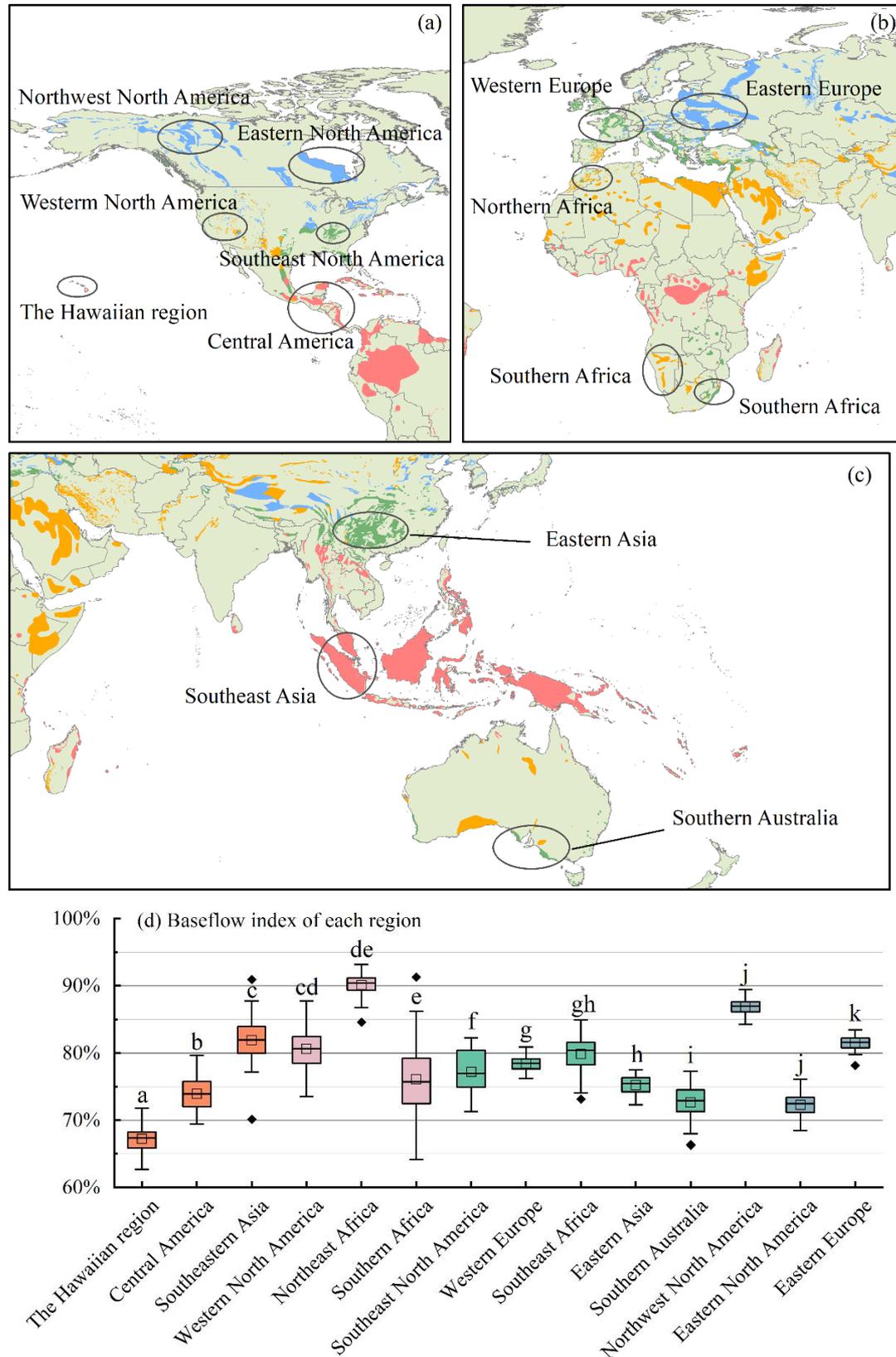
Reply to Anonymous Referee 1

L27: What is the baseflow index? Please clarify!

