Dear Editors,

We sincerely thank the Editor, Associate Editor, and Reviewers for their constructive and helpful comments for our manuscript entitled ***“Regional Base-Flow Index in Arid Landscapes Using Machine Learning and Instrumented Records”***. We have revised the manuscript accordingly and provide a detailed point-by-point response below. Reviewer comments are reproduced in *italics*, followed by our responses. Line numbers refer to the revised manuscript.

**Editor’s Comments**

1. *In the Introduction, explicitly formulate your research questions or hypothesis, emphasizing the novelty of your work, and in the final paragraph, clearly state the specific objectives of the paper.*
   * The Introduction explicitly describes the research questions and clearly stated objectives in the final paragraph (Lines 98-110).
2. *The journal adheres to the FAIR Data principles (please see author guidelines) of transparency and accessibility. Please add a statement and provide access to the data either through links or Supplementary Materials.*
   * Data is available at https://github.com/CaelumMroczek/baseflow-AZ .

**Reviewer #2 Comments**

1. Throughout the document: The word "baseflow" or "base flow" (including "base-flow + noun" for the latter) is used inconsistently throughout the document. Either: (a) Change all "baseflow" to "base flow", including in the highlights and any applicable figure; or (b) Change all "base flow" and "base-flow" to "baseflow".
   * We standardized terminology throughout the manuscript, using “base flow” as a noun and “base-flow” as an adjective.
2. Line 27 (Study Focus) and Figure 1 caption: Please define "HUC" in both places.
   * We defined “HUC” as *Hydrologic Unit Code* at first use and in the Figure 1 caption (Line 27, 127; Figure 1 caption).
3. *Lines 56–57: “…changes in land cover/land use…” The spacing around the slash is inconsistent. Consider either (i) land cover/use or (ii) land cover / land use. Choose the form that best reflects the intended meaning*
   * We revised the text to “land cover / land use” for consistency (Lines 56-57).
4. Throughout the document: There is a mix of "gauge" and "ungauged", but "streamgage" is also used. I suggest changing all instances of "streamgage" to "stream gauge" for consistency.
   * All instances of “streamgage” were changed to “stream gauge”.
5. Line 119: Please define "masl" at first use.
   * We define “masl” as “meters above sea level” at the first mention (Line 119).
6. *Line 131: “…evaporation and dry preceding soil properties leads to most…” The verb should be plural.*
   * Corrected “leads” to “lead” (Line 131).
7. Lines 324–325, 338–340, 485–486 and Figure 4 caption: “Statistically significant decreasing trends are observed at 16.1% of sites, while increasing trends are found at 8.8% of sites.” The terms "decreasing" and "increasing" could be interpreted ambiguously.
   * Revised wording to “downward” and “upward” trends for clarity (Lines **322, 335-348, 479-480**; Figure 4 caption).
8. Table 1 vs Figure 6 vs Line 391: “Land Cover – Herbaceous,” “grassland,” and “shrubland” are not used consistently.
   * Standardized terminology to “herbaceous” across the text and Figure 6. We also changed “Water Body” to “Open Water” in Figure 6 for consistency (Line **388**; Figure 6).
9. Line 396: I suggest spelling out "USDA" at first mention.
   * We defined “USDA” as “United States Department of Agriculture” is now spelled out at first mention (Line 393).
10. *Line 435: “The downstream decline in BFI Figure 3 is consistent…” This sentence is incomplete or awkwardly phrased.*
    * Revised to: “The downstream decline in BFI illustrated in Figure 3 is consistent with the transition from headwater recharge zones to more arid, losing stream reaches. This highlights the importance of hydrogeologic setting in shaping base-flow dynamics (Winter 2007)” (Lines 430-433).
11. Line 451: Please spell out "CONUS" at first mention.
    * We defined “CONUS” as “continental United States” at first mention (Line 446).
12. *The reviewer suggested Rumsey et al. (2015) and Jaffrés et al. (2021) as relevant literature on elevation and baseflow/BFI relationships.*
    * We reviewed these papers and added reference to them in the Discussion to highlight comparable settings (Jaffrés et al. 2021; Line 446) and previous studies (Rumsey et al. 2015; Line 447).

