Authors’ Response to Editors/Reviewers of

Response to Reviewer 1 on “A Raster–Vector Framework for Multi-Scale Hydrological–Hydraulic Modeling Across Large Domains”

GMD,

RC: ***Reviewers’ Comment***, AR: Authors’ Response, ☐ Manuscript Text

We sincerely thank Reviewer 1 for the careful reading of our manuscript and for the constructive comments provided. Below, we respond point by point to the issues raised and indicate how the manuscript will be revised accordingly.

1. Main comments

RC: ***The manuscript presents a timely and novel study, proposing a method to transforming high-resolution vector river networks into coarse-resolution raster river grids addresses a methodological challenge that has not yet been well resolved in the existing research. This aspect represents a potentially contribution to large-scale hydrological hydrodynamic modeling. However, the aspects of the framework present several limitations regarding maturity, scalability and demonstrated benefits.***

AR: We thank the reviewer for acknowledging the timeliness and novelty of the proposed framework. We respectfully disagree, however, with the characterization of the framework as presenting limitations in terms of maturity, scalability, and demonstrated benefits. As explained in our detailed responses below, these comments highlight the need to emphasize more explicitly the scope and objectives of the manuscript. The objective of the study is not to deliver an optimized forecasting system or to demonstrate improved predictive performance, but to introduce and evaluate a reproducible raster–vector H&H framework addressing the spatial and structural mismatch between grid-based hydrological models and vector-based river hydraulic representations. The framework demonstrates encouraging computational efficiency and is successfully applied to a 50,100 km^2 basin, which already represents a significant domain for such high-resolution H&H modeling and is relevant to a wide range of operational and research applications. The detailed responses below address each of the reviewer’s points in turn and indicate how the manuscript will be revised to make this scope and the demonstrated contributions clearer.

RC: ***The overall workflow appears rather complex, involving multiple intermediate processing steps. The river-network transformation procedure has not been adequately integrated into a unified and streamlined framework, which reduces both the transparency and reproducibility of the methodology.***

AR: We thank the reviewer for this comment. However, we respectfully disagree with this assessment. The preprocessing workflow and the H&H model are both integrated within the open-source `smash` modeling platform (Colleoni et al., 2025), whose source code is openly available. The frozen archive associated with this manuscript, including all preprocessing and analysis scripts, is deposited on Zenodo (Berkaoui et al., 2026), ensuring full reproducibility of the results presented.

As explained in our responses to the detailed comments below, the preprocessing chain was designed as a generic and modular workflow applicable to different DEM sources, target hydrological resolutions, and mapped hydrographic networks, as illustrated in this study through its application at multiple spatial scales to both a national DEM (BDALTI) and a global DEM (MERIT), combined with a reference mapped vector network.

In the revised manuscript, and as detailed in our responses to the methodology comments below, we will streamline the description of the preprocessing workflow and further emphasize its core methodological contributions to improve overall clarity. The corresponding figures and captions will also be revised accordingly.

RC: ***The selected study area is comparatively small (approximately 50,000 km²), which limits the manuscript's ability to demonstrate the stability and transferability of the proposed method for broader large-domain applications. The framework does not yet appear to be readily applicable to large-scale basins or multi-river complex systems, as the computational burden may become prohibitive. This brings concerns regarding its scalability and practical applicability beyond the relatively limited study domain.***

AR: We thank the reviewer for this remark, but we respectfully disagree with the characterization of the Garonne basin as a relatively limited or low-complexity study domain, particularly given the relatively high resolution of the coupled H&H modeling. As detailed in our responses below, the modeled domain covers 50,100 km^2, spans a strong topographic gradient from the Pyrenean headwaters to the alluvial plain, includes contrasted hydro-climatic regimes, and contains a complex and highly anthropized hydrographic system. A basin with these characteristics constitutes a significant and already challenging application domain for a coupled H&H framework at this level of spatial and structural integration.

The computational efficiency demonstrated in this study for the H&H simulations, with simulation times dropping from a few minutes at 1 km resolution to a few seconds at 5 and 10 km resolutions (Table 2), already supports the practical applicability of the framework to larger spatio-temporal domains. We agree, however, that the scalability of the full workflow, including preprocessing, should be discussed more explicitly. In the revised manuscript, we will therefore add information on the computational cost of the preprocessing workflow, complementing the H&H simulation times already reported, and strengthen the discussion on applicability to larger domains. Future testing at larger spatial scales, including regional to national domains, will be clearly identified as an important perspective.

RC: ***While the study emphasizes methodological novelty, it remains unclear whether the proposed framework leads to a substantial improvement in hydrological simulation skill or process representation compared to existing methods. In particular, the overall presentation of the manuscript, including figures and visualizations, requires substantial improvement to meet the standards of clarity. And the results and discussion require stronger evidence to support the claimed methodological advantages.***

AR: We thank the reviewer for this comment. We emphasize that the objective of this manuscript is not to demonstrate improved predictive performance through calibration or optimization, but to introduce and evaluate a consistent raster–vector hydrological–hydraulic (H&H) framework across spatial resolutions.

The methodological contribution lies in the preprocessing and grid–vector coupling strategy, and in its integration within a unified modeling framework, enabling consistent spatial reconciliation between hydrological and hydraulic components across spatial scales.