L36: Please provide the full name of the BFIs.

Reply: We have combined the responses to these two comments as they both address the clarity and definition of the Baseflow Index in the Abstract.

We agree that the full term and a clear definition were necessary for the first mention. We have revised the Abstract to explicitly provide the full name "Baseflow Index" (BFI) along with its definition. The revised sentence in the Abstract is as follows:

Graphical and digital filtering methods were employed to estimate the Baseflow Index (BFI, defined as the ratio of baseflow to total streamflow) and to analyze its spatial patterns and trends.

L45-46: "as a slow recharge component of ..."and "as a hydrological stabiliser"wore both presented. Please rephrase this sentence.

Reply: We appreciate the reviewer pointing out the awkward phrasing in the original sentence. We have rephrased it to improve clarity and flow by using a parallel structure. The revised sentence is as follows:

"Baseflow serves as both a slow groundwater recharge component to runoff and a vital hydrological stabiliser (Mukherjee et al., 2018; Chen et al., 2019)."

This revision effectively integrates the origin and function of baseflow without the repetitive use of "as".

L60-66: This sentence was too tedious and long. Please improve this sentence.

Reply: We agree with the reviewer that the original sentence was overly long and tedious.   We have rewritten this section by breaking the long sentence into shorter, clearer statements to improve readability while retaining the specific regional examples.   The revised text is as follows:

Regional-scale studies further refine these spatial patterns. For instance, the BFI is observed to be higher in the eastern regions of both the United States and India compared to their western counterparts (Mei et al., 2024; Sharma and Mujumdar, 2024). In China, the Yellow River Basin exhibits a 'high-low-high' trend from upstream to downstream, whereas the Wei River Basin shows a gradual decline (Lyu et al., 2023; Zhang et al., 2019).

L108: For the daily-scale runoff data from 1375 watersheds within the karst region,

Reply: The raw daily runoff data compiled in this study cover a broad historical period from 1880 to 2024. However, data availability varies significantly over time. To ensure the reliability of the global assessment, we applied a screening criterion: only years with valid data from at least 500 basins globally were included in the trend analysis (which effectively covers the period from 1960 to 2015).

In the revised manuscript, we have:

Clarified the time span in the Introduction to provide immediate context. We have clarified the data period in Section 2.1.1. The revised text states: "This process resulted in a final dataset of 1,375 valid karst watersheds covering the period from 1960 to 2015."

Highlighted the temporal distribution of valid data in the lower-right subplot of Figure 1, where the red horizontal line indicates the 500-basin threshold.
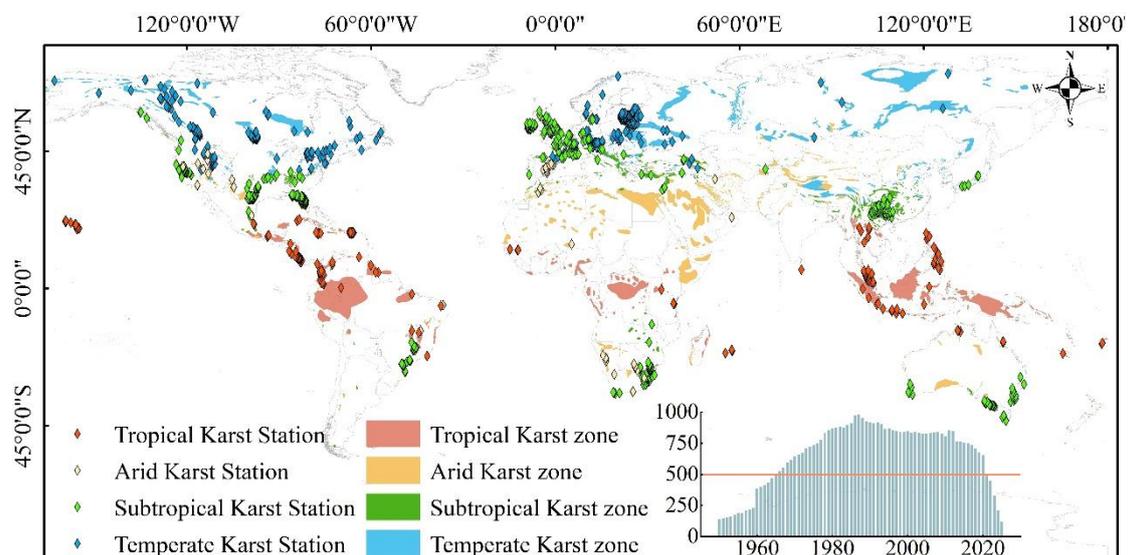
L122: 1412 watersheds? You can directly present the 1375 stations since 37 stations were not used in present study.

