Answer to reviewer 01

General response

We thank the reviewer for the thorough and constructive evaluation of our manuscript. We appreciate the positive assessment of the study's relevance and the many suggestions provided to improve its clarity, robustness, and scientific interpretation. We have carefully considered all comments and will substantially revise the manuscript accordingly. The revised manuscript will include five major categories of improvements:

1. **Clearer articulation of the study motivation, novelty, and objectives.** We will substantially revise the Introduction to focus more directly on runoff forcing uncertainty, streamflow simulation, and trend evaluation. The revised text will explicitly highlight the novelty of comparing ERA5-Land and ISIMIP3a runoff within a common CaMa-Flood framework and clarify the scientific questions addressed by the study.
2. **More cautious interpretation of the results and clearer discussion of limitations.** Following the reviewer's recommendations, we will revise the Discussion and Conclusions to avoid overly broad recommendations and causal interpretations. Throughout the manuscript, we will emphasize that the analysis compares two complete runoff-product configurations rather than formally attributing differences to individual sources of uncertainty. We will also clarify the shared ERA5 heritage of ERA5-Land and ISIMIP3a and discuss differences in anthropogenic-process representation.
3. **Additional robustness and sensitivity analyses.** Several new analyses will be included to assess the robustness of the conclusions. These include: (i) evaluation using additional performance metrics (NSE, logNSE, and RMSE); (ii) sensitivity of trend results to stations with small reference flow values; (iii) assessment of anthropogenic influences using additional ISIMIP3a experiments; (iv) evaluation of the impact of the CaMa-Flood reservoir module; (v) complementary Mann–Kendall trend significance analyses; and (vi) additional analyses of meteorological drivers to support interpretation of differences between ERA5-Land and ISIMIP3a.
4. **Stronger integration of daily streamflow performance and trend evaluation.** To better connect the two major components of the study, we will introduce new analyses examining the relationship between daily streamflow skill and trend errors. These additions will help establish a clearer link between short-term hydrological performance and long-term trend reproduction and provide a more coherent interpretation of the results.
5. **Improved presentation and methodological transparency.** We will expand the description of station filtering and allocation procedures, report the number of stations retained after major processing steps, clarify the use of river geometry parameters and regional aggregation frameworks, remove duplicated text, and add new supplementary figures illustrating station-level variability and the sensitivity of conclusions to alternative regional classifications.

Overall, these revisions will substantially strengthen the manuscript by improving transparency, clarifying the scope and limitations of the analysis, and providing additional evidence supporting the principal conclusions.

Reviewer comment 01: *The manuscript presents a comparison between era5land and ISIMIP3a runoff to address an important research gap, but the novelty is not yet convincingly established. Many previous studies have already evaluated global runoff products, streamflow simulations, hydrological model ensembles,* *reanalysis products, and routing uncertainty. The authors need to explain more explicitly what is new in this study and how it advances beyond existing benchmark comparisons.*

Response 01: We thank the reviewer for this comment and agree that the novelty of the study was not sufficiently emphasized in the original manuscript. While previous studies have evaluated global runoff products, hydrological models, reanalysis datasets, and routing schemes, our study addresses a different question: how two contrasting runoff-production workflows translate into river discharge when evaluated within a common routing framework, and whether their relative performance depends on the aspect of streamflow being considered. This comparison is particularly relevant because ERA5-Land and ISIMIP3a represent two contrasting yet widely used approaches to runoff generation: a high-resolution, reanalysis-based product and a bias-adjusted, multi-model hydrological ensemble.

Planned revision 01: We will revise the Introduction and Discussion sections to more clearly articulate the study's contribution. Specifically, we will highlight three key advances: (i) a direct comparison of ERA5-Land and ISIMIP3a runoff using an identical CaMa-Flood configuration, thereby minimizing routing-related differences and focusing on runoff-forcing uncertainty; (ii) a global-scale evaluation against more than 5,000 gauging stations spanning diverse hydroclimatic regions; and (iii) a joint assessment of both historical streamflow performance and long-term trends in low, mean, and high streamflow, complemented by targeted diagnostics of meteorological forcing, pre-routing runoff, and routed discharge to better interpret where the observed differences emerge. These revisions will clarify how the study extends beyond previous benchmark evaluations and provides practical guidance on the strength and limitations of two widely used runoff products for hydrological and climate-impact assessments.