The main evidence supporting these contributions concerns the geometric consistency of the derived sub-grid river networks across scales, the reduction of catchment-size errors through the sub-grid area representation, the preservation of mass conservation between the hydrological and hydraulic domains, and the computational efficiency of the coupled H&H simulations. In this context, the fact that satisfactory discharge reproduction is achieved at multiple gauging stations without any H&H recalibration at coarser resolutions, alongside the geometric and numerical consistency tests presented throughout the study, constitutes a positive demonstration of the framework's methodological soundness.

We acknowledge, however, that the manuscript should make this positioning clearer to avoid any ambiguity. In the revised manuscript, we will clarify the scope, objectives, and contributions of the study, and will also improve figure clarity, captions, terminology, and visual layouts, as detailed in our responses to the specific comments below.

RC: ***Overall, although the framework appears technically consistent, the manuscript does not yet convincingly demonstrate clear improvements in predictive performance. For these reasons, I do not recommend publication in GMD in its current form.***

AR: We thank the reviewer for this overall assessment. We acknowledge that the scope and objectives of the manuscript should be stated more explicitly. As explained throughout our responses, the present work does not aim to demonstrate optimized predictive performance, but to introduce and evaluate a reproducible raster–vector H&H framework for multi-scale modeling across large domains. We will revise the manuscript to make this positioning clearer from the outset, in order to avoid any misinterpretation of the paper's objectives and contributions.

2. Detailed comments

2.1. Abstract

RC: ***Line 9: define DEM at its first occur.***

AR: We thank the reviewer for pointing this out. We will define DEM at its first occurrence in the abstract.

2.2. Introduction

RC: ***Lines 60–63: the authors clearly identify the primary H&H challenge as the scale mismatch between coarse-grid hydrological models and fine-resolution vector river networks. However, from line 78 onwards, the manuscript shifts to a detailed discussion of channel geometry and bathymetry parameterization (e.g., depth, friction), which is not clearly linked back to this core scale-mismatch problem. The authors should better integrate this discussion into the main methodological narrative.***

AR: We thank the reviewer for this remark. Our intention in the Introduction was to position the present work within the broader set of methodological challenges encountered in large-scale hydrological–hydraulic (H&H) modeling by distinguishing two major challenges. The first challenge is the spatial and structural mismatch between coarse-grid hydrological models and fine-scale river hydraulic representations, which is the primary focus of the present manuscript and is directly addressed by the raster–vector coupling methodology developed here. The second challenge concerns hydraulic model parameterization, particularly river bathymetry, which remains a major limitation for large-scale H&H applications.

We agree that the positioning of the second challenge was not sufficiently explicit. This parameterization challenge is closely linked to the geometric consistency of the drainage-network structure across scales. Indeed, by preserving physically consistent drainage areas at coarse resolutions through the sub-grid representation, the proposed framework effectively mitigates compensation effects inherent to regular grid discretizations, where drainage area overestimation is typically compensated during calibration by less physically meaningful model parameters. The proposed framework thereby provides a more physically consistent basis for future model optimization and data assimilation perspectives. We will add an explicit transitional sentence in the Introduction to clarify the relationship between the two challenges, and will further strengthen this connection in the revised manuscript.

2.3. Methodology

RC: ***Section 2.2.1-2.2.2: These describe standard DEM preprocessing steps which some of these are not part of the methodological novelty and there is redundancy with Section 2.2.3. Remove redundant descriptions. Section 2.2.3-2.2.4: These contain the core methodological contribution. Emphasize and clarify.***

AR: We thank the reviewer for this constructive remark. We agree that the original presentation of Section 2.2 did not sufficiently distinguish between standard preprocessing operations and the core methodological contributions of the manuscript. In the revised version, we will streamline Sections 2.2.1 and 2.2.2, as they mainly describe standard DEM preprocessing and flow-direction computation steps, while retaining the key elements needed to understand how the fine-to-coarse link established by the IHU upscaling is later exploited in Sections 2.2.3 and 2.2.4. By contrast, Sections 2.2.3 and 2.2.4 will be revised to more explicitly emphasize the main methodological steps of the proposed raster–vector coupling workflow, and redundant descriptions across Section 2.2 will be removed or shortened accordingly.

RC: ***Line 219: define GR4 at first instance.***

AR: We thank the reviewer for pointing this out. We will define the acronym at its first occurrence in the revised manuscript.

2.4. Study sites and data

2.4.1 Garonne basin

RC: ***The selected basin (50,000 km^2) is relatively small and predominantly flat. What is the rationale for selecting such a low-relief basin? How would the framework perform in larger and more complex basins? The river network appears simplified, with limited representation of small tributaries.***

AR: We thank the reviewer for this remark, but we respectfully disagree with the characterization of the Garonne basin as a small or low-complexity study domain. As described in the manuscript, the modeled domain covers 50,100 km^2 and spans a strong topographic gradient, from elevations exceeding 3000 m in the Pyrenees to a few hundred meters in the alluvial plain. It also encompasses contrasted hydro-climatic and hydrological regimes, including nival, pluvial, and mixed influences, as well as marked discharge variability ranging from extreme floods to severe low flows. In addition, the basin is highly anthropized, with numerous dams and diversion channels, even though the present study focuses only on natural flow processes. These characteristics make the Garonne a challenging and representative testbed, appropriate in scale, physiographic heterogeneity, and hydrological complexity for evaluating the proposed H&H framework. The Garonne basin is also a well-instrumented reference site, with an extensive gauging network and access to water surface observations from multiple sensors, including the SWOT satellite mission, for which it serves as one of the primary French validation domains. This further supports its relevance as a study site, especially in view of future developments toward the assimilation of water surface observations from space, as discussed in the manuscript.