Reply: We accept the reviewer's suggestion to clarify the number of stations.   In the revised manuscript, we have explicitly stated that the final analysis is based on 1,375 watersheds.

We have also refined the description of the data cleaning process in Section 2.1.1 to explain that while 1,412 stations were initially considered, strict quality control (excluding stations with severe data gaps or prolonged zero-flow periods) reduced the final valid dataset to 1,375. This ensures transparency in our data selection process.

186-190: Eight methods to calculate the baseflow should be described in detail. Alternatively, you can add a Table to exhibit these eight methods.

Reply: We thank the referees for pointing this out. We acknowledge that the original description of the separation methods was too brief.

In the revised Section 2.2.1, we have added a comprehensive summary table (Table 2) that details the background and principles of all 12 methods used in this study (including the 8 digital filters mentioned by the reviewer and 4 graphical methods). We also clarified the parameter estimation process (e.g., recession constant estimation) in the text to ensure reproducibility.

2: The colors for these four karst regions were too similar. Please improve the color.

Reply: The color scheme in the figure has been modified to achieve better discrimination.
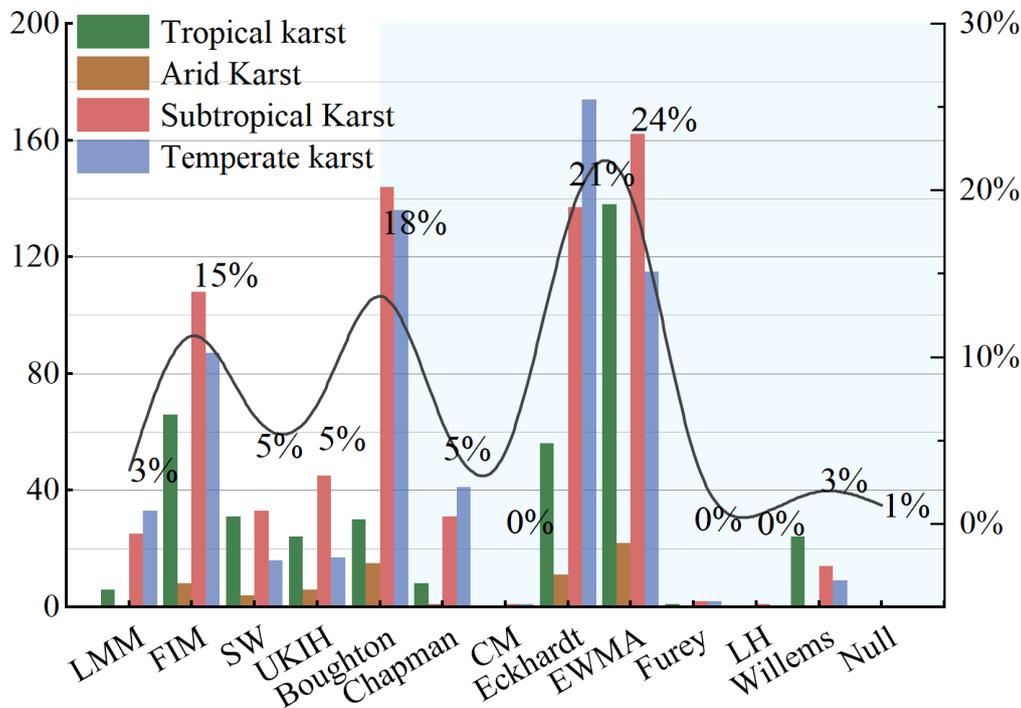
3-4: Significant difference test should be added.

