

**A Summary of Revisions and Responses on “A spatially hybrid hydrological modeling approach using subbasin-specific model structures”  
(Ref: ENVSOFT-D-25-02328)**

**With regards to comments from the Reviewer #1**

*Reviewer #1: The authors present a well written manuscript but there are questions that must be answered before the paper may be published. There are also spelling and grammar errors listed after the main questions and comments.*

We sincerely thank the reviewer for the encouraging evaluation of the manuscript and for raising these critical issues that have significantly improved the rigor and clarity of our study. In the revised manuscript, we have made substantial revisions to address all the issues raised in the comments.

(1) We have clarified the mechanisms driving the performance improvement of the spatially hybrid model structure. The revision emphasizes that this improvement stems from the mechanistic complementarity of the hybrid structure rather than merely an increase in parameter degrees of freedom due to regionalization.

(2) We have explicitly defined the topographic criteria used for subbasin classification. The revision acknowledges the proof-of-concept nature of the current classification, while outlining potential directions for future research, such as domain knowledge-driven methods to facilitate constructing appropriate model structures for specific application contexts.

(3) We have justified the rationale of the commonly adopted assumption of neglecting lateral groundwater exchange across subbasin boundaries in the case study. Furthermore, we discussed the extensibility of the SEIMS framework to incorporate grid-based groundwater simulation modules in future work.

(4) We have clarified that the spatially hybrid model constructed in the case study achieves a balanced improvement representing a trade-off between a higher NSE and a lower PBIAS, rather than claiming overall superiority.

(5) We have provided the definition of “semi-physically-based” models within the context of this study and explained how frozen soil processes are handled through simplified physical constraints.

(6) We have elaborated on the framework’s design regarding parameter consistency, specifically detailing the “parameter group” mechanism that allows users to enforce constraints across subbasins.

Furthermore, we have carefully proofread the entire manuscript to correct the spelling

and grammar errors. We believe these revisions have substantially addressed the reviewer's concerns and strengthened the contributions of this study.

### *Main Comments/Questions*

*1. Since your own results demonstrate that regional calibration (without internal data) can degrade mass balance performance (PBIAS), does the success of the Hybrid Model (HybM) rely on error cancellation between the upstream and downstream subbasins? How can you attribute the NSE improvement to the "hybrid structure" rather than simply the increased degrees of freedom allowed by splitting the watershed's parameter sets? This needs to be addressed in the manuscript.*

Thank you for this insightful question. We appreciate the opportunity to clarify the mechanisms behind the model performance, particularly regarding the concerns about error cancellation and the influence of parameter degrees of freedom.

We acknowledge that calibration relying solely on data from a single watershed outlet may inherently introduce error compensation between upstream and downstream subbasins. This is a common challenge in hydrological modeling where internal observational data are unavailable, rather than a unique feature of our specific experiments. We agree that the degradation of PBIAS in the regional calibration strategies reflects this issue. In the revised manuscript, we have explicitly discussed this limitation by adding the following statement to the last paragraph of Section 4.2: *"Specifically, incorporating multi-site observational data would help constrain individual subbasin simulations, thereby reducing the reliance on potential error compensation between upstream and downstream subbasins."*

Regarding the attribution of the NSE improvement in the case study, we respectfully submit that the performance gain in the HybM is primarily driven by the complementary mechanisms of the hybrid structure rather than simply increased parameter degrees of freedom (i.e., more calibrated parameters). This conclusion is supported by the comparison of parameter counts and model performance. According to the experimental design, the regional calibration of PhyM2 involved 40 calibrated parameters, representing the highest degree of freedom among all experiments. However, its NSE (0.59) remained stable and showed negligible improvement over the universal calibration strategy (PhyM1; 0.58). In contrast, the HybM uses only 26 calibrated parameters yet achieved a significantly higher NSE of 0.72 during the calibration period compared to all other experiments (range 0.41–0.59). If the improvement were solely due to increased parameter degrees of freedom, PhyM2 should outperform HybM. In the revised manuscript, we have incorporated this comparison into the last paragraph of Section 3.3.1: *"Crucially, the NSE improvement of HybM is driven by the complementary mechanisms of the hybrid structure rather than simply increased parameter*

*degrees of freedom (i.e., more calibrated parameters). This is evidenced by the fact that HybM achieved a significantly higher NSE (0.72) using fewer calibrated parameters (26) than the regional PhyM2 (40 parameters)."*

Furthermore, the qualitative analysis of hydrographs in Section 3.3.2 provides further evidence. The HybM successfully reproduced the sharp rising limbs and peak flows characteristic of the PhyM, which directly drives the NSE improvement. Simultaneously, it retained the baseflow overestimation tendency of ConM, resulting in a less favorable PBIAS. This pattern of inherited strengths and weaknesses demonstrates that HybM faithfully reflects its hybrid configuration. It suggests that the HybM is operating based on mechanistic complementarity rather than mathematically smoothing out all errors to force a statistical fit.