Regarding the apparent simplification of the river network, the selection of main rivers and major tributaries only was a deliberate methodological choice, as justified in Section 3.2.1: *"This filtering ensures consistency between the mapped river network and the grid resolutions considered (up to 10 km), while preserving realistic drainage patterns and avoiding stream collision artifacts that may arise when upscaling narrow, low-order streams, thereby achieving an optimal balance between hydrographic accuracy and computational efficiency for large-scale H&H modeling."*

RC: ***Lines 65–66: the author refer to the preprocessing step of H&H framework remains computationally intensive and presents ongoing efficiency challenges. The discussion of scalability and computational efficiency is essential.***

AR: We thank the reviewer for this remark. We agree that the computational performance of our preprocessing workflow is not explicitly quantified in the manuscript. We will add in the revised manuscript information on the computational cost of our preprocessing workflow, to complement the computational efficiency of H&H simulations reported in Table 2, in order to support the scalability of the proposed framework.

2.4.2 Mapped river network

RC: ***Figure 4: Add north arrow and scale bar.***

AR: We thank the reviewer for this remark. We note that a scale bar is already present in Figure 4. A north arrow will be added in the revised manuscript.

2.4.3 Digital Elevation Models (DEMs)

RC: ***Line 144 & 296: Define MERIT and BD ALTI at first instance.***

AR: The MERIT acronym will be defined at its first occurrence in the revised manuscript (line 144). Regarding BD ALTI, we note that this is a proper dataset name rather than an acronym: "BD" stands for "*Base de Données*" (i.e., database) and "ALTI" refers to "*altimétrique*" (i.e., elevation). It is the official name of the national elevation dataset produced by the National Institute of Geographic Information (IGN). We will nevertheless clarify this point at its first occurrence to avoid ambiguity.

RC: ***Clarify consistent spatial resolution (MERIT resampled to 100 m, BD ALTI resolution unclear).***

AR: We thank the reviewer for pointing this out. We agree that the description of the BD ALTI DEM resolution should be made clearer. In the revised manuscript, we will explicitly state that BD ALTI is used at its native 25 m spatial resolution on a regular grid, instead of the less direct current formulation referring to elevation data provided on a regular 25 m × 25 m grid (Line 297).

RC: ***If the workflow involves extracting fine-resolution vector rivers and aggregating them to coarse grids, would it be more appropriate to use pre-upscaled DEM products (e.g., MERIT Hydro IHU) to reduce resampling-induced uncertainty?***

AR: We thank the reviewer for this relevant comment. We acknowledge that resampling MERIT DEM from its native 3 arc-second resolution ($\sim$90 m) to a regular 100 m grid may introduce local uncertainties in elevation gradients and flow-direction derivation. This resampling step was nonetheless necessary in our study to ensure multiplicity between native and target spatial resolution during the upscaling step of flow directions to the target hydrological grid resolutions (0.5, 1, 5, and 10 km). This enabled a consistent scale-based comparison between BD ALTI and MERIT DEMs in both the geometric assessment of the preprocessing and the H&H model assessment. We will further highlight in the revised manuscript the resampling procedure applied to MERIT DEM and its potential implications on MERIT-based results. Regarding MERIT Hydro IHU (Eilander et al., 2020), we clearly acknowledge its relevance as a global hydrography dataset providing multiple hydrography products, from flow directions to sub-grid hydrographic variables, at $\sim$1 km, $\sim$10 km, and $\sim$30 km resolutions. We note that our preprocessing workflow is built on the same algorithms used for producing MERIT Hydro IHU, namely the IHU approach (Eilander et al., 2021) and the open-source Python `pyflwdir` package (Eilander, 2022). However, one of the objectives of our work is to develop and evaluate a generic preprocessing workflow that is applicable to different DEM sources, target spatial resolutions, and mapped hydrographic networks. We therefore recomputed the preprocessing chain from the input DEMs rather than relying directly on available hydrography products at predefined spatial resolutions such as MERIT Hydro IHU. This allows us to keep full control over the overall preprocessing chain and its outputs, from DEM conditioning to flow-direction computation and upscaling, hydrography delineation, and grid-to-vector network coupling. Nevertheless, the flexible and modular implementation of the preprocessing workflow within the open-source `smash` framework allows pre-upscaled products such as MERIT Hydro IHU to be integrated as alternative inputs in future applications. We will mention this perspective in the revised manuscript.