Reply: We have added a significant difference test. Using the Kruskal-Wallis test, we confirmed the statistical significance of the differences in both the KGE and NSE coefficients among the 12 baseflow separation methods. Accordingly, we will enhance the main text by adding a discussion on the performance differences between different types of methods, along with further interpretation of the effectiveness of each separation method.

In the figure below, the letters denote the results of multiple comparisons based on the Mann-Whitney U test, while the horizontal lines at the bottom of the figure indicate

significant differences between the graphical methods and digital filtering methods. Methods sharing the same letter show no significant difference at the p < 0.05 level, whereas methods with different letters are statistically significantly different. Methods assigned multiple letters do not differ significantly from multiple groups.
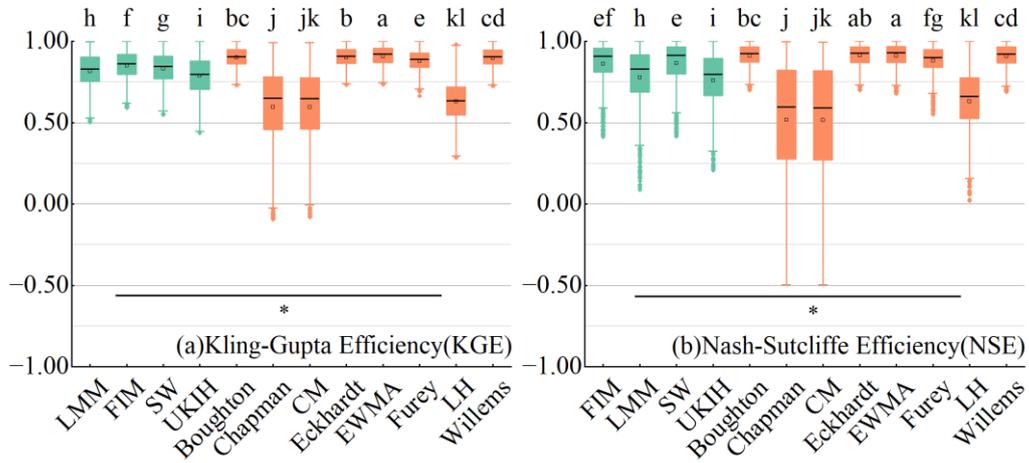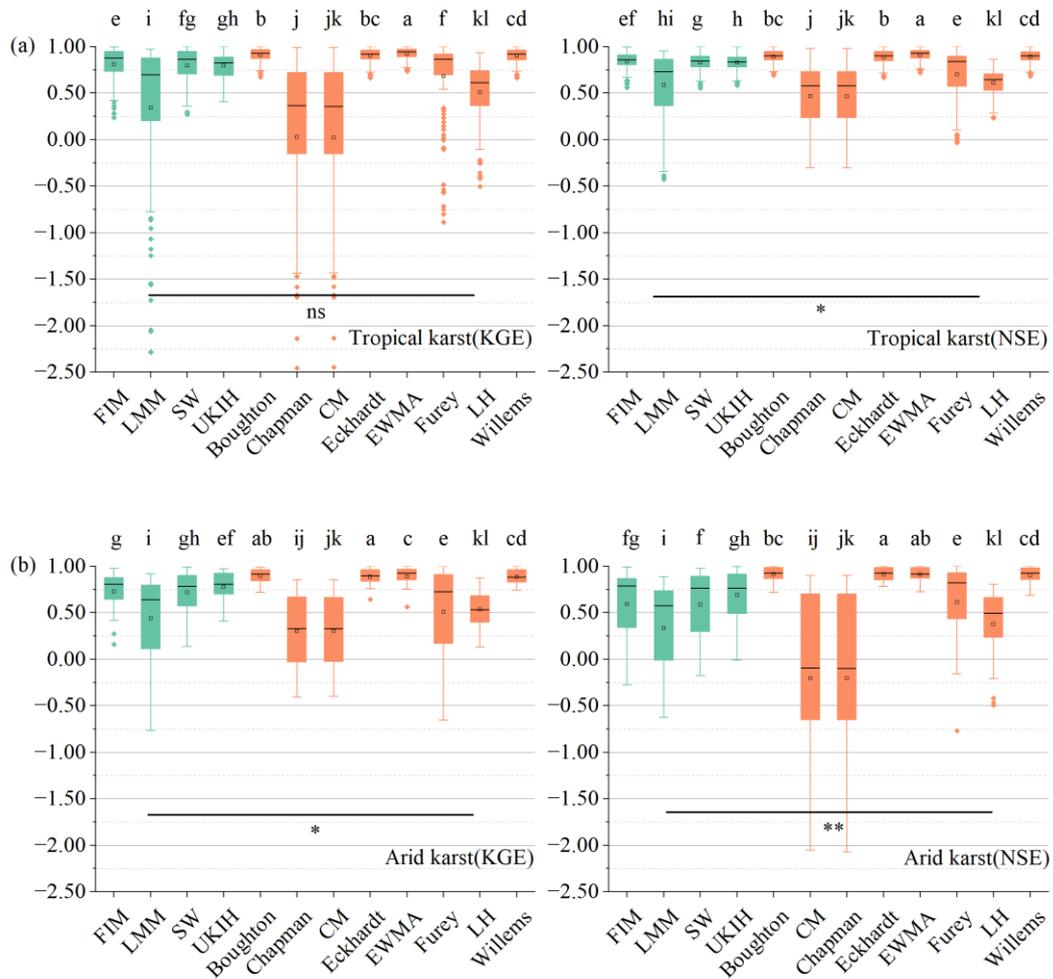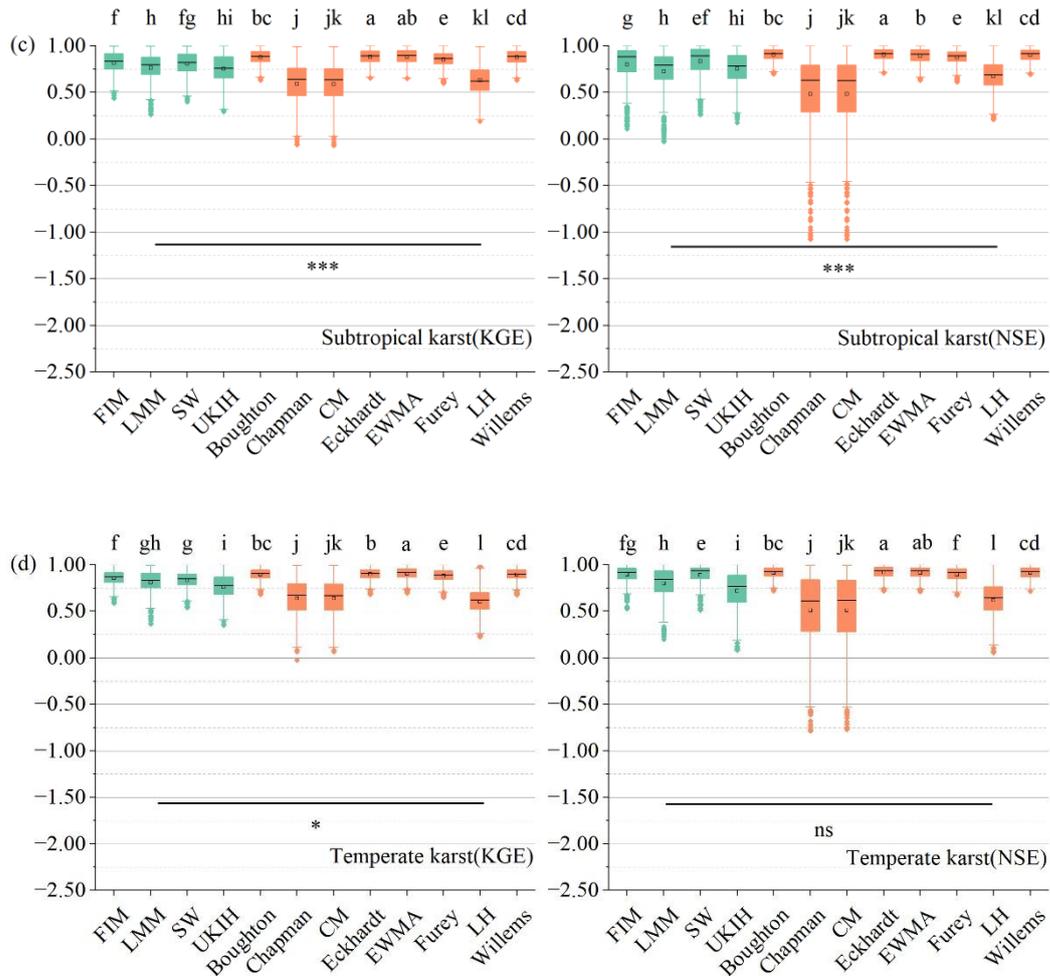