*2. What quantitative metric or threshold (e.g., mean slope degree, elevation variance) was used to categorize the subbasins? Without a specific, reproducible definition of "gently sloping," is the decision boundary arbitrary? If a different threshold were used (shifting 2-3 subbasins from one category to the other), how sensitive is the model performance to this structural choice? In the revised manuscript, these questions need to be discussed.*

Thank you for raising this important methodological point regarding the robustness of subbasin classification.

In the current case study, the classification was explicitly based on the topographic statistics of the subbasins. Specifically, we identified headwater subbasins with a mean elevation above 3400 m and a mean slope greater than  $10^\circ$  as mountainous subbasins and assigned them the fully-distributed semi-physically-based (PhyM) model structure. The remaining downstream subbasins falling below these thresholds were identified as gently sloping subbasins and assigned the lumped conceptual (ConM) model structure. We have added these specific criteria to Section 3.2.3 of the revised manuscript: *"Subbasins were first categorized based on the topographic statistics. Specifically, headwater subbasins with a mean elevation above 3400 m and a mean slope greater than  $10^\circ$  were classified as mountainous subbasins (13 subbasins, shown in Figure 4), while the remaining downstream subbasins were classified as gently sloping subbasins (16 subbasins). Accordingly, PhyM was assigned to the mountainous subbasins, while ConM was applied to gently sloping subbasins."*

Regarding the sensitivity of model performance to this choice, we acknowledge that shifting the classification threshold would reassign subbasins and inevitably alter the hybrid model structure and simulation results. However, the primary aim of this study was to demonstrate the feasibility and potential benefits of the proposed spatially hybrid modeling approach. The current case study serves as a proof-of-concept to validate that the

implementation based on SEIMS can successfully couple different model structures to achieve an improved performance, rather than to identify the single optimal hybrid structure for this specific watershed. We have added a statement to Section 3.2.3 of the revised manuscript acknowledging this scope: *“We acknowledge that these specific classification thresholds and corresponding model assignments serve primarily as a proof-of-concept for the proposed approach and may require further verification in practical modeling contexts.”*

Furthermore, we regard the classification thresholds used in this case study as a preliminary representation of the spatial heterogeneity of the watershed. Conducting a sensitivity analysis on this threshold essentially involves comparing different perceptual models or multiple working hypotheses. This relates to the broader issue of model structural uncertainty, which we have acknowledged and discussed as a prospect for future research in the fourth paragraph of Section 4.2. Additionally, we highlighted a promising future direction in the third paragraph of Section 4.2 regarding the development of domain knowledge-driven methods. These methods and tools aim to reduce the substantial modeling burdens on users (particularly non-expert users), when determining appropriate model structures for specific application contexts. We have added a statement to the third paragraph of Section 4.2 in the revised manuscript: *“For instance, in this study, subbasins were manually classified using specific topographic thresholds to assign different model structures. While this serves as one feasible example, it may not represent the optimal or universally applicable criteria.”*

*3. You explicitly admit that the current architecture ignores inter-subbasin groundwater flow and assumes exchange only happens via the channel network: how does this assumption impact the validity of your current case study in the Babao River watershed, which is highly mountainous? Aren;t you forcing the “physically-based” modules to adopt unrealistic parameter values to compensate for the missing lateral groundwater flow paths?*

Thank you for raising this insightful question regarding the inter-subbasin groundwater flow connectivity.

As stated in the manuscript, the constructed model structures in the case study assume that mass exchange between subbasins occurs only through the channel network and does not explicitly account for lateral groundwater flow exchange across subbasin boundaries. This simplification aligns with standard practices in most semi-physically-based hydrological models, such as SWAT. Its widespread adoption is primarily driven by high data requirements for physically-based groundwater routing algorithms. Therefore, employing this assumption represents a necessary trade-off between data limitations and modeling feasibility in the current case study. We have modified the statement in the third paragraph in Section 2.1 of the revised manuscript: *“Each subbasin model simulates hillslope processes and channel routing. In this study, it is assumed that mass and energy exchanges between subbasins occur*

*exclusively through the channel network, aligning with standard practices in most semi-physically-based hydrological models, such as SWAT.”*

Regarding the concern that this simplification may lead to unrealistic parameter compensation, we have examined the calibrated parameter sets for all experiments as detailed in Table S2 of the Supplementary Material. The analysis confirms that the calibrated parameter values are well-distributed within their predefined rational ranges and do not exhibit clustering at the maximum or minimum bounds. Although parameter values vary across different experimental cases, this reflects the parameter interactions among different hydrological processes, common to hydrological models.