2.5. Hydro-meteorological data

RC: ***Precipitation (1 km) and PET (8 km) have inconsistent resolutions. Clarify final resolution of all forcing datasets. Provide detailed description of runoff validation datasets in this section.***

AR: We thank the reviewer for these remarks. We agree that the final spatial resolution of potential evapotranspiration (PET) should be clarified. PET is derived from SAFRAN Météo-France temperature data (Quintana-Seguí et al., 2008; Vidal et al., 2010), originally available at 8 km resolution and downscaled to a 1 km grid, using the Oudin formula (Oudin et al., 2005), to obtain daily interannual potential evapotranspiration at a final spatial resolution of 1 km. In the revised manuscript, we will clarify this point explicitly and state that both atmospheric forcing datasets are ultimately provided at a consistent 1 km spatial resolution. Regarding observed discharge, we agree that the corresponding dataset should be described more clearly in this section. In the revised manuscript, we will provide a clearer description of the observed discharge dataset, including its temporal resolution, the spatial coverage of the gauging stations, and the range of associated drainage areas. The distinction between stations used for geometric preprocessing assessment and those retained for H&H discharge evaluation is described later in Section 4.2, where the corresponding quality assessment methods are introduced.

2.6. Numerical experiments

2.6.1 Spatio-temporal configuration

RC: ***The author only uses the two years, the year of 2018 for spin, the year of 2019 for formal simulations, the author should use longer time series forcing data for the validation of the simulation of runoff by H&H framework or justify this choice.***

AR: We thank the reviewer for this comment. As clarified in our responses to the main comments, the objective of the present manuscript is not to optimize or benchmark long-term hydrological predictive performance, but to evaluate the methodological robustness of the proposed H&H framework and demonstrate its applicability across spatial scales. As explicitly stated in Section 4.1 of the manuscript, the model setup relies on previously calibrated hydrological parameters, hydraulic geometry derived from empirical relationships, and uniform friction, with no recalibration performed at coarser resolutions (5–10 km). This simple setup provides a controlled and sufficient baseline for evaluating the internal consistency of the proposed framework, rather than for delivering an optimized forecasting system. Accordingly, the core claims of the manuscript concern the geometric consistency of the preprocessing in terms of hydrographic network and drainage-area representation across spatial scales, the preservation of mass conservation through the implemented grid–vector coupling framework, and the computational efficiency of the overall framework, rather than predictive performance gains through calibration. In this context, a one-year evaluation period (2019), preceded by a spin-up year (2018) to initialize model states, constitutes a sufficient and controlled experimental basis for evaluating the methodological robustness of the H&H framework under identical hydrological and hydraulic parameterizations across all tested spatial resolutions and drainage-area delineation methods. We will clarify this scope more explicitly in the revised manuscript to justify the chosen simulation period with respect to the paper's objectives.

2.6.2 Aggregation of meteorological forcings

RC: ***This section (4.1.2) is a methodological data description and should be merged into Section 3.2.3.***

AR: We agree with this remark. In the revised manuscript, the content of Section 4.1.2 will be merged into Section 3.2.3, consistently with the clarifications provided in the response above, to further specify that both precipitation and PET are provided at a native 1 km spatial resolution for the baseline simulation, and are subsequently aggregated by arithmetic mean to the coarser model resolutions considered in the experiments

(5 km and 10 km), in line with the spatio-temporal configuration described in Section 4.1.1.

2.6.3 Hydraulic model parameterization

RC: ***Consider scale-dependent or spatially variable Manning coefficients, as they strongly influence discharge simulations.***

AR: We thank the reviewer for this relevant comment. We acknowledge that spatially variable Manning coefficients would better reflect the diversity of channel roughness conditions across the Garonne basin, from steep headwater reaches in the Pyrenees to lowland alluvial channels downstream, and would therefore be important to consider in a more advanced hydraulic modeling perspective. However, as stated previously, the objective of the numerical experiments presented in this manuscript is not to optimize predictive performance, but to evaluate the methodological robustness of the proposed H&H framework. In addition to the demonstrated consistent spatial accuracy of the sub-grid networks across spatial scales (Sections 5.1 and 6.1 of the manuscript), the use of a uniform friction parameterization, together with a scale-consistent hydraulic geometry derived from geomorphological relationships applied to sub-grid drainage areas (Equation 6 in the manuscript), enables us to better isolate the sensitivity of the model to drainage-area representation across scales (i.e., grid-based vs. sub-grid). This is supported by the stable correlation coefficient r across scales (Sections 5.3 and 6.3 of the manuscript), while the bias component β is identified as the primary driver of the performance differences observed between area delineation methods across spatial resolutions.

2.6.4 Geometric assessment of the preprocessing

RC: ***The using of visual inspection is not scalable for large domains. It is unclear how the H&H framework handles: (1) closely connected tributaries; (2) overlapping or merged coarse-grid representations. Provide a systematic or automated evaluation metric, and clarify how river connectivity is preserved during rasterization.***

AR: We thank the reviewer for these relevant comments. Regarding visual inspection, this qualitative evaluation is widely used in the assessment of DEM-derived river networks at various spatial scales, and is recognized as one of the most direct methods for evaluating flow-direction upscaling algorithms, as reviewed by Sousa and Paz (2017). That said, in our study, visual inspection is used as an illustrative complement to the quantitative analysis, in order to help the reader interpret the quality of the derived river networks and to support the discussion of representative local spatial patterns.

