Point by Point Response to Comments for manuscript entitled: “Hydrologic Model Parameter Estimation in Snow-Dominated Headwater Catchments Using Multiple Observation Datasets”

North et al., in review at HESS

Reviewer 1 Comments

I reviewed the manuscript “Hydrologic Model Parameter Estimation in Snow-Dominated Headwater Catchments Using Multiple Observation Datasets” by North et al. This study investigates the value of integrating diverse observational datasets (streamflow, satellite-derived snow, soil moisture, and evapotranspiration) for parameter sensitivity and estimation in the pywatershed hydrological model across four snow-dominated headwater catchments. Using Morris screening and Monte Carlo filtering, we find that while alternative observations consistently identify more informative parameters than streamflow alone, their impact on final streamflow performance is highly catchment-specific. Only actual evapotranspiration (AET) data reliably improved simulations, whereas snow and soil moisture datasets yielded inconsistent results, sometimes degrading performance. The work underscores the context-dependent utility of multi-observational calibration, highlighting challenges such as equifinality, spatial representativeness, and model-observation alignment that must be addressed to effectively leverage new data sources in hydrological forecasting.

This is a timely, well-executed, and intellectually rigorous study that makes a valuable contribution to the field of hydrological model calibration. The experimental design is robust, leveraging state-of-the-art sensitivity and uncertainty analysis techniques (Morris method, LHS Monte Carlo filtering) applied to a relevant and modern modeling code (pywatershed). The manuscript is clearly written, and the analysis convincingly demonstrates both the potential and the pitfalls of integrating diverse, often spatially mismatched, observational datasets. The conclusions are supported by the data presented. I only have a few specific comments in the annotated manuscript file for further clarify the narrative and implications of the work. I would recommend publication after minor revisions.

> AC1 Thank you for this review. We’re glad to hear that you found the study to be a valuable contribution to the hydrologic modeling literature. We’ve incorporated your detailed comments from the annotated PDF, and we reply to the comments that are more complex than simple typographical edits below. In our responses, we have denoted reviewer comments in italics and our responses in plain text.

In-text comments:

Line 12: Delete “only”, add “alone”.

> Deleted, added.

Line 13: Add “-”.

> Added.

Line 87: Add “the”.

> Added.

Line 103: Add “-”.

> Added.

Line 106: Replace “How does the use” with “What is the impact”.

Replaced.

Line 118: Replace "homogenous" with "homogeneous", add "-"

Replaced, added.

Line 119: Add "-"

Added.

Line 140: Why not putting them in one panel so that there will be four panels left in Figure 1. In this case, the panels may align better?

We created a new study site visualization that reduces clutter, provides improved geographical context, and reduced ancillary labeling.

Figure 1: This can't be seen.

These ancillary labels have been removed in the revised figure 1.

Line 151: Delete "-"

Deleted.

Line 154: Replace "was" with "were".

Replaced.

Line 179: Replace "limits the realization" with "constrained the identification".

Replaced.

Line 213: Replace "statical" with "statistical".

Replaced.

Line 214: Replace "was" with "were".

Replaced.

Line 271: Replace "10 to 16" with "10-16".

Replaced.

Line 282: This [black squares and triangles] should be displayed in a legend in the figure.

Figure 2 has been updated to include these elements in the legend.

Line 288: related to what? Cold, wet conditions?

Updated the text to read: "For example, jh_coef was more sensitive in warm, dry conditions than in cold, wet conditions."

Line 295: I would suggest to show the original value on the color bar rather than the log values.

We updated figure 3 to show the original values.

Line 306: What is the significant level?

We use $\alpha < 0.05$ to determine significance and updated the manuscript to reflect that.

Figure 5: Perhaps a better name that can reflect the intersections?

We updated the legend title to “Criteria” to be consistent with the labeling in Table 2.

Line 362: Replace “is influenced by” with “for different”

Replaced.

Line 362: Replace “with” with “between streamflow and”

Replaced.