Reviewer comment 02: *Currently, the paper mostly comes across as a benchmark comparison. However, it would be helpful to clearly explain the main goal of the regional performance assessment. While the authors share regional KGE maps, they don't specify the scientific question this analysis aims to address. maybe clarify whether this part is intended to evaluate the spatial consistency of Ear5land's benefits, identify regions where both runoff products may not perform well, or explore regional relationships among ISIMIP3a models may provide some help.*

Response 02: We thank the reviewer for this helpful comment and agree that the purpose of the regional performance assessment was not sufficiently articulated in the original manuscript. The regional analysis was not intended solely as a descriptive presentation of *KGE* patterns, but rather to assess whether differences between ERA5-Land and ISIMIP3a were spatially consistent across contrasting hydroclimatic settings. While global summary statistics provide an overall assessment, they can obscure important regional variations in runoff generation processes, forcing quality, and model behavior. The regional evaluation therefore helps identify where differences between runoff products are robust and where uncertainty remains particularly large. In addition, performance was summarized using the IPCC SREX regions for two reasons: i) to facilitate comparison with previous large-scale assessments of observed and simulated streamflow trends and ii) to present the results within a framework commonly used in climate-impact and policy-relevant assessments.

Planned revision 02: We will revise both the Materials & Methods and Discussion sections to explicitly state the scientific objectives of the regional assessment. In particular, we will clarify that the analysis is intended to evaluate the spatial consistency of ERA5-Land and ISIMIP3a performance, identify regions where both runoff products exhibit persistent limitations, and assess how forcing-related uncertainties vary across hydroclimatic regimes.

Reviewer comment 03: *The experiment does not isolate the effect of forcing dataset, hydrological model structure, bias correction, calibration, human impacts, or spatial resolution. Yet the Discussion, especially Section 4.2, frequently suggests causal explanations. The authors should avoid causal overinterpretation unless additional experiments are added. At minimum, they should clearly state that the comparison is between two runoff-product configurations rather than a controlled attribution of individual uncertainty sources.*

Response 03: We thank the reviewer for this important observation and agree that parts of the original Discussion implied a stronger level of causal attribution than is supported by the experimental design. The objective of this study is to compare two complete runoff-product configurations rather than to isolate individual sources of uncertainty. ERA5-Land and ISIMIP3a differ simultaneously in several aspects, including meteorological forcing, bias-correction procedures, hydrological model structure, ensemble representation, calibration strategy, and spatial resolution. Consequently, performance differences cannot be uniquely attributed to any single factor.

Planned revision 03: We will revise the beginning of Section 4.1 to explicitly acknowledge these limitations and clarify the scope of the analysis. Throughout the Discussion, we will avoid causal language and instead present proposed mechanisms as plausible interpretations consistent with the known characteristics of the datasets. We will also add an explicit limitation statement emphasizing that the results reflect the combined influence of multiple interacting factors and should not be interpreted as a formal attribution of individual uncertainty sources.

Reviewer comment 04: *The authors should explain why the SREX regional scale is appropriate for evaluating model performance. SREX regions are broad climate-impact regions rather than hydrological units, and aggregation at this scale may mask substantial basin-level variability. The authors should justify this choice and discuss whether using major river basins, hydroclimatic zones, or aridity-based regions would affect the conclusion.*

Response 04: We thank the reviewer for this valuable comment. We agree that SREX regions are not hydrological units and that aggregation at any broad spatial scale inevitably masks some basin-level variability. However, the objective of the regional analysis was not to characterize performance within individual river basins, but rather to assess whether differences between ERA5-Land and ISIMIP3a are spatially consistent across large-scale climatic and environmental settings.

We selected SREX regions because they are widely used in climate-change impact assessments, facilitate comparison with previous global hydrological studies, and provide sufficient station density for robust estimation of regional statistics and uncertainty bounds. Nevertheless, we agree that the choice of regionalization framework could potentially influence the interpretation of results.

Planned revision 04: To assess the sensitivity of our conclusions to the selected aggregation scheme, we will conduct an additional analysis using Köppen–Geiger climate regions. We will add a new supplementary figure summarizing *KGE* and its components across Köppen–Geiger climate regions. We will also revise Section 2.2.3 and the Discussion to justify the use of SREX regions, acknowledge their limitations, and demonstrate that the principal findings are not fundamentally altered when an alternative climate-based regionalization framework is adopted.