As for the quantitative metrics, we would like to emphasize that the metrics used here, namely the mean separation distance (MSD) and the percentage within buffer (PWB), are systematic, fully automated, and scalable to any domain size. The MSD quantifies the spatial displacement between derived and reference networks across all resolutions, while the PWB assesses network completeness with respect to the reference hydrography. Both metrics are widely used in the assessment of DEM-derived networks, as reviewed in Sousa and Paz (2017), and together with the visual inspection directly support the interpretation and conclusions in Sections 5.1 and 6.1. The code to compute all metrics used for the *"geometric assessment of the preprocessing"* is available in the frozen archive on Zenodo at `https://doi.org/10.5281/zenodo.18924808` (Berkaoui et al., 2026).

Regarding closely connected tributaries and overlapping or merged coarse-grid representations, these are indeed well-known limitations of coarse-resolution network derivation methods, and are directly related to the consistency between the level of detail of the reference vector network and the computational grid resolution. In the present study, this issue is addressed upstream of the workflow through the filtering strategy of the reference hydrographic network described in Section 3.2.1, where only the main rivers and major tributaries are retained in order to ensure consistency between the level of detail of the vector reference network and the coarse grid resolutions considered (up to 10 km), thereby avoiding stream-collision artifacts that arise during upscaling when low-order and closely connected tributaries are retained. We will further clarify this point in the revised manuscript while explicitly acknowledging this limitation.

Regarding river connectivity preservation, this is ensured at two levels. The first is at the coarse-resolution grid scale, where river connectivity is guaranteed by the coarse D8 flow directions map produced by the IHU upscaling algorithm. The second level is at the sub-grid scale, where the sub-grid network inherits the connectivity and topology of the coarse-scale river network delineated at the first level, as it is constructed by connecting the outlet pixels of the coarse river network cells by tracing the fine-scale flow direction map. This is described in Section 2.2.3 and will be further clarified in the revised manuscript, in view of the planned revisions of the different parts of Section 2.2 (Preprocessing chain for raster–vector coupling), as outlined by the reviewer.

RC: ***Line 381: clarify definitions of "sub-grid" and "grid-based" methods and ensure consistent usage throughout in the manuscript.***

AR: In order to avoid any confusion, we will systematically use in the revised manuscript the terms "grid-based area" and "sub-grid area" to refer to the two drainage-area delineation methods. We will further emphasize that "grid-based area" refers to drainage areas delineated from the coarse hydrological grid, whereas "sub-grid area" refers to drainage areas computed from fine-resolution flow directions embedded within the coarse computational grid. We will ensure consistent use of these terms throughout the text, figures, and tables.

2.7. Results

2.7.1 Quality assessment of DEM-derived networks across scales

RC: ***Lines 427–428: provide figure or table references to support reported percentages.***

AR: We thank the reviewer for this remark. The reported percentage reductions in MSD (66–95% for BDALTI and 69–83% for MERIT) are derived from the values shown in Figure 6a, by comparing the MSD of sub-grid networks with that of coarse networks for each DEM and spatial resolution. In the revised manuscript, we will provide a table reporting MSD values for both coarse and sub-grid networks across all resolutions and DEMs, along with the corresponding relative improvement.

RC: ***Lines 428–430: current scatter plots (Figure 6a) do not clearly show differences. Consider alternative visualization (e.g., boxplots).***

AR: We thank the reviewer for this remark. We agree that the current representation of Figure 6a can make some differences difficult to distinguish, particularly at fine resolutions (0.5 km and 1 km). We will improve the readability of Figure 6a in the revised manuscript by enhancing the visual contrast between coarse and sub-grid networks across DEMs and spatial resolutions.

Regarding the suggestion of boxplots, we note that MSD is computed at the whole-network scale, as the mean distance between all vertices of the reference network and their nearest projection on the derived network. Therefore, for each DEM, spatial resolution, and network representation, the current analysis produces one aggregated MSD value. That said, we will explore whether the distance distributions used to compute MSD could provide an informative boxplot visualization. The revised figure will also be supported by the summary table described above.

RC: ***Figure 5: All sub-plot adds north arrow and scale bar.***

AR: We thank the reviewer for this remark. As stated in the current caption of Figure 5, the scale bar shown applies to the four main overview maps. In the revised manuscript, a north arrow will be added to the overview maps, and scale bars will be added to the inset zoom panels for each sub-figure.

RC: ***Figure 6: Align the term of "coarse" and "sub-grid" in the figure with that used in the manuscript. Clarify the sample size used in subplots (a), (b), and (d); the number of points appears unexpectedly low. Why not metrics are computed at the sub-catchment or tributary level?***

AR: We thank the reviewer for this comment.

In the revised manuscript, we will systematically use "coarse river network" and "sub-grid river network" when referring to the two network representations derived from the preprocessing, and distinguish them from "grid-based area" and "sub-grid area", used to refer to the two drainage-area delineation methods, as defined in our response to Line 381 above.

Regarding sample size in subplots (a), (b), and (d), we clarify that each plotted point corresponds to one aggregated network-scale metric for a given DEM, spatial resolution, and network representation. As stated above, MSD is computed as the mean distance between all vertices of the reference network and their nearest projection on the derived network. Compared to subplot (a), subplot (d) uses fewer reference network vertices, as omitted features of the reference network are excluded from the computation of adjusted MSD values. We will further clarify this in both the results interpretation and the Figure 6 caption, and specify the number of reference vertices used to compute the MSD metric in subplots (a) and (d). Similarly, for subplot (b), PWB is computed from the full reference and derived sub-grid network geometries and is reported as one value for each DEM and spatial resolution.