Line 363: It seems that this is repeating with your legend.

Statement deleted.

Line 396: Based on?

Replaced “in terms of” with “for soil water” to clarify the subject of discussion.

Line 412: Replace “or” with “and”

Replaced.

Line 425: Replace “(Markstrom et al., 2016)” with “Markstrom et al. (2016)”

Replaced.

Line 429: Add “-”.

Added.

Line 434: Replace “to no” with “, if any, “.

Replaced.

Line 435: Replace “to” with “for“.

Replaced.

Line 435: Replace “large” with “strong“.

Replaced.

Line 436: Replace “identified as” with “associated with“.

Replaced.

Line 456: Add “of”.

Added.

Line 525: Replace “have” with “has“.

Replaced.

Line 527: Replace “catchments that border each other (the East and Taylor River)” with “adjacent East and Taylor River catchments“.

Replaced.

<u>Reviewer 2 Comments</u>

This manuscript presents an interesting and timely evaluation of how multi-source observational datasets can be used to constrain the pywatershed in snow-dominated headwater catchments. The integration of diverse data products (such as ASO SWE and OpenET) to address the pervasive issue of model equifinality is a significant undertaking that holds substantial value for the hydrological community. The study's focus on identifying which observations provide truly informative vs. potentially misinformative constraints is particularly important for advancing operational modeling practices. However, while the research is promising, there are several points that warrant further discussion and clarification before the paper can be considered for publication. I recommend Major Revision.

Thank you for the review and for your suggestions. We're pleased to hear that you found the study important for advancing operational modeling practices. We've responded to your detailed comments below, and think the manuscript is substantially improved. We look forward to any additional feedback as needed.

Major Comments:

Q1: The sensitivity analysis identifies forcing correction parameters, such as tmax_cbh_adj and tmax_allsnow, as highly sensitive across all target observations. Please explain whether this heavy reliance on "tuning parameters" to correct meteorological input errors might be masking structural deficiencies in the model's snowmelt or runoff generation mechanisms.

AC2.1 Thank you for this comment. It is not unexpected that these forcing correction parameters rank highly in the identifiability analysis, given their impact on many hydrological processes. We think there are a few important factors to consider here, and we include language in section 4.2 to clarify the plausibility of structurally compensating errors, and it reads as follows:

> "It is plausible that the high sensitivity/reliance on forcing adjustments could mask structural deficiencies in the model's snowmelt and runoff generation mechanisms. In that case, we would expect the parameter values to exhibit concentration at the bounds, pronounced shifts when SWE constraints are introduced, or dramatic shifts across similar catchments. Instead, we see that the inter-quartile ranges are similar between catchments and calibration criteria, and that the range of acceptable values is wide. This signals substantial equifinality, rather than compensation for structural bias. The similarities between the discharge-only and discharge + SWE criteria suggest that most of the information about the parameter value is embedded in the discharge signal. We acknowledge that structural uncertainty cannot be assessed without testing additional model formulations; however, the current evidence does not suggest that these forcing correction parameters mask deficiencies in the snowmelt or runoff mechanisms."

In section 4.2 of the manuscript, we first note that these highly identifiable parameters also exhibit high Morris variances, indicating strong interaction and non-linearity (Figure S16). Then, we have added language that further interprets our results in the context of structural errors versus equifinality. Altogether, these results are consistent with interaction-driven equifinality. However, we acknowledge that structural error remains a possibility. In this case, we emphasize that the value of the work lies primarily in assessing the behavior of a model that is widely used for national assessments and may be representative of other models that are structurally similar (i.e. process-based, lumped models).

Finally, we argue that the inclusion of these forcing correction parameters provides a unique opportunity: substantial literature argues that forcing uncertainty is one of the most important sources of hydrologic model uncertainty (e.g. Tang et al., 2023). While not a holistic assessment of the impact of forcing vs parameter uncertainty, our findings regarding the importance of tmax_all_snow also support the importance of forcing uncertainty.