Reviewer comment 05: *ERA5-Land runoff does not include the same human impact representation as ISIMIP3a. Therefore, it is not clear whether the two products are fully comparable in regions strongly affected by reservoirs, water abstraction, irrigation, or other human interventions. The authors should explain how human impacts are represented in each product and whether this difference may bias the comparison.*

Response 05: We thank the reviewer for this important comment. We agree that differences in the representation of anthropogenic influences are a potential source of uncertainty when comparing ERA5-Land and ISIMIP3a runoff products.

Human influences are represented at different stages of the modelling framework. In the present study, all runoff products were routed through the same CaMa-Flood configuration, including the reservoir operation module, thereby providing a consistent representation of major flow-regulation processes across all simulations. Consequently, differences associated with reservoir routing are largely controlled for in the comparison.

However, the runoff products themselves differ in their treatment of anthropogenic influences prior to routing. The ISIMIP3a simulations used in this study correspond to the "obsclim + histsoc" experiment, which includes time-varying representations of water abstraction and water management. The specific implementation of these processes varies among the individual hydrological models that comprise the ISIMIP3a ensemble. In contrast, ERA5-Land runoff does not explicitly represent many of these anthropogenic processes. As a result, differences in human-impact representation may contribute to some of the observed performance differences, particularly in heavily managed basins.

Planned revision 05: We will explicitly discuss this limitation in the manuscript and clarify that the comparison reflects the combined influence of meteorological forcing, model structure, and anthropogenic-process representation. In addition, we will perform an additional analysis using the ISIMIP3a "obsclim + 1901soc" simulations, in which anthropogenic water management is fixed at pre-development conditions. Comparing these simulations with the "obsclim + histsoc" experiment will allow us to quantify the influence of human water management on streamflow performance globally and across SREX regions.

Reviewer comment 06: *Related to the previous point, the manuscript states that the Cama-Flood dam module is enabled. However, ISIMIP3a simulations may already include some human influences depending on the hydrological model and experimental setup. The authors should clarify whether applying the Cama-Flood dam module to ISIMIP3a runoff could create inconsistencies or double count some forms of regulation. A sensitivity experiment without the dam module would strengthen the study.*

Response 06: We thank the reviewer for this helpful suggestion. We agree that the interaction between anthropogenic influences represented within the runoff products and reservoir regulation represented by the CaMa-Flood dam module should be clarified.

We note that the dam module operates during the routing stage and is applied consistently to all runoff products. Nevertheless, because some ISIMIP3a hydrological models include representations of water management and abstraction processes, there is potential for overlap between processes represented during runoff generation and those represented during routing. To assess the importance of this issue, we conducted preliminary analyses comparing simulations with and without the CaMa-Flood dam module. These analyses showed that differences in station-level *KGE* are generally centered near zero across SREX regions and runoff products. Although reservoir regulation can substantially affect individual stations, particularly in heavily managed river systems, the overall comparison between ERA5-Land and ISIMIP3a is largely unchanged.

Planned revision 06: We will update the method section to make clear that at what stages streamflow regulation are consider in ISMIP models, clarifying that there is no double count some forms of regulation. In addition, we will include a dedicated sensitivity analysis comparing simulations performed with and without the CaMa-Flood dam module. The results will be presented as an additional supplementary figure and summarized in the manuscript. This analysis will demonstrate the extent to which the study conclusions depend on the representation of reservoir regulation and will provide evidence that the principal findings are robust to the inclusion or exclusion of the dam module.

Reviewer comment 07: *The Discussion needs some significant rewriting. Right now, it repeats the Results and includes some speculative interpretations. It would be helpful for the authors to concentrate the Discussion on the main scientific question. Why does Era5land seem to do better for daily streamflow changes, while ISIMIP3a appears to be more conservative or more consistent when looking at long-term trends?*

Response 07: We thank the reviewer for this insightful comment and agree that the original Discussion devoted too much space to summarizing the results and did not sufficiently focus on the broader scientific implications of the findings. We also agree that the contrasting behavior of ERA5-Land and ISIMIP3a with respect to daily streamflow performance and long-term trend representation deserves a more focused discussion.