Regarding not computing these metrics (i.e., MSD and PWB) at the sub-catchment or tributary level, we acknowledge that a reach-by-reach analysis will provide a finer spatial analysis. However, this requires a robust, automated, and hydrologically consistent correspondence between each reference reach and its corresponding derived reach, which is technically challenging to implement in a fully automated and scalable way over large domains.

We therefore adopted a whole-network evaluation approach, providing a systematic and automated assessment of spatial alignment and completeness across DEMs and spatial resolutions, which is consistent with previous studies (e.g., McMaster, 2002; Davies and Bell, 2009; Sousa and Paz, 2017). This provides a sufficient basis for the conclusions drawn in Sections 5.1 and 6.1 regarding the spatial accuracy of coarse and sub-grid networks across scales.

We will clarify this methodological choice in the revised manuscript.

RC: ***Subplot (c): the stacked bar chart is not an effective way to compare internal vs. headwater differences between DEMs. Clarify the meaning of "omitted reference network length". All subplots should avoid grey backgrounds and instead use a white or very light background for clarity.***

AR: We thank the reviewer for these remarks. Regarding the term "omitted reference network length", we clarify that it refers to the portion of the reference hydrographic network that is not reproduced by the sub-grid derived network, i.e., the parts of the reference network that fall outside the buffer placed around the sub-grid network and are therefore not captured by it. It is the complement of the PWB metric (PWB + omitted length % = 100%). We will provide in the revised manuscript a more explicit definition of this term, along with clearer distinction between headwater and internal omissions in both the results section and the caption of Figure 6.

Regarding the stacked bar chart in subplot (c), we will revise this subplot in the revised manuscript to improve the visual comparison between internal and headwater omissions across DEMs and spatial resolutions, in addition to removing the grey background from all subplots of Figure 6.

RC: ***Lines 439–441: the definitions of sub-grid and grid-based methods are unclear. Clarify the distinction to avoid confusion.***

AR: As noted in our response to the reviewer comment on Line 381 above, we will add explicit definitions of the "grid-based" and "sub-grid" drainage-area delineation methods at their first occurrence in the manuscript, and ensure consistent usage of these terms throughout the text, figures, and tables.

2.7.2 H&H model performance across scales

RC: ***Line 451: Locations of the 18 gauging stations are not clearly shown.***

AR: We thank the reviewer for this remark. We agree that the locations of the 18 gauging stations retained for H&H performance evaluation are not explicitly shown in the main manuscript, and currently appear only in Figures D2 and D3 of Appendix D. As outlined in our previous responses, we will add the locations of the gauging stations to the main manuscript, clearly distinguishing those used for the geometric preprocessing assessment from those retained for H&H model assessment.

RC: ***Unify the term naming convention. The term used in Table 1 (Grid and Sub-grid), Figure 7 (grid-based and sub-grid), and Figure 6 (coarse vs. sub-grid) is inconsistent and may confuse readers.***

AR: We thank the reviewer for pointing out this inconsistency. As already noted in our responses above, we will unify the terminology throughout the text, figures, and tables in the revised manuscript, systematically using "coarse river network" and "sub-grid river network" for the network representations, and "grid-based area" and "sub-grid area" for the drainage-area delineation methods.

2.8. Discussion

2.8.1 Coarse vs. sub-grid networks spatial accuracy across scales

RC: ***Line 487: The omission of headwater segments in zoom2 likely has limited impact on discharge. In contrast, the internal omissions observed in zoom1 should be more thoroughly discussed. Specially, the manuscript should explain why BDALTI DEM fails to capture meandering river network, while MERIT DEM appears to better represent them (Figure 2). The origin of these differences between DEMs after sub-grid processing should be deeper analyzed.***

AR: We thank the reviewer for this relevant comment.

We would like, however, to first clarify that the internal omissions discussed in the manuscript are observed exclusively for MERIT-derived sub-grid networks, rather than for BDALTI-derived networks. As shown in Figure 6c, internal omissions are essentially absent for BDALTI ($<0.1\%$), while they represent approximately 1% for MERIT. This pattern is also illustrated in Zoom 1 of Figure 5d, where the MERIT-derived sub-grid network (blue) locally fails to capture sharp meanders of the reference network (red), while the corresponding Zoom 1 of Figure 5b shows that the BDALTI-derived sub-grid network (blue) preserves them more accurately. These internal omissions are directly linked to the native DEM resolution, with MERIT having a relatively coarser native resolution (resampled at 100 m) compared to BDALTI (25 m). This analysis therefore highlights the sensitivity of the quality of derived sub-grid networks to the native DEM resolution.

That said, as demonstrated in the manuscript, overall omissions (internal + headwater) remain very low across scales for the two DEMs (up to 6%), confirming the spatial accuracy consistency of sub-grid river networks and the preservation of the reference network topology. Given their limited extent, these omissions are expected to have minimal influence on basin-scale H&H simulations.

In the revised manuscript, we will clarify these explanations more explicitly in Sections 5.1 and 6.1, improve the caption of Figure 5 in relation to omissions, and improve the readability of Figure 6c, as mentioned in our previous response.