Q2: The Morris analysis indicates that almost no runoff-related parameters were identified as sensitive in this study. Please explain if the pywatershed model overemphasizes meteorological forcing corrections while potentially neglecting the regulatory role of catchment topography and geological attributes on runoff flow paths.

AC2.2 Thank you for this comment. We think it's closely aligned with the first comment above. We note that we did find a few parameters related to runoff generation that were sensitive; these include soil_moist_max and gwflow_coef.

It is certainly true that runoff generation is simplified in pywatershed: while many components of the model are physics-based, the conceptual soil storage buckets necessarily yield a more conceptual, less physics-based approach to runoff generation pathways. In the revised manuscript, specifically section 4.2, we now include additional reasoning as to why runoff parameters are not identified as sensitive. The modified text reads

> "For the third reason, we find parameters outside of the runoff module, such as soil_moist_max and gwflow_coef, are more identifiable with respect to discharge than parameters within the runoff module itself. Other studies have arrived at similar conclusions, noting that the runoff process requires the largest number of parameters to explain 90% of its sensitivity, while other processes only rely on a few (Markstrom et al., 2016). Parameters within other process groups, such as soil and PET, explain runoff sensitivity in a greater proportion of CONUS than runoff parameters like smidx_coef and smidx_exp (Markstrom et al., 2016). Since soil_moist_max and gwflow_coef represent processes relevant to catchment geology, like infiltration and groundwater flow velocity (Table S1), this finding is logically coherent and supported by previous work."

Q3: The results show that the adjacent East and Taylor River catchments respond quite differently to the same observational datasets. The author suggests that the performance discrepancy between the adjacent East and Taylor River catchments may result from differences in hydroclimatic variables such as runoff ratio or aridity. However, these variables are not explicitly quantified or compared for the two study basins in the manuscript. To support this claim, the author should provide a table or figure comparing the key physiographic and hydroclimatic attributes (e.g., mean slope, forest cover, runoff ratio, and aridity index) of these two basins to explain why the model's physical consistency varies so significantly between them?

AC2.3 We recognize that Table S2 could have been referenced to better support this discussion. The data in Table S2 presents several differences between the two, such as the Taylor catchment having nearly half of the drainage area of the East catchment, a higher mean elevation, and lower annual average precipitation. Forest cover is different by about 6%. While all catchments in this study encompass headwaters, the Blue and Taylor catchments have similar attributes in terms of drainage area and elevation, and they exhibited the poorest performance with respect to discharge. The largest catchment, the Dolores, exhibited the best performance – this leads us to suspect that catchment size and proportion of high elevation headwater area influence the model's performance. However, a study with a much larger number of catchments would be necessary to assess this. We agree that additional hydroclimatic and physiographic attributes should be computed to further support this finding. In the revised manuscript, we add mean slope, annual average discharge, runoff ratio, aridity index to Table S2 to provide further insight into the differences of the neighboring East and Taylor catchments. This is used to add to the discussion in section 4.4, which reads:

> "We find that runoff ratio and aridity are similar across the catchments in this study; however, we note a 7% difference in runoff ratio between the East and Taylor catchment, which may influence

> the differences we observe in streamflow performance when including soil moisture data (Table S2)."

Q4: Does the structure of the pywatershed model may struggle to remain compatible with the simultaneous integration of high-spatial-resolution (ASO) and high-frequency (SMAP) remote sensing data? Table 2 reports several instances where the number of behavioral parameter sets is zero when certain observations are included. My primary concern lies in the potential for model structural failure under these advanced constraints.

AC2.4: As far as spatial compatibility, we will clarify that we computed area averages with respect to the hydrologic response units for ASO and SMAP. We made note of this aggregation for SMAP (line 173) and OpenET (line 181) but appear to have omitted this detail for ASO and MODIS-SCA. This detail is added to the revised manuscript on line 165. Area averaged gridded observations are spatially analogous to what is represented in the model.