While the experimental design does not permit formal attribution of the observed differences to individual factors, we agree that additional analyses can provide useful insights into potential mechanisms. In particular, we will investigate differences in key hydroclimatic drivers, including precipitation and temperature, between ERA5-Land and ISIMIP3a. These analyses will help contextualize the observed streamflow differences and assess whether contrasting representations of meteorological forcing may contribute to the stronger daily streamflow performance of ERA5-Land and the more conservative trend behavior of ISIMIP3a.

Planned revision 07: We will substantially restructure the Discussion to focus on the central scientific question addressed by the study: the trade-offs associated with the contrasting characteristics of the two runoff-production frameworks and their different performance in reproducing historical streamflow dynamics and long-term trends. We will reduce repetition of the Results section, introduce additional analyses of precipitation and temperature characteristics, and use these results to provide a more evidence-based interpretation of the observed differences between ERA5-Land and ISIMIP3a. Throughout the Discussion, interpretations will be framed as plausible explanations rather than causal attributions, consistent with the limitations of the experimental design.

Reviewer comment 08: *The Cama-Flood description is duplicated. The text describing CaMa-Flood appears in Section 2.1.2 and then again at the start of Section 2.2.*

Response 08: We thank the reviewer for identifying this issue.

Planned revision 08: We will remove the duplicated description appearing at the beginning of Section 2.2 and retain a single, consolidated description of the CaMa-Flood model in Section 2.1.2.

Reviewer comment 09: *The Introduction spends too much space on global flood exposure, climate pledges, and future flood risk. Should be shortened and refocused on runoff forcing uncertainty, streamflow simulation, and trend evaluation.*

Response 09: We thank the reviewer for this helpful comment and agree that the Introduction can be more tightly focused on the scientific motivation of the study. In the original manuscript, the broader context of flood risk and climate adaptation occupied a disproportionate amount of space relative to the central research question concerning runoff forcing uncertainty and streamflow simulation performance.

Planned revision 09: We will substantially revise the Introduction to streamline the broader flood-risk background and place greater emphasis on runoff forcing uncertainty, global streamflow simulation, and the challenges associated with reproducing both historical discharge dynamics and long-term streamflow trends. We will also revise the latter part of the Introduction to more clearly establish the research gap, articulate the novelty of the study, and explicitly state the scientific objectives and hypotheses being tested.

Reviewer comment 10: *The authors have done an interesting job estimating TS slopes for annual Q10, mean flow, and Q90, and then comparing regional medians. To make their trend analysis even more robust and trustworthy, they might consider incorporating MK tests at the station level. Discussing field significance or multiple testing, and checking whether the modeled and observed regional trends differ significantly, could also add valuable insights. The suggestion that ISIMIP3a offers more observation-aligned trend estimates is promising, and with a bit more statistical backing, it can be even more convincing.*

Response 10: We thank the reviewer for this constructive suggestion. We agree that additional statistical diagnostics can help strengthen the interpretation of the trend analysis. In the original manuscript, uncertainty in regional trend estimates was quantified using a bootstrap framework. However, we acknowledge that complementary station-level significance information will provide additional context regarding the robustness of the identified trends.

Planned revision 10: We will complement the existing Sen's slope analysis with Mann–Kendall trend tests at the station level. Specifically, we will summarize the proportion of stations exhibiting significant positive, significant negative, and non-significant trends for Q10, mean flow, and Q90 in both observations and simulations. This additional analysis will provide an independent assessment of trend robustness while remaining consistent with the non-parametric framework already adopted in the study. We will also discuss the relationship between the Mann–Kendall and Sen's slope results and demonstrate that the principal conclusions regarding the relative performance of ERA5-Land and ISIMIP3a remain unchanged.

Reviewer comment 11: *KGE is useful, but KGE alone cannot fully characterize hydrological performance. Possible additions include NSE, logNSE, flow duration curve bias, or others. This is very important because the manuscript is motivated by flood modeling, but the current evaluation does not directly assess flood peaks or floodplain inundation.*

Response 11: We thank the reviewer for this valuable comment. We agree that no single performance metric can fully characterize hydrological model skill and that complementary metrics can provide additional insights into model behavior across different parts of the flow regime. While *KGE* was selected as the primary metric because it simultaneously evaluates correlation, variability, and bias, we have already computed several additional performance measures, including *NSE* and *logNSE*.