Furthermore, we consider neglecting lateral groundwater flow between subbasins to be a delimitation of this specific case study, rather than an inherent flaw of the SEIMS framework. On the contrary, the modular architecture of SEIMS provides the flexibility to implement grid-based physically-based groundwater flow modules. We have emphasized this point and identified it as a future direction in the second paragraph of Section 4.2 of the revised manuscript: *“Future work should explore flexible inter-subbasin connectivity mechanisms to account for realistic lateral interactions, such as integrating watershed-scale grid-based groundwater module similar to gwflow in SWAT+ (Bailey et al., 2022).”*

*4. Why is the Hybrid Model considered superior when its PBIAS is substantially higher than the spatially consistent physical model? In hydrological modeling, particularly for water resource management, isn't the preservation of mass balance (low bias) as critical as the NSE? Doesn't this result suggest the Hybrid model is “curve fitting” at the expense of physical mass conservation?*

Thank you for this observation regarding the trade-off between NSE and PBIAS.

We fully agree with you that water balance preservation (low PBIAS) is as critical as capturing flow dynamics (high NSE), especially for water resources management applications. In the revised manuscript, we have removed absolute assertions like “superior” and instead describe the performance of HybM as a “balanced improvement.” We explicitly state that HybM *“leverages the strengths of both model structures while inheriting some of their respective limitations”* (e.g., in Section 3.3.1 and Section 5). Its primary advantage is reflected in capturing runoff dynamics, which leads to higher NSE and lower RSR, while its drawback lies in the relatively higher PBIAS (water balance error) during the validation period.

Regarding the concern about curve fitting, in the case study, we employed a multi-objective optimization algorithm for the parameter calibration that treated NSE, RSR, and PBIAS simultaneously as objective functions. The final parameter set for each experimental

case represents a solution on the Pareto front that balances these three objectives.

**Importantly, the experimental results demonstrate that the HybM is capable of maintaining water balance.** As shown in Table 4, during the calibration period, the PBIAS of HybM (10.02%) is actually better than that of both PhyM1 (10.63%) and PhyM2 (17.34%). In the validation period, HybM's PBIAS shifts from positive to negative (-18.28%), indicating a transition from underestimation to overestimation of total water volume. The magnitude of this bias in validation is greater than that of both PhyM1 (2.58%) and PhyM2 (13.9%). This pattern mirrors the behavior of ConM2, which shifts from 10.96% to -21.03%. This indicates that HybM achieves a balanced improvement, trading off gains in NSE against a less favorable PBIAS in the validation period, a characteristic largely inherited from the conceptual module. We have modified the comparative description of PBIAS in Section 3.3.1 of the revised manuscript to avoid misleading conclusions: *“However, its PBIAS performance (10.02%/-18.28%) fell between ConM and PhyM, i.e., it was superior to both ConM cases but inferior to both PhyM cases.”*

Furthermore, we acknowledge that the relatively higher PBIAS of HybM in the validation period is primarily due to the non-conservative and highly parameterized “reservoir” expressions used for baseflow simulation in the integrated GR4J conceptual module. However, this finding precisely highlights the core contribution of the methodological framework proposed in our study. The framework provides the technical feasibility to implement spatially heterogeneous modeling, which gives users the choice to customize model structures to fit specific application contexts.

*5. Given that frozen soil dynamics are a dominant control on infiltration and runoff in this specific watershed, does omitting this process render the “semi-physically based” modules effectively conceptual? If the physical equations (e.g., Darcy’s Law for interflow) are missing the governing variable (frozen ground permeability), can the model truly be classified as “physically based” in this context, or is it a more complex conceptual model?*

Thank you for raising this critical issue regarding the physical realism of the model structure.

The second paragraph of the Introduction in the manuscript explicitly defined the model classification used in this study to distinguish different levels of hydrological process representation. We define “fully-physically-based” representations as those applying fundamental physical laws by solving coupled partial differential equations. In contrast, the “semi-physically-based” representations adopted in this study are defined as employing simplified physical equations for individual hydrological processes that are generally simulated independently. We respectfully submit that a model structure need not be composed entirely of physically-based simulation algorithms for every process to be classified as

physically-based model. The PhyM model structure used in this study includes several “semi-physically-based” process representations, such as the interflow algorithm based on a kinematic wave approximation of Darcy’s law. Therefore, despite lacking a complex frozen soil dynamics module, the model fundamentally operates on simplified physical laws rather than the empirical reservoirs typical of conceptual model structures.