2.8.2 Mitigating catchment size errors

RC: ***Lines 495–497: add the figure reference to support the sentence. Lines 504–505: The authors should further discuss why MERIT DEM shows larger errors at coarse resolution than BDALTI. Given that MERIT DEM is generally expected to perform better in hydrological applications, this result is counterintuitive. A likely explanation is that the original MERIT DEM (~90 m, 3 arc-second) was resampled to 100 m using bilinear resampling, which may introduce additional errors in flow routing and catchment delineation. In contrast, BDALTI has a native resolution of 25 m, which may explain its superior performance in this study. Therefore, the observed differences may not reflect intrinsic DEM quality but rather preprocessing effects. Would using preprocessed DEM products (e.g., MERIT Hydro IHU) reduce this bias? Given the relatively small study area, the conclusion that BDALTI outperforms MERIT DEM may not be applicable to larger basins.***

AR: We thank the reviewer for these relevant comments.

Regarding Lines 495–497, we will add a reference to Figure 7 in the revised manuscript to support this sentence.

Regarding the larger errors obtained with MERIT DEM at coarse resolutions for the grid-based area method, we agree that this result deserves further discussion.

As outlined in the manuscript (Lines 504–505), these differences reflect the inherent sensitivity of the grid-based delineation method to the native DEM resolution: a finer native resolution yields more accurate flow directions, which in turn improves the quality of flow-direction upscaling and resulting catchment delineation at coarser grid resolutions. Since BDALTI has a native resolution of 25 m, it is expected to produce more accurate grid-based catchment delineation than MERIT DEM, which was resampled from its native ~90 m to a regular 100 m grid prior to flow-direction computation.

As already discussed in our response above, this resampling step was necessary to ensure a consistent multiplicity between native and target resolutions during flow-direction upscaling, but may locally alter elevation gradients and flow-direction accuracy, potentially amplifying grid-based catchment area errors at coarser resolutions (5–10 km). We will clarify this interpretation in the revised discussion, emphasizing that the observed differences between BDALTI and MERIT at coarse resolutions may reflect the combined effect of the coarser effective resolution of MERIT and its resampling to a regular 100 m grid, rather than the intrinsic quality of MERIT DEM alone.

However, the main objective of this section is to demonstrate the benefit of the sub-grid area method over the grid-based area method. For both DEMs, the grid-based area method shows increasing catchment area errors as resolution is coarsened, whereas the sub-grid area method maintains consistent accuracy across scales. This supports one of the key findings of this study: propagating fine-resolution drainage information into the hydrological mesh effectively mitigates the catchment size problem across spatial resolutions.

Regarding the potential use of preprocessed products such as MERIT Hydro IHU, as explained in our response above, we deliberately recomputed the full preprocessing chain from the input DEMs to maintain methodological control and generality across DEM sources, target resolutions, and hydrographic networks. The integration of pre-upscaled products such as MERIT Hydro IHU as alternative direct inputs remains a relevant perspective and will be mentioned in the revised manuscript.

2.8.3 Understanding multi-scale performance patterns

RC: ***Figures D1 and 8 do not convincingly demonstrate improved discharge simulation performance of the HH framework. Across resolutions (1 km, 5 km, 10 km), the distributions of KGE and Pearson correlation (r) are nearly identical, with similar medians. Differences in bias are minor and do not clearly indicate***

superior performance of the sub-grid method. The only noticeable improvement appears in KGE, while the overall spread of boxplots remains similar, with differences largely driven by the removal of extreme values. Figures D2 and D3: Would classify results by sub-basin and DEM improve interpretation (Figures D2, D3)? Current plots do not sufficiently show these distinctions.

AR: We thank the reviewer for these relevant comments.

We would first like to highlight that Figures 8 and D1 are not intended to demonstrate an improvement in predictive discharge performance of the sub-grid area method over the grid-based area method. As stated in our responses above, the objective of the numerical experiments is not to optimize or benchmark hydrological predictive performance, but to evaluate the methodological robustness and internal consistency of the proposed H&H framework across spatial resolutions.

The preservation of flow timing across scales for both delineation methods, demonstrated through the stable Pearson correlation r, further supports the spatial accuracy of the preprocessing workflow (as discussed in Sections 6.1 and 6.2), since both drainage-area methods use consistent sub-grid river network, hydraulic geometry, and uniform Manning coefficient.

The use of KGE$'$ over the standard KGE was motivated by its ability to decouple bias from variability. The differences between the grid-based and sub-grid area methods are primarily expressed through the bias component β, with the sub-grid area method showing more stable and bounded bias distributions across scales (Figure 8d,h). These differences do not indicate that one area method outperforms the other, but rather reflect distinct physical patterns across spatial scales. As discussed in the manuscript, the grid-based area method tends to yield bias values closer to the ideal value of 1, since drainage area overestimation inflates precipitation volumes and compensates for precipitation aggregation and hydrological parameter smoothing during spatial upscaling (i.e., no recalibration was performed). Conversely, the sub-grid area method constrains these precipitation volume biases by preserving more accurate drainage areas across scales (as demonstrated in Figure 7), explaining the lower but more stable median β values across resolutions.

