Answer to reviewer 02

General response

We thank the reviewer for the careful and constructive assessment of the manuscript. We agree that the two central issues are: (i) the scientific rationale for comparing ERA5-Land and ISIMIP3a runoff was not developed clearly enough, and (ii) the manuscript reported contrasting performance more strongly than it explained the mechanisms that could give rise to those contrasts. In a revised manuscript, we will therefore restructure the Introduction around runoff-product uncertainty and long-term streamflow/flood-change assessment, rather than broad flood-impact background. We will also clarify that the study is not designed to isolate a single causal factor, such as precipitation forcing alone or routing alone. Instead, it evaluates how two widely used but structurally different runoff-production workflows behave when routed through a common CaMa-Flood framework. The revised manuscript will make this scope explicit and will add targeted diagnostics to provide a more process-based interpretation of the results.

Reviewer comment 01: *The scientific question is not sufficiently developed. The manuscript mainly reports that ERA5-Land performs better for daily streamflow KGE, while ISIMIP3a gives more conservative long-term streamflow trends. More diagnostic analysis is needed to identify whether the differences arise from forcing construction, precipitation bias correction, model structure, ensemble averaging, human impacts, calibration, or routing interactions.*

Response 01: We agree. The original manuscript did not sufficiently distinguish between reporting product differences and explaining why such differences may arise. In a revised version, we will frame the central scientific question as follows: to what extent do streamflow performance and long-term streamflow-change diagnostics depend on the choice of runoff-production workflow when the river-routing framework is held fixed? This framing recognizes that ERA5-Land and ISIMIP3a differ not only in the runoff values they provide, but also in their underlying meteorological forcing, bias adjustment, model structure, ensemble design, spatial resolution, and treatment of human impacts.

Planned revision 01: To strengthen the mechanism discussion without claiming formal attribution, we will add targeted diagnostics along the modelling chain. First, we will compare key forcing characteristics, especially precipitation and temperature climatology/trends, between ERA5 and W5E5 where feasible, because these variables directly influence runoff generation and long-term drying or wetting signals. Second, we will compare pre-routing runoff characteristics, including runoff climatology, runoff trends, high-flow runoff metrics, and runoff ratio (R/P), between ERA5-Land and the ISIMIP3a ensemble. This will help distinguish cases where streamflow differences are consistent with differences in meteorological forcing from cases where differences emerge primarily during runoff generation. Third, we will compare selected pre-routing runoff diagnostics with post-routing discharge diagnostics within CaMa-Flood to assess whether the main spatial patterns are already present in runoff or are substantially modified by routing. These analyses will be presented as diagnostic evidence supporting plausible mechanisms, not as a complete decomposition of all uncertainty sources.

Reviewer comment 02: *The rationale for comparing ERA5-Land and ISIMIP3a is not clearly established. The authors should more clearly justify why these two products are selected and explain their fundamental differences in terms of data assimilation versus bias correction, land-surface versus hydrological model structures, single-product versus ensemble design, and their expected implications for streamflow simulation.*

Response 02: We agree that the original Introduction made the comparison appear too much like a comparison of two available datasets. In a revised manuscript, we will move the rationale into the Introduction and make it explicit before the dataset descriptions. The reason for selecting ERA5-Land and ISIMIP3a is that they are both spatially complete runoff products that can be routed consistently through the same river model, but they represent contrasting runoff-production workflows used in large-scale hydrological applications. ERA5-Land provides runoff from a single high-resolution land-surface model driven by ERA5 reanalysis forcing. ISIMIP3a provides runoff from multiple hydrological and land-surface models driven by bias-adjusted W5E5 forcing. This contrast allows us to evaluate whether daily streamflow performance and long-term streamflow-change diagnostics are robust across two influential runoff-generation pathways.

Planned revision 02: We will revise the Introduction to explain this distinction more concretely and will add a concise table in the Data section summarizing the major product differences: forcing source, bias adjustment, runoff model family, spatial/temporal resolution, single-product versus ensemble design, and human-water-management representation. The Introduction will state the scientific rationale, whereas the Data section will provide the technical details. The Discussion will then return to these differences only to interpret the results cautiously.

Reviewer comment 03: *The conclusions may be too dependent on a single routing model. All main results are derived from CaMa-Flood simulations, so it is difficult to separate runoff-forcing effects from routing-model effects. The authors should either test whether the conclusions are robust using other routing models or clearly limit their claims to the CaMa-Flood framework.*

Response 03: We agree that the current study uses a single routing framework and therefore cannot support routing-model-independent conclusions. We do not intend to claim that the same quantitative results will necessarily be obtained with another routing model. A multi-routing-model intercomparison will be valuable but is beyond the scope of this study, which is designed to hold the routing framework fixed while comparing two runoff-production workflows.

Planned revision 03: In a revised manuscript, we will consistently limit the conclusions to the CaMa-Flood framework. Statements such as "ERA5-Land performs better than ISIMIP3a" will be revised to "ERA5-Land-forced CaMa-Flood simulations perform better than ISIMIP3a-forced CaMa-Flood simulations" where appropriate. The Discussion will explicitly acknowledge that streamflow timing, peak attenuation, high-flow behaviour, and trend propagation may depend on CaMa-Flood's routing scheme, river network, floodplain storage representation, parameter settings, and dam treatment. We will not claim to have tested routing-model uncertainty.