Figure 3

Figure 4.

Reply: We appreciate the reviewer's suggestion to strengthen the statistical interpretation of Figure 8. To verify the trend significance, we performed a Mann-Kendall trend test, which yielded a P-value of 0.00002 (P<0.001), confirming a highly significant increasing trend. We further cross-verified this with linear regression (Durbin-Watson statistic ≈ 1.8, P ≈ 0.00003), ensuring the trend is not due to random chance.

The following is the modified graph after increasing the p-value:

y = 0.00033x+0.116
$R^2 = 0.62$
NSE = 0.825
KGE = 0.817
P = 0.00002 (<0.05)
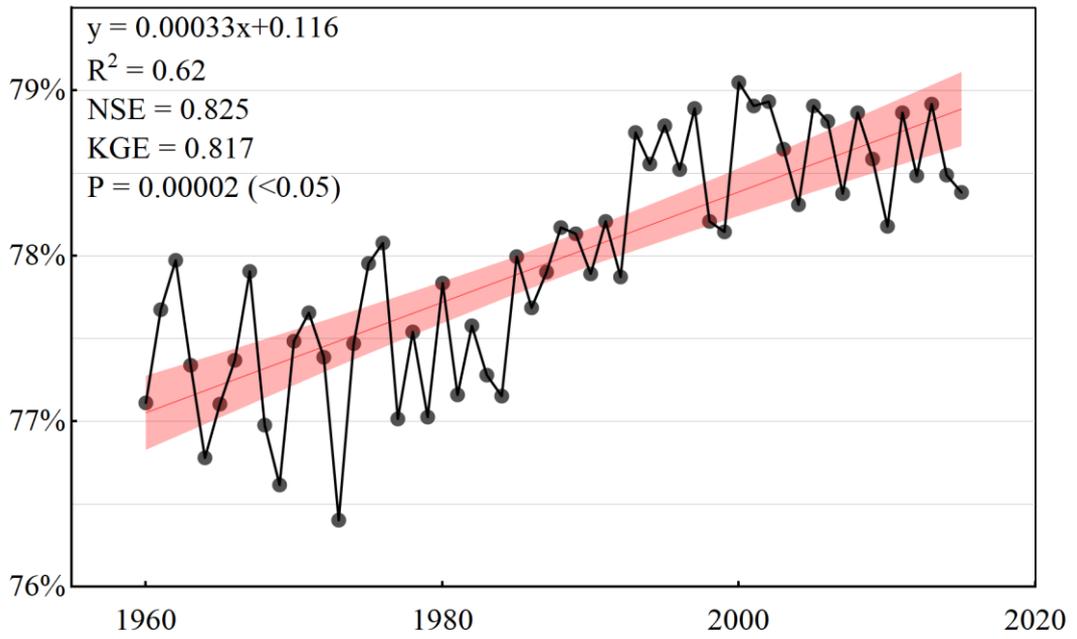
Comments on Dr. Yang Lu:

