Point-by-point reply to referees' comments

Article: EGUSPHERE-2025-244

Anonymous Referee #1:

General Comments:

This manuscript presents a well-executed and innovative study on baseflow separation in small rural catchments, focusing on the calibration of three Recursive Digital Filters (Eckhardt, Lyne and Hollick, and Chapman and Maxwell). The integration of dissolved silica as a tracer and the event-specific calibration approach are both novel and valuable contributions to the field. The writing is clear, and engaging, making the technical content accessible and enjoyable to read. The study's methodology is robust, and the results convincingly demonstrate the advantages of dynamic, event-based calibration-particularly for the Eckhardt filter, which outperforms the others in accuracy when parameters are tailored to rainfall event intensity.

Response: Thank you for your thorough and positive feedback on our study. We're pleased that our novel integration of dissolved silica and dynamic calibration approach has been recognised as a valuable contribution to the field.

Specific Comments:

While the manuscript is strong overall, there are areas where it could be further improved. The authors provide a thorough explanation of performance metrics such as NSE and RMSE, but this level of detail may be unnecessary for the target audience, who are likely already familiar with these standard evaluation tools. Streamlining these sections would help maintain the manuscript's focus and momentum.

Response: We thank the reviewer for their constructive feedback. We acknowledge the point that the target audience is likely familiar with the model performance metrics. So, to streamline the manuscript and maintain its momentum, we have condensed the descriptions of NSE, RMSE, and related metrics. We have reduced section 2.4 by 200 words. See revised section 2.4.

The presentation of results relies heavily on numerical values in the text. Incorporating more plots and figures would greatly enhance the reader's ability to interpret and appreciate the findings.

Response: We agree that more visual aids would enhance clarity and reader engagement. We have revised the results section (section 3), incorporating two additional multi-plot figures (Figures 7 and 8).

This is a timely contribution with clear novelty in its event-specific calibration strategy and use of chemical tracers. The manuscript is exceptionally well-written and methodologically sound. To maximize its impact, I recommend reducing the over-explanation of standard metrics and enhancing the presentation of results with more visual aids. These improvements would make the findings even more accessible and persuasive to a broad hydrology audience.

Response: Thank you again for the constructive feedback, which has improved the quality of our manuscript. As described in the above comments, we have incorporated both suggestions, reducing the over-explanation of performance metrics and including two new figures.

Anonymous Referee #2

General Comments

The manuscript describes an analysis in which three baseflow separation methods (specifically types of recursive digital filters) are tested against a silica-based estimate of baseflow for a small watershed in Brazil. The authors compare RDFs fit against the entire dataset as well as RDFs fit to events falling into one of three magnitude categories. The analysis shows that the Eckhardt filter tends to provide better results across events and that the use of categorized parameters can provide some improvement in estimating baseflow across a range of event types. The work

presented is thorough and is explained with a commendable level of detail. The manuscript is organized nicely and is well written, making it easy to quickly understand what was done and the nature of the results. The authors provide a good explanation of the methods and the figures and tables support the interpretation and conclusions given.

Response: Thank you for your positive and encouraging feedback. We are pleased to hear that you found the manuscript well-organized, clearly written, and methodologically sound, and we are grateful for your support in strengthening the manuscript.

One criticism I might offer is that this work is based on and potentially limited in relevance to a single (and comparatively small) watershed. The manuscript would be more impactful if the value of the analysis is situated in a more generalized context - what about the results would be useful to a researcher or practitioner needing to understand or estimate baseflow somewhere else? I would suggest a potential expansion of the discussion and interpretation to address this broader context a bit more.

Response: We appreciate the reviewer's insightful suggestion to strengthen the broader relevance of our findings. To address the limitation of focusing on a single, small watershed, we have substantially revised ***Section 5 (Conclusions and Recommendations)*** *by adding five new paragraphs. These additions provide practical guidance for researchers and practitioners on how to adapt the event-specific calibration and tracer-based validation approach to a variety of catchment types, particularly under variable hydroclimatic conditions. We highlight the transferability of the recommended BFImax value (0.653) as a suitable starting point for catchments with similar characteristics and suggest alternative strategies for data-scarce regions, including the use of more widely available tracers such as electrical conductivity. Additionally, we expanded the final two paragraphs of* ***Section 5*** *to outline at least four key opportunities for future research. These include applying the method across diverse catchments and seasons, integrating machine learning for automated calibration, exploring regionalization techniques, and developing systematic protocols to guide the use of event-based versus general calibration. Together, these additions aim to position our study as a foundation for advancing practical, scalable baseflow separation strategies.*

Specific and Technical Comments

While the authors describe the physical characteristics of the watershed, a more specific explanation of the conceptual model of the watershed would be helpful in providing a basis for interpretation of the baseflow separation results. For example, it would be helpful to discuss what the Si data (or other previous work in the watershed) suggests about how baseflow contributes to streamflow. This can be used to provide context for the different methods and their interpretation too.