For the case of ASO, our criteria expects the model to represent snow water equivalent on the days ASO is acquired. We believe this strict criterion is justified by the water resource implications of peak snow water equivalent, and that the ASO acquisitions used in this study occur near the day of peak snow water equivalent. For the case of SMAP (which is natively at a 3-hour cadence), we compute daily averages to match the temporal resolution of the model. However, there is conceptual misalignment between observed soil moisture and what is represented in the model, which is discussed in detail in section 4.1. We posit that the lack of behavioral parameter sets for soil moisture observations is attributable to this misalignment. Because pywatershed has a daily timestep, we do not think that temporal misalignment is likely to be a major source of structural uncertainty.

Q5: While Actual Evapotranspiration (AET) is the only variable to consistently improve simulation performance, it is evaluated using monthly averages. Could this improvement be an artifact of the temporal averaging smoothing out daily error variances?

AC2.5 Thank you for bringing attention to this important caveat. It is possible that this improvement could result from the temporal smoothing of AET, and we made note of this in section 4.3. In the revised manuscript, we build upon the discussion of temporal resolution in modeling choices, and how that may impact our calibration results. The modified text reads:

> "However, previous work has noted that models calibrated to monthly streamflow yield unrealistic daily values while models calibrated to daily data (Adla et al., 2019) and others have highlighted the relationships between model complexity, process timescale, and the spatial and temporal resolution of calibration data (Kavetski et al., 2011). We hypothesize that parameter estimation for daily streamflow is constrained beneficially by monthly AET because it provides more holistic information about the catchment water balance, but the examination of this hypothesis warrants further work and is not currently possible with OpenET."

Minor Comments:

Q1: The layout of Figure 1 appears somewhat cluttered and would benefit from a clearer visual hierarchy to help readers quickly identify the study's geographic context.

AC2.1m Thank you for this comment. As suggested by the other reviewer 1, we created a new study site visualization that reduces clutter, provides improved geographical context, and reduced ancillary labeling.

Q2: In Figure 4, the Spearman correlation R values and p-values overlap with data points in several subplots, hindering legibility. Please adjust the placement of these statistical annotations to improve clarity.

AC2.2m Thank you for this comment. In the revised manuscript, we manually adjust the positioning of the labeling for the R values and denoted statistically significant correlations with an asterisk to reduce label clutter. The significance threshold is declared in the caption.

References

Adla, S., Tripathi, S., and Disse, M.: Can We Calibrate a Daily Time-Step Hydrological Model Using Monthly Time-Step Discharge Data?, Water, 11, https://doi.org/10.3390/w11091750, 2019.

Hay, L. E., LaFontaine, J. H., Beusekom, A. E. V., Norton, P. A., Farmer, W. H., Regan, R. S., Markstrom, S. L., and Dickinson, J. E.: Parameter estimation at the conterminous United States scale and streamflow routing enhancements for the National Hydrologic Model infrastructure application of the Precipitation-Runoff Modeling System (NHM-PRMS), Techniques and Methods, U.S. Geological Survey, https://doi.org/10.3133/tm6B10, 2023.

Kavetski, D., Fenicia, F., and Clark, M. P.: Impact of temporal data resolution on parameter inference and model identification in conceptual hydrological modeling: Insights from an experimental catchment, Water Resour, https://doi.org/10.1029/2010WR009525, 2011.

Markstrom, S. L., Hay, L. E., and Clark, M. P.: Towards simplification of hydrologic modeling: identification of dominant processes, Hydrology and Earth System Sciences, 20, 4655–4671, https://doi.org/10.5194/hess-20-4655-2016, 2016.

Tang, G., Clark, M. P., Knoben, W. J. M., Liu, H., Gharari, S., Arnal, L., Beck, H. E., Wood, A. W., Newman, A. J., and Papalexiou, S. M.: The Impact of Meteorological Forcing Uncertainty on Hydrological Modeling: A Global Analysis of Cryosphere Basins, Water Resources Research, 59, e2022WR033767, https://doi.org/10.1029/2022WR033767, 2023.