We also agree that flood modeling provides an important motivation for evaluating runoff products. However, the primary objective of the present study is to assess the ability of different runoff products to reproduce observed streamflow dynamics and long-term trends. Streamflow constitutes the principal input to large-scale hydrodynamic flood models such as CaMa-Flood, making discharge performance a necessary first step in evaluating the suitability of runoff products for flood-related applications. We acknowledge that the present study does not directly evaluate flood peaks, inundation extent, or floodplain dynamics.

Planned revision 11: We will include additional supplementary figures presenting *NSE* and *logNSE* across simulations and regions. These analyses will be briefly discussed to demonstrate that the principal conclusions are robust across multiple performance metrics and are not dependent on the choice of *KGE* alone. We will also clarify in the Introduction and Discussion that the scope of this study is limited to streamflow evaluation, and that direct assessment of flood peaks and inundation performance represents an important avenue for future work.

Reviewer comment 12: *The manuscript presents trends as percent changes per decade relative to the long term mean for each flow metric. This can produce very large or unstable values when mean low flow is small, especially in arid and semi-arid regions.*

Response 12: We thank the reviewer for highlighting this important issue. We agree that expressing trends as percentage changes relative to the long-term mean can lead to inflated or unstable values when the reference flow magnitude is very small. This concern is particularly relevant for low-flow metrics in arid and semi-arid regions, where even small absolute changes may correspond to large relative changes.

Planned revision 12: To assess the sensitivity of our results to this issue, we will perform an additional robustness analysis excluding stations with very small reference flow values. We will then repeat the regional trend analysis and compare the resulting distributions with those presented in the original manuscript. Preliminary analyses indicate that the principal conclusions, including the relative performance of ERA5-Land and ISIMIP3a in reproducing observed trends, are not materially affected by the exclusion of these stations. We will document the filtering criterion and provide the corresponding sensitivity analysis in the Supplementary Material.

Reviewer comment 13: *The conclusion that spatial resolution has limited influence is too broad. The experiment only shows that, within the present Cama-Flood configuration and for the selected streamflow metrics, upscaling era5land runoff from 0.1° to 0.5° has little effect.*

Response 13: We thank the reviewer for this important observation and agree that the original wording was overly broad. The experiment does not provide a general assessment of the influence of runoff spatial resolution on streamflow simulations. Rather, it evaluates the sensitivity of simulated discharge to the aggregation of ERA5-Land runoff from 0.1° to 0.5° within a specific modelling framework.

In particular, all simulations were performed using the same CaMa-Flood configuration, which operates on a 0.1° river network and internally redistributes runoff to the routing grid. Consequently, the analysis does not isolate the full impact of spatial resolution across the entire modelling chain. Furthermore, the experiment only considers a relatively modest reduction in runoff resolution (0.1° to 0.5°), and the results cannot be assumed to apply to substantially coarser runoff products.

Planned revision 13: We will revise the relevant sections of the Results, Discussion, and Conclusions to clarify that the observed sensitivity is specific to the tested resolutions, evaluated streamflow metrics, and CaMa-Flood configuration. We will replace broad statements regarding the influence of spatial resolution with more precise language indicating that the aggregation of ERA5-Land runoff from 0.1° to 0.5° had a limited effect on simulated streamflow performance within the modelling framework considered in this study.

Reviewer comment 14: *The regional median KGE values don't fully capture how well the model performs in each region. Keep in mind that within each SREX region, the performance at individual stations can differ quite a bit, highlighting the importance of looking at more detailed data.*

Response 14: We thank the reviewer for this valuable comment and agree that regional median values provide only a partial summary of model performance. While medians are useful for identifying broad regional patterns and reducing the influence of outliers, they inevitably mask the substantial station-to-station variability that exists within each SREX region.

Planned revision 14: To better characterize within-region variability, we will add supplementary figures showing the distribution of station-level performance metrics within each region. Specifically, we will present boxplots of KGE and its individual components (correlation, variability ratio, and bias ratio) across SREX regions. These figures will complement the regional median summaries by explicitly illustrating the spread and heterogeneity of station-level performance. We will also provide analogous results for the Köppen–Geiger climate regions introduced in response to Reviewer Comment 04. Together, these additional analyses will demonstrate that the manuscript conclusions are based not only on regional median values but also on the broader distribution of station-level performance.

