Response to reviewers

We greatly appreciate the reviewers providing valuable and constructive comments on our manuscript. We seriously considered each comment and revised the original manuscript accordingly. The individual comments are replied below. In the following, the reviewer comments are black font and our responses are blue, and the green texts are the quotes of the revised manuscript.

Reviewer #2

The manuscript 'The general formulation for runoff components estimation and attribution at mean-annual time scale' proposes a concise MPS framework for partitioning total runoff into surface and baseflow components. The topic is timely and the presentation generally clear. With several focused revisions—mainly on scope, definitions, robustness checks, and uncertainty, the paper will be suitable for publication.

Reply: We sincerely appreciate your constructive suggestions. We have carefully addressed each comment and incorporated corresponding revisions into the revised manuscript.

Concerns:

Please explicitly delimit applicability to small/medium catchments and justify the exclusion (or stratified analysis) of large basins, where digital filters can misclassify delayed stormflow as baseflow. Provide a short area-threshold sensitivity (e.g., ≤500/1,000/2,500/5,000 km²) showing effects on BFI and on MPS fits; discuss implications for scaling to large rivers (cf. recent global assessments).

Reply: Thank you for this important comment. According to Xie et al. (2024), the underlying assumptions of digital filter baseflow separation methods may not be appropriate for large basins. For example, headwater stormflow of large basins may take weeks to reach the basin outlet and become the low-frequency component of downstream flow. Consequently, these separation methods typically overestimate baseflow in large basins because they misidentify upstream stormflow as baseflow (Rutledge, 1998). Therefore, we focus our analysis on small and medium catchments with an area⩽500,000 km² to minimize the influence of channel routing.

Furthermore, we conducted the area-threshold sensitivity analysis in China as recommended. We systematically tested the effects of varying area thresholds on the performance of the fitted MPS curves. The results showed that the goodness of fit for the MPS relationships remained robust and did not exhibit significant degradation across these different area thresholds (Table A1). We interpret the stability of the MPS fits to mean that the functional relationship between available water and runoff components (as captured by the MPS model) may be scale-invariant within the range of basin sizes studied.

Table A1 The coefficient of determination (R^2) and model parameters for the MPS curve fittings under different area thresholds for selecting catchments in China

Area thresholds	Number of	R^2			Parameters (mm)		
(km^2)	catchments	Q_s	Q_b	Q	W_p	V_p	U_p
2,000	67	0.85	0.62	0.89	3220	2794	1439
5,000	135	0.84	0.63	0.89	3004	2651	1356
10,000	180	0.84	0.69	0.90	3098	2614	1375
20,000	219	0.85	0.68	0.90	3138	2585	1376
80,000	257	0.85	0.69	0.90	3207	2487	1364
500,000	295	0.85	0.69	0.91	3278	2428	1362

Clarify the boundary between baseflow/slow flow and surface/fast flow. At minimum, acknowledge that baseflow aggregates multiple processes (groundwater discharge, hyporheic/subsurface flow, delayed snowmelt, and—if relevant—deep leakage).

Reply: Thank you for this suggestion. We acknowledge that the term "baseflow" aggregates multiple delayed flow processes, including groundwater discharge, hyporheic exchange, subsurface stormflow, delayed snowmelt, and deep leakage with distinct origins, timescales and physical mechanisms. In response, we have expanded the Discussion section (Line 536-553) to explicitly recognize that baseflow represents an integrated concept encompassing these heterogeneous components: "It is important to acknowledge several uncertainties in this study. First, the definition of "baseflow" itself introduces uncertainty. Although widely used as a collective term for delayed streamflow components, baseflow encompasses contributions from hydrologically distinct sources such as groundwater drainage, hyporehic exchange, snowmelt, and deeper subsurface leakage-each with distinct origins, timescales, and sensitivities to environmental factors. For instance, groundwater flow and deep leakage are strongly controlled by geological heterogeneity, including the distribution of rock types, porosity, permeability, faults, and fractures (Schiavo et al., 2023). In contrast, snowmelt baseflow, on the other hand, is mainly driven by temperature variations within interannual to decadal climate cycles. Future studies could combine isotope tracing with hydrological modeling to better quantify the contributions of these different sources".

Strengthen the interpretation of Wp, Vp, and Up: (1) outline hypothesized controls (soil/rock properties, storage capacity, seasonality); (2) report basic identifiability/collinearity checks (against Budyko-type indices); (3) add a cross-region transfer test (China→CONUS and vice versa) to show portability; (4) explain how to calculate the changes in parameters when attributing the variations in runoff components, such as ΔUp.

Reply: We thank for these insightful suggestions.

(1) We agree that a clearer physical interpretation of the parameters is beneficial. In the revised Discussion section, we have added the following paragraph: "W_p is influenced by soil properties and available storage capacity, determining the fraction of precipitation that rapidly becomes surface runoff versus what is stored (Line 503-504)"; "The parameter V_p is the upper limit of the fraction of wetting returned to the atmosphere as water vapor (Ponce and Shetty, 1995), and is likely responds to subsurface characteristics such as aquifer permeability and geological layering (506-508)".

(2) We have compared the results with Budyko equations in Section 5.1.