1. The baseflow separation methods seem to generate distinct BFIs and event contrasting trends in certain time periods. How do you tackle the different separation results?

Reply: We agree with the reviewer that different separation methods can yield varying results. To tackle this uncertainty, rather than relying on a single method or a simple average, we adopted an automatic optimization strategy to identify the most suitable method for each individual watershed.

Our approach is based on the hydrological principle that during recession periods (after surface runoff has ceased), streamflow is dominated by baseflow. Therefore, the separated baseflow should match the observed total streamflow during these periods. We utilized the Nash-Sutcliffe Efficiency (NSE) and Kling-Gupta Efficiency (KGE) to evaluate the fit between the separated baseflow and the observed streamflow specifically during these recession periods.

To ensure surface runoff had fully receded, we focused on watersheds with an area < 2,500 km², where the concentration time is short. For each watershed, the baseflow Python library calculated the KGE and NSE for all 12 methods, The method with the highest accuracy scores was automatically selected as the "optimal method" for that specific watershed. This ensures that the final BFI and trend analyses are based on the most physically representative separation result for each basin. We have clarified this optimization logic in Section 2.2.1.

2. In Figure 7, the authors compared BFI in karst regions and the global average. However, the global average is impacted by the coverage ratio of karst regions. Consider replace the global mean with non-karst area mean.

Reply: We appreciate this insightful comment. We agree that comparing Karst BFI directly with non-Karst BFI (rather than the Global average) provides a more scientifically rigorous contrast.

Accordingly, we have revised Figure 7 to replace the "Global Mean" with the "Non-Karst Mean." To establish this baseline, we utilized data from over 6,000 non-karst watersheds. It is worth noting that due to the large sample size of non-karst watersheds (over 6,000) compared to karst watersheds (1,375), the numerical difference between the previously used "Global Mean" and the "Non-Karst Mean" is actually very small (<1%). However, to eliminate any potential bias from the coverage ratio as pointed out by the reviewer, we have adopted the non-Karst BFI in the revised figure to ensure a strictly independent comparison.

Reply: We appreciate this insightful comment. We agree that focusing solely on relative importance in the initial draft failed to clarify whether the effects of various factors on the Base Flow Index (BFI) were positive or negative. Following your suggestion, we have incorporated SHapley Additive exPlanations (SHAP) into the revised manuscript to comprehensively analyze both the direction and magnitude of each factor's influence on BFI. Specifically, we have made the following modifications:

We added a SHAP summary plot (beeswarm plot) to visualize the global importance of each feature. This plot ranks features by importance and uses color coding to demonstrate how high or low feature values impact the BFI (positively or negatively). We included SHAP dependence plots (or detailed analysis) to further illustrate the specific relationship between key feature values and their corresponding SHAP values, providing a more granular view of their influence on BFI.

The relevant figures (Figure 11) and detailed interpretations have been added to Section 3.4 of the revised manuscript. Please see the updated figures and text below.

(b)