Response: We appreciate the referee's suggestion to include a conceptual model of the Arvorezinha catchment to contextualize the baseflow separation results. We included a new paragraph in ***Section 2.1*** *(last paragraph), describing our conceptual knowledge and understanding of the catchment's flow pathways, supported by previous studies carried out in the catchment. We hope these additions clarify the hydrological basis for our methods and enhance the interpretation of results for broader applications.*

What does the difference in performance from the tiered/categorized RDF parameter fits tell us about (potential) mechanisms in the watershed? This is an interesting result and I would appreciate a little more discussion on the topic.

Response: We agree that this is a crucial point and have added two new paragraphs in ***Section 3.5*** *to elaborate on the hydrological mechanisms underlying the variation in optimal BFImax and Beta parameters across different event magnitudes. Specifically, we explain that higher BFImax values observed during low-magnitude events reflect dominant groundwater contributions, as baseflow sustains most of the streamflow under these conditions. In contrast, lower BFImax values during high-magnitude events indicate increasing dilution of baseflow by rapid surface runoff. For the Beta parameter, we discuss how lower values during low-magnitude events suggest streamflow is largely governed by sustained groundwater discharge, while higher Beta values associated with high-magnitude events point to streamflow dominated by quickflow inputs and a reduced baseflow component.*

The description indicates the area receives relatively consistent precipitation, perhaps more than many other places in the world. This would suggest fairly constant GW levels and perhaps less variation in "old" water. Yet the results show a lot of variability in shape and timing of baseflow. I'm curious how this is to be interpreted, and whether the authors view this tier-type approach as something than can be useful in disentangling seasonal variation in groundwater-surface water connection? Again, some additional discussion addressing this would be appreciated.

Response: We respectfully disagree with the assumption that consistent annual precipitation necessarily results in stable groundwater levels or minimal variation in the "old water" (baseflow) contribution. While the Arvorezinha catchment exhibits relatively stable annual rainfall totals (average of ~1,938 mm), this precipitation is distributed unevenly throughout the year and often delivered in short, high-intensity events, particularly in the spring. Such intra-annual variability — combined with the catchment's steep topography, shallow soils, and fractured basalt geology — promotes rapid drainage and episodic recharge, leading to dynamic groundwater levels and highly variable baseflow contributions. This is evident in the diverse shapes and timings of the baseflow hydrographs across events (Figure 5), even within the same season. Therefore, consistent rainfall at the annual scale does not equate to uniform hydrological responses at the event scale, and variability in "old water" contributions remains an important feature of this system.

Regarding seasonal variation, the event-based (tier-type) classification (low, medium, high magnitude) is driven by runoff response — not necessarily aligned with seasons. For example, low-magnitude events occurred in both spring (e.g., Event 5) and winter (e.g., Event 4), suggesting that hydrological behaviour, not season, governed parameter variation. So, while the event-based calibration successfully reveals variability in BFImax and Beta as a function of event magnitude, it does not support inference about seasonal variability in these parameters. Drawing seasonal conclusions would require: A larger number of events, more balanced seasonal coverage (including summer), and stratified analysis of events by both season and magnitude. We have added this as a recommendation for further research (see the second-to-last paragraph of ***Section 5 – Conclusions and Recommendations****).*

Much of the information shown in error metrics tables is repetitive between the two sections of analysis and results. I'd suggest condensing some of this and/or moving some of it to an appendix/supplement.

Response: We appreciate your suggestion, and we have moved Tables 9 and 10 (from the original manuscript) to the Appendices (now Appendices A1 and A2). This suggestion aligns with the first referee's comment about improving the presentation of results by using more visuals than tables, so we added two new figures in the results section.

Table 1 (page 8) - Units for Qmax and Qmin are different. Check flow units for consistency.

Response: Thank you for your suggestion. To maintain consistency with the units for Qmax, we have expressed Qmin in m3/s.

Figure 2 and Figure 5 - I appreciate the information density on these plots. However, the line widths on the RDF results (colored lines) came through rather faintly on the PDF I reviewed, making it difficult to readily differentiate among the results. Additionally the figure seemed a bit blurry. To the extent that the line widths and resolution could be adjusted, these figures will be much more effective.

Response: We have increased the resolution of Figures 2 and 5.

Line 558: "events 2,9, and 9 showed..." - I suspect the second "9" was in error here.

Response: We have fixed this typo.