Reviewer comment 15: *The interpretation of the uncertainty overlap in Table 1 might benefit from a little bit of clarification. Keep in mind that a model with broader uncertainty intervals tends to overlap with observations more frequently, but this doesn't automatically mean it performs better.*

Response 15: We thank the reviewer for this important observation and agree that uncertainty overlap should not be interpreted in isolation. A model with broader uncertainty intervals may indeed exhibit higher overlap with observations simply because a larger range of values is encompassed, rather than because the underlying trend estimates are more accurate.

Planned revision 15: To address this concern, we will expand Table 1 to also report the mean width of the uncertainty intervals. This additional information allows the overlap statistics to be interpreted alongside the magnitude of the associated uncertainty. In the revised Discussion, we will explicitly note that higher overlap does not necessarily imply better performance when accompanied by substantially wider intervals. However, in the present analysis, the uncertainty intervals of ERA5-Land and ISIMIP3a are of comparable width, while ISIMIP3a exhibits substantially higher overlap with observations. This suggests that the improved overlap cannot be explained solely by broader uncertainty bounds and instead reflects closer agreement with the observed trend estimates.

Reviewer comment 16: *The station filtering and allocation procedure needs more detail.*

Response 16: We thank the reviewer for this comment. We agree that several aspects of the station filtering and allocation procedure were not described in sufficient detail in the original manuscript. In particular, the definition of a "valid year" was not explicitly provided, and the procedure used to identify and evaluate candidate model grid cells for station allocation was described too briefly.

Planned revision 16: We will expand the methodological description of the station filtering and allocation procedure and provide additional quantitative information regarding the number of stations retained after each filtering step. Specifically, we will clarify that: "A valid year was defined as a calendar year containing at least 355 daily discharge observations. Stations with fewer than five valid years of data during 1979–2019 (not necessarily consecutive), or with a reported drainage area smaller than 360 km², were excluded."

We will also provide a more detailed description of the station allocation procedure: "For the remaining 5,099 eligible stations, an initial candidate grid cell was identified from the reported station coordinates. Neighboring grid cells within a predefined search window were then evaluated to identify the optimal allocation based on drainage-area consistency. The search radius was adapted according to the reported station drainage area (Table S2), allowing larger rivers to be matched across a broader spatial window."

In addition, we will report the number of stations remaining after each major filtering stage and revise the text to make the allocation workflow fully reproducible.

Reviewer comment 17: *The manuscript should better connect the daily streamflow performance analysis with the long-term trend analysis. For example, do regions with high KGE also show better trend agreement? Do regions with low KGE show larger trend errors? The current manuscript treats these two analyses mostly separately, but their relationship is central to the paper's interpretation.*

Response 17: We thank the reviewer for this insightful comment and agree that the relationship between daily streamflow performance and long-term trend reproduction deserves greater attention. In the original manuscript, these analyses were presented largely independently, which may have obscured potentially important connections between short-term hydrological skill and long-term trend fidelity.

To address this point, we conducted an additional analysis relating station-level *KGE* values to errors in simulated streamflow trends. The results indicate that higher daily streamflow skill tends to be associated with smaller trend errors.

Planned revision 17: We will add a new figure examining the relationship between *KGE* and trend error across stations and runoff products. We will also expand the Discussion to explicitly connect the daily-performance and trend-analysis sections. The revised manuscript will highlight that daily streamflow skill and trend reproduction evaluate different aspects of model performance and will incorporate the new station-level analysis to more explicitly examine their relationship.

Reviewer comment 18: *The manuscript should be more cautious when stating that Era5land is well suited for reproducing historical streamflow dynamics and highlights that ISIMIP3a is especially valuable for long-term trend assessment. These are reasonable hypotheses, but the current evidence isn't yet strong enough to endorse such broad recommendations.*

Response 18: We thank the reviewer for this important comment and agree that some statements in the original manuscript were phrased too strongly relative to the evidence presented. The objective of the study is to compare the performance of ERA5-Land and ISIMIP3a with respect to historical streamflow dynamics and long-term streamflow trends, rather than to provide definitive recommendations for all potential applications.