**Reviewer #2 Comments**

**Major Comments**

1. While the final model was trained and applied across the entire state, did the authors evaluate performance differences across physiographic regions (e.g., Colorado Plateau vs. Basin and Range)? If not, would a regionally stratified validation improve understanding of model bias or spatial overfitting?
   * This is mentioned in paragraph 3 of the Machine Learning section and paragraph 2 of the Discussion. We did compare the state-wide model with multiple models trained separately for physiographic regions. The region-specific models both showed overall worse performance than the state-wide model. We predict that this is an instance where the state-wide model having access to more data allows it to perform better than the regional-specific models which each have less data.
2. *The manuscript describes the use of SHAP values for predictor selection in the reduced XGBoost model. Were highly correlated predictors (e.g., precipitation and ET) assessed for multicollinearity? Could collinearity have influenced SHAP-based feature importance rankings?*
   * We evaluated multicollinearity of predictors using variance inflation factors (VIF). All predictors had VIF values < 5, with precipitation (4.3) and AET (4.2) being the highest. These results indicate moderate but not problematic collinearity, consistent with common thresholds (VIF < 5). As such, SHAP rankings are unlikely to be influenced by multicollinearity.
3. The study uses calendar-year PRISM data instead of a water-year format. Could inter-annual snow storage effects still introduce systematic error in high-elevation, snow-dominated basins? Would using water-year data improve robustness?
   * We acknowledge that water-year formatting may improve robustness in snow-dominated basins. However, in practice, the model consistently performed better in snow-dominated and Plateau basins than in monsoon-dominated basins. Additionally, snow in northern Arizona typically melts within the same winter, reducing systematic bias.
4. It would strengthen transparency if the authors discussed predictors that were considered but excluded (e.g., specific PRISM variables like temperature).
   * We added a Supplemental Table listing the full set of initial predictors. This addition is referenced in the Machine Learning section (Line 255).
5. Table 4 indicates the model underpredicts annual BFI in monsoon-dominated (pbias ≈ –14 %) and warm-dry (pbias ≈ –12 %) basins. Have the authors examined residual plots or partial-dependence profiles to identify which predictors drive this bias?
   * We examined residual plots and partial-dependence profiles. In both basin types, residuals showed only weak and inconsistent associations with predictors (|ρ| ≤ 0.05). This suggests that the modest bias is not driven by a single factor but rather reflects combined interactions among multiple predictors. Residual analysis is presented below for support.

A screenshot of a graph

AI-generated content may be incorrect.

1. Could the authors partition trend magnitudes by dominant hydrogeologic setting (karst vs. alluvial vs. volcanic) to test whether geologic framework, rather than climate regime, explains spatial scatter in trend directions?
   * We initially considered hydrogeologic setting, but geologic variables were removed during dimensionality reduction of the machine learning predictors. Additionally, statewide geologic maps lack the resolution needed for this analysis.
2. Could the authors elaborate on how predictive uncertainty might be reduced in the absence of improved streamgage coverage? For example, could remotely sensed, citizen science, or proxy data be incorporated?
   * In the Discussion (Lines 451-462), we note that predictive uncertainty could be reduced by incorporating additional data sources and the current limitations for improving model performance.

**Minor Comments**

1. *Could the authors clarify the criteria used to include HUC-8 units in the analysis? Specifically, was there a minimum percentage of area within Arizona required for a HUC to be considered?*
   1. No minimum area threshold was applied; any HUC-8 partially within Arizona was retained.
2. *The manuscript mentions different climate types (like "monsoon-dominated" and "warm-dry"). Consider using consistent labels or explaining how these zones were defined.*
   1. The classifications are defined in the Classification Trends section (Lines 325-334).
3. *Figures 3 and 4 could use a bit more detail in their captions to help readers understand the key patterns without reading the main text.*
   1. More result-oriented details were added to the captions of Figures 3 and 4.
4. *Please remove the tracked changes and comments from the manuscript. Only the clean version should be submitted.*
   1. We submitted both a clean version and a tracked-changes version; the clean version removes all tracked edits and comments.

We hope that the revisions address the reviewers’ and editor’s concerns. Both a tracked-changes version and a clean version of the revised manuscript have been submitted. Thank you for your consideration. We look forward to the opportunity to contribute to your journal.

Sincerely,  
**Caelum Mroczek** (corresponding author)  
*on behalf of all co-authors*