(3) In the doctoral thesis of the first author (He, 2025), the explicit equations relating the parameters (W_p, V_p and U_p) to catchment attributes (e.g., rainfall intensity, snow fraction, topographic indices, elevation, permeability) have been established using a large dataset of Chinese catchments. These relationships have been validated within China and shown to provide reliable runoff components estimates for ungauged catchments. While a direct cross-region transfer test (e.g., China → CONUS) is beyond the scope of this paper, the attribute-based parameterization approach provides a strong foundation for geographical generalizability. We will explicitly recommend and undertake this important validation in future work.

(4) The changes in parameters between two periods (e.g., ΔU_p) are calculated as follows: First, the U_{p1} and U_{p2} are inversely estimated from the observed total runoff

using Equation (14) for period 1 and period 2, respectively. Then, the change of U_p is computed simply as the difference between two periods ($\Delta U_p = U_{p2} - U_{p1}$). Similarly, ΔP represents the change in mean annual precipitation between the two periods. These derived changes (ΔP, ΔU_p) are then used in the attribution framework (Equation 17(b)) to quantity variations in total runoff to climatic and environment changes.

Discuss how known aquifer heterogeneity and preferential flow may map onto parameter dispersion (notably Vp).

Reply: Thank you for this important comment. In response, we have added the following discussion to Section 5 (Line 507-514) of the revised manuscript: "..., and is likely responds to subsurface characteristics such as aquifer permeability and geological layering. For instance, in highly heterogeneous aquifers with well-developed preferential pathways (e.g., fractured rock or karst systems), water is rapidly drained toward the stream, leading to a higher efficiency of baseflow production and thus a lower V_p value (as less water is retained for evaporation). Conversely, in catchments with more homogeneous, porous media (e.g., sandy aquifers), water movement is slower and more diffuse, potentially allowing for a greater fraction of stored water to be evaporated, resulting in a higher V_p".

Minor comments

Unify color scales/units; add 95% confidence bands to CDF/scatter plots.

Reply: Done.

Provide a concise symbol table (first occurrence) and standardize terminology ('runoff components' vs 'flow components'; 'baseflow/slow flow').

Reply: We thank the reviewer for the suggestion regarding terminology and symbols. We have thoroughly reviewed the manuscript to ensure standardized terminology. The terms "runoff components" and "baseflow" are now used consistently throughout the text.

Regarding the symbol table, we have defined each symbol upon its first occurrence in the text. We believe this approach provides clarity to readers without a symbol table.

Briefly document missing-data criteria, period lengths by region, and QC steps.

Reply: Thank you for this important suggestion. We have supplemented the catchment screening criteria in Section 3.1, with detailed procedures available in He et al. (2025).

In Table 1, state whether exponents/capacities are calibrated or empirical and, where possible, cite numerical/observational backing.

Reply: We have added the sources of parameters in Table 1.

Add a short analysis or paragraph on precipitation seasonality effects on BFI and on partitioning assumptions at the annual scale.

Reply: Thank you for this comment. We have added some discussion in Line 604-609: "In addition, the seasonality of precipitation measures the concentration of precipitation within a year. The more concentrated the precipitation, the more likely it is to generate surface runoff, resulting in greater intra-annual fluctuations in the BFI and a lower annual BFI. In contrast, in catchments with evenly distributed precipitation, soil water and groundwater are replenished consistently and gradually, leading to relatively stable intra-annual BFI and a higher annual BFI".

For the phrase 'As for ΔQ attribution' on line 394, perhaps 'attribution' should be removed.

Reply: Done.

References

Beck, H.E., van Dijk, A., Miralles, D.G., de Jeu, R.A.M., Bruijnzeel, L.A., McVicar, T.R. & Schellekens, J. (2013). Global patterns in base flow index and recession based on streamflow observations from 3394 catchments. Water Resources Research, 49(12): 7843-7863. DOI:10.1002/2013wr013918

Bloomfield, J.P., Gong, M., Marchant, B.P., Coxon, G. & Addor, N. (2021). How is Baseflow Index (BFI) impacted by water resource management practices? Hydrology and Earth System Sciences, 25, 5355-5379. DOI: 10.5194/hess-25-5355-2021

Gonzales, A. L., Nonner, J., Heijkers, J. & Uhlenbrook, S. (2009). Comparison of different base flow separation methods in a lowland catchment. Hydrology And Earth System Sciences, 13, 2055-2068. DOI:10.5194/hess-13-2055-2009

Lott, D. A. & Stewart, M. T. (2016). Base flow separation: a comparison of analytical and mass balance methods. Journal of Hydrology, 535, 525-533. DOI:10.1016/j.jhydrol.2016.01.063

Rutledge, A. T (1998). Computer Programs for Describing the Recession of Ground-Water Discharge and for Estimating Mean Ground-Water Recharge and Discharge from Streamflow Records-Update (USGS); https://doi.org/10.3133/wri984148

Wang, Y., Chen, Y. & Chang, H. (2021). Seasonal dynamic identification of Eckhardt digital filter parameters based on isotopes. Water Resources and Hydropower Engineering, 52(12): 99-110. DOI:10.13928/j.cnki.wrahe.2021.12.010

Xie, J., Liu, X., Jasechko, S., Berghuijs, W.R., Wang, K., Liu, C., Reichstein, M., Jung, M. & Koirala, S. (2024). Majority of global river flow sustained by groundwater. Nature Geoscience, 17(8): 770-777. DOI:10.1038/s41561-024-01483-5