Our results indicate that ERA5-Land generally exhibits higher skill for reproducing observed daily streamflow variability, while ISIMIP3a often shows closer agreement with observed trend characteristics. However, we agree that these findings should be interpreted within the context of the specific datasets, modelling framework, evaluation metrics, and analysis period considered in this study. They do not establish universal superiority of either product for a given application.

Planned revision 18: We will revise the Discussion and Conclusions to adopt more cautious language when interpreting the comparative strengths of ERA5-Land and ISIMIP3a. In particular, we will replace broad recommendations with statements that explicitly acknowledge the scope and limitations of the analysis. We will emphasize that the results suggest potential trade-offs between reproducing historical discharge dynamics and representing long-term streamflow trends within the present modelling framework, rather than establishing general guidance applicable to all hydrological or climate-impact studies.

Reviewer comment 19: *The authors should also address the non-independence between ear5land and ISIMIP3a since ISIMIP3a W5E5 partly uses ERA5 data and applies bias adjustments based on observational products, and era5land is similarly driven by ERA5 atmospheric forcing, this connection needs acknowledgment.*

Response 19: We thank the reviewer for highlighting this important point. We agree that ERA5-Land and ISIMIP3a are not fully independent products. Both datasets ultimately derive from a common meteorological foundation through the use of ERA5 atmospheric data. In the case of ISIMIP3a, the W5E5 forcing dataset combines ERA5 information with bias adjustments based on observational products before being used to drive multiple global hydrological models. ERA5-Land, in contrast, is generated directly from the ERA5 reanalysis through a land-surface modelling framework.

Consequently, the comparison presented in this study should not be interpreted as a comparison between fully independent runoff products. Rather, it represents a comparison between two contrasting approaches for generating runoff from a shared meteorological foundation: a reanalysis-based land-surface modelling framework and a bias-adjusted multi-model hydrological ensemble. Despite the common forcing ancestry, substantial differences remain in the treatment of meteorological inputs, runoff generation processes, model structure, calibration strategy, and anthropogenic influences.

Planned revision 19: We will revise Section 2.1 and the Discussion to explicitly acknowledge the shared ERA5 heritage of ERA5-Land and W5E5/ISIMIP3a. We will also clarify that the study evaluates differences arising from the subsequent processing and modelling chains rather than comparing two fully independent meteorological datasets. This limitation will be considered when interpreting the magnitude of differences between ERA5-Land and ISIMIP3a throughout the manuscript.

Reviewer comment 20: *Additionally, the paper should avoid suggesting that the results directly apply to flood hazard modeling unless flood-specific metrics are incorporated. Currently, the analysis focuses only on streamflow skills and trends.*

Response 20: We thank the reviewer for this important comment and agree that the scope of the present study is limited to the evaluation of streamflow simulations. While flood hazard modelling represents an important application of runoff products and hydrodynamic routing models, the analyses presented here focus exclusively on discharge performance and streamflow trends. Consequently, the results should not be interpreted as a direct evaluation of flood hazard modelling skill.

At the same time, streamflow remains a fundamental input to large-scale flood models such as CaMa-Flood. Assessing the ability of runoff products to reproduce observed discharge dynamics and long-term streamflow trends therefore provides important information regarding their suitability for flood-related applications, even though flood hazards themselves are not evaluated.

Planned revision 20: We will revise the Introduction, Discussion, and Conclusions to clearly distinguish between streamflow evaluation and flood hazard assessment. In particular, we will remove or soften statements that imply a direct assessment of flood hazard modelling performance and instead frame our findings in terms of runoff-product suitability for streamflow simulation and as a prerequisite for subsequent flood modelling applications. We will also explicitly acknowledge that evaluation using flood-specific metrics, such as flood peaks, inundation extent, or flood frequency characteristics, lies beyond the scope of the present study and represents an important direction for future research.

Reviewer comment 21: *Cama-flood simulation uses some river width and depth for era5land and ISIMIP3a? Or they calculate separately.*

Response 21: We thank the reviewer for this question. River geometry parameters were identical across all simulations.

Planned revision 21: We will clarify in the manuscript that all CaMa-Flood simulations used the same default river width and river depth datasets provided with the standard CaMa-Flood setup. These parameters were not recalculated separately for ERA5-Land or ISIMIP3a simulations. Consequently, differences in simulated streamflow arise from differences in the runoff forcing datasets rather than from differences in river geometry parameterization.