Beyond the scope of the present study, the sub-grid area method mitigates the compensation effects inherent to regular grid discretization, where drainage area overestimation is typically compensated by less physically meaningful model parameters through calibration. By ensuring physically consistent drainage area representation at coarse scales, the sub-grid area method thereby provides a more physically consistent basis for future model optimization and data assimilation perspectives.

Therefore, the results in Section 6.3 should be interpreted solely in terms of physical consistency and stability across spatial resolutions, and not as a benchmark of predictive performance. We acknowledge that this distinction is not sufficiently explicit in the current manuscript and will clarify this interpretation in the revised manuscript to avoid any ambiguity.

Regarding Figures D2 and D3, these figures provide the spatial distribution of KGE$'$ performance across stations and are already separated by DEM: Figure D2 corresponds to BDALTI and Figure D3 to MERIT. The panels also distinguish spatial resolution (columns: 1, 5, and 10 km) and drainage area delineation method (rows: grid-based and sub-grid). We will improve the caption and layout of these figures in the revised manuscript to make these distinctions clearer.

2.8.4 Conclusion

RC: ***The manuscript emphasizes a "unified raster-vector H&H framework" with improved physical consistency and scalability. However, the presented simulation of discharge results mainly demonstrates stable correlation (r) and redistribution of bias, without convincingly showing a significant improvement in simulation performance compared to existing methods. Moreover, the reported advancement in catchment delineation appears limited given the limited basin size, lack of testing in complex river systems, and***

inconsistencies in DEM resolution and preprocessing.

AR: We thank the reviewer for this comment. As clarified in detail in our responses to the main comments and the corresponding detailed comments above, the scope of this study is not to demonstrate improved predictive performance or to benchmark the framework against existing methods. We therefore respectfully disagree with the characterization of the discharge results as insufficient, as this assessment does not align with the stated objectives of the manuscript.

We also respectfully disagree with the statement that the reported advances in catchment delineation are limited by the study basin or by the DEM preprocessing choices. As discussed in our responses above, the Garonne basin constitutes a significant and hydrologically contrasted test case for evaluating the proposed methodology. The results show consistent geometric performance of the sub-grid river networks across scales, and a clear reduction of catchment-size errors when using the sub-grid area method instead of the grid-based area method. These results directly support the methodological contribution of the preprocessing workflow.

Regarding the mentioned inconsistencies in DEM resolution and preprocessing, the methodological choices related to MERIT resampling and river network filtering were deliberate and are explained in the manuscript, as well as in our responses to the corresponding comments above. We acknowledge, however, that their implications for the interpretation of the geometric and H&H results should be stated more explicitly. These points will therefore be clarified in the revised manuscript, together with the associated limitations. However, these preprocessing choices do not undermine the main conclusion of the study, which concerns the internal methodological consistency and scalability of the proposed raster–vector H&H framework across spatial resolutions.

References

Berkaoui, M. A., Saadi, M., Colleoni, F., Huynh, N. N. T., Akhtari, A., Larnier, K., Roux, H., and Garambois, P.-A.: A Raster–Vector Framework for Multi-Scale Hydrological–Hydraulic Modeling Across Large Domains, Zenodo [code and data], , 2026.

Colleoni, F., Huynh, N. N. T., Garambois, P.-A., Jay-Allemand, M., Organde, D., Renard, B., De Fournas, T., El Baz, A., Demargne, J., and Javelle, P.: SMASH v1.0: a differentiable and regionalizable high-resolution hydrological modeling and data assimilation framework, Geoscientific Model Development, 18, 7003–7034, , 2025.

Davies, H. N. and Bell, V. A.: Assessment of methods for extracting low-resolution river networks from high-resolution digital data, Hydrological Sciences Journal, 54, 17–28, , 2009.

Eilander, D.: pyFlwDir, , 2022.

Eilander, D., Winsemius, H. C., Van Verseveld, W., Yamazaki, D., Weerts, A., and Ward, P. J.: MERIT Hydro IHU, , 2020.

Eilander, D., van Verseveld, W., Yamazaki, D., Weerts, A., Winsemius, H. C., and Ward, P. J.: A hydrography upscaling method for scale-invariant parametrization of distributed hydrological models, Hydrology and Earth System Sciences, 25, 5287–5313, , 2021.

McMaster, K. J.: Effects of digital elevation model resolution on derived stream network positions, Water Resources Research, 38, 13–1–13–8, , 2002.

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., and Loumagne, C.: Which potential evapotranspiration input for a lumped rainfall–runoff model?: Part 2—Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling, Journal of Hydrology, 303, 290–306, , 2005.

Quintana-Seguí, P., Moigne, P. L., Durand, Y., Martin, E., Habets, F., Baillon, M., Canellas, C., Franchisteguy, L., and Morel, S.: Analysis of Near-Surface Atmospheric Variables: Validation of the SAFRAN Analysis over France, Journal of Applied Meteorology and Climatology, 47, 92–107, , 2008.

Sousa, T. M. I. and Paz, A. R.: How to evaluate the quality of coarse-resolution DEM-derived drainage networks, Hydrological Processes, 31, 3379–3395, , 2017.

Vidal, J.-P., Martin, E., Franchistéguy, L., Baillon, M., and Soubeyroux, J.-M.: A 50-year high-resolution atmospheric reanalysis over France with the Safran system, International Journal of Climatology, 30, 1627–1644, , 2010.