Regarding the specific model structure for the Babao River in this case study, we acknowledge that this high-altitude and cold-region watershed is significantly influenced by freeze-thaw dynamics. In the current implementation of PhyM, this process is not entirely ignored but is addressed through a simplified physical constraint. Specifically, in the interflow module, subsurface flow generation is restricted when soil temperature drops below freezing (0°C). This serves as a simplified physical approximation representing the reduction of hydraulic conductivity in frozen soil.

We fully agree that pursuing higher physical realism for specific watersheds would require developing or integrating more rigorous physically-based SEIMS modules. This is explicitly listed as a priority for future research in Section 4.2. In the revised manuscript, we have modified the text in the first paragraph of Section 4.2: “*For example, while simplified representations of cold-region processes (e.g., disabling subsurface flow when soil temperature is below freezing) were used in this study, a dedicated, physically-based frozen soil dynamics module is not yet available.*”

*6. Beyond the setup burden, how does the framework guard against parameter inconsistency? If two adjacent subbasins have similar characteristics but are assigned different structures (or separate configuration files), what prevents a user from assigning conflicting parameter values (e.g., vastly different soil porosity for the same soil type) across the boundary? Does the framework enforce any parameter continuity constraints?*

Thank you for raising this important point regarding parameter consistency within the framework.

The current design philosophy prioritizes modeling flexibility, allowing users to balance physical consistency with the need to capture spatial heterogeneity. At the model setup stage, consistency is maintained by assigning identical initial values to parameters associated with the same soil or land use/cover types across all simulation units. However, during the parameter calibration stage (e.g., using the multi-objective NSGA-II algorithm in this case study), the adjustment of the same parameter is allowed to vary across subbasins. We believe this divergence is reasonably based on two practical considerations. First, the actual physical properties of the same soil or land use/cover type often exhibit spatial variability within a watershed. Second, allowing differential adjustment serves as a rational way to compensate

for inherent uncertainties in source data (e.g., large-scale soil database).

To address cases where strict consistency across specific subbasins is required, the framework provides a configuration mechanism to enforce parameter continuity constraints. Users can define “parameter groups” that link specific subbasins together. Within such a group, a shared parameter is calibrated to a single and synchronized value across all member subbasins during calibration. This ensures that the user can enforce cross-subbasin consistency when necessary.

In summary, the proposed modeling framework adopts a flexible strategy characterized by initial consistency, permitted spatial differentiation, and user-configurable enforcement of constraints. We have added a clarification on this mechanism in Section 2.2.3 of the revised manuscript: *“This subbasin-specific configuration mechanism naturally facilitates subbasin-specific parameter calibration. To manage parameter consistency across different subbasins, the framework allows users to define “parameter groups.” Within each group, a designated set of subbasins can be linked such that a shared parameter is constrained to a single, synchronized value across all member subbasins during the optimization process. This feature enables users to enforce cross-subbasin parameter consistency based on hydrological understanding or data quality, while still permitting differential calibration in other areas to account for spatial heterogeneity or compensate for source data uncertainties.”*

### *Corrections*

*Page 6 line 101: spatially consistent method and spatially varying method -> spatially consistent methods and spatially varying methods*

*Page 10 line 191: each subbasins -> each subbasin*

*Page 16 line 297: generate -> generates*

*Page 16 line 298: convey -> conveys*

*Page 13 line 233: such the -> such as the*

*Page 4 line 59: vis -> via*

*Page 22 line 405: Law and, with the -> Law and with the*

*Page 26 line 458: assumption that successfully -> assumption that successful*

*Page 30 line 535: spatial heterogenous -> spatial heterogeneity*

*Page 36 line 646: remains central challenges in hydrology community -> remain central challenges in the hydrology community*

*Page 36 line 662: studies in watershed -> studies in watersheds*

Thank you for your careful reading and for pointing out these typographical and grammatical errors. All corrections have been made in the revised manuscript. We have thoroughly proofread the entire manuscript to ensure its clarity and accuracy.

**With regards to comments from the Reviewer #2**

*Reviewer #2: This manuscript demonstrates a new approach that integrates physically-based simulation algorithms with spatially explicit distributed simulation units. The experimental results show improvement through comparison. While the manuscript is well-written, it contains several grammatical errors that require attention prior to publication.*

Thank you for your positive evaluation of our work and recognition of the proposed hybrid modeling approach. As suggested, we have conducted a thorough proofreading and language polishing of the entire manuscript to ensure clarity, accuracy, and compliance with publication standards.