Reviewer comment 04: *The conclusion that spatial resolution has limited influence is too broad. The manuscript only compares the native ERA5-Land runoff resolution with one aggregated resolution, and this single sensitivity test is insufficient to support a general statement that resolution has little impact on streamflow simulations.*

Response 04: We agree. The original wording was too broad. The aggregation experiment does not test the full role of spatial resolution in runoff generation, because ERA5-Land runoff is still generated at its native high resolution before being aggregated. Therefore, this experiment cannot answer whether a land model run natively at coarser resolution will produce similar runoff. It only tests whether providing the high-resolution ERA5-Land runoff field directly to CaMa-Flood, rather than an aggregated version, substantially changes the evaluated streamflow metrics.

Planned revision 04: We will retain the aggregation experiment because it is still informative, but we will revise the interpretation. If the ERA5-Land advantage remains after aggregation to the ISIMIP3a resolution, the appropriate conclusion is not that spatial resolution has little influence in general. Rather, the result suggests that the performance advantage is not solely due to the resolution of the runoff field supplied to the routing model. We will remove broader statements about spatial resolution and clearly state that the result is specific to one aggregation level, one routing framework, and the selected evaluation metrics.

Reviewer comment 05: *The Introduction devotes considerable space to the general impacts of flood hazards, but gives relatively limited attention to the uncertainty of runoff forcing products, which is the central issue of this study.*

Response 05: We agree. In a revised manuscript, the Introduction will be reorganized to reduce the broad discussion of flood impacts and to focus more directly on runoff-product uncertainty in global streamflow and flood simulations. The revised Introduction will first explain why long-term streamflow and flood-change assessments increasingly rely on spatially complete model-based runoff or discharge products, then discuss how different forcing products and runoff-generation workflows can affect simulated discharge. This will lead directly to the motivation for comparing ERA5-Land and ISIMIP3a within a common routing framework.

Planned revision 05: The revised Introduction will distinguish ERA5-Land and ISIMIP3a from already-routed discharge products such as GloFAS or GRADES. Such discharge products are valuable benchmarks, but they already embed their own hydrological and routing models. By contrast, ERA5-Land and ISIMIP3a provide runoff fields that can be routed through the same CaMa-Flood configuration, making them more suitable for evaluating sensitivity to runoff-production workflows under a controlled routing setup.

Reviewer comment 06: *Sections 2.1.2 ("Global flood simulation") and 2.2 ("Evaluation of simulations") contain duplicated descriptions of the CaMa-Flood model.*

Response 06: We thank the reviewer for identifying this redundancy. We agree that the duplicated description of CaMa-Flood unnecessarily lengthens the Methods section and reduces its clarity.

Planned revision 06: We will remove the duplicated CaMa-Flood description and reorganize the Methods section so that the model description, simulation setup, and evaluation methodology are presented only once and in a clear and logical sequence.

Reviewer comment 07: *In Section 3.1.1, the use of regional median KGE alone may not adequately represent model performance within each SREX region.*

Response 07: We thank the reviewer for this important observation. We agree that regional median KGE alone may mask substantial station-level variability, particularly within large and hydroclimatically heterogeneous SREX regions. Providing information on the distribution of station-level performance will therefore give a more complete representation of model skill within each region.

Planned revision 07: We will supplement the regional median KGE with additional distributional information, including the interquartile range, station-level KGE distributions, and the fraction of stations with positive KGE values. This will allow readers to assess both the central tendency and the spatial variability of model performance within each SREX region.

Reviewer comment 08: *In Section 3.2.1, the statement that the leave-one-out sensitivity tests indicate "moderate dependence on individual regions with pronounced hydroclimatic signals" is not clearly supported by Table S2.*

Response 08: We agree that the original interpretation was too strong. Table S2 reports the range of leave-one-out cross-regional correlations but does not by itself identify which excluded regions are responsible for the minimum or maximum values. Therefore, attributing the observed sensitivity to regions with pronounced hydroclimatic signals is not sufficiently supported by the analysis.

Planned revision 08: We will revise the relevant statement in Section 3.2.1 to avoid attributing the observed sensitivity to specific hydroclimatic characteristics that are not directly demonstrated by the analysis. We will instead describe the results more neutrally as indicating the degree of sensitivity of the cross-regional relationship to the exclusion of individual regions. We will also revise the Discussion accordingly to ensure that the interpretation remains consistent with the evidence provided by the leave-one-out analysis.

Reviewer comment 09: *Figure 4 presents trend comparisons using regional averages, but this aggregation may mask important station-level variability. The authors should consider adding a station-level comparison using all available gauging stations.*

Response 09: We agree that regional aggregation can obscure station-level variability and may mask geographically coherent differences between the two runoff products. We will therefore retain the regional averages as a concise summary of large-scale behavior, while supplementing them with a station-level comparison of observed and simulated trends across all GRDC stations included in the trend analysis. This spatial representation will allow readers to assess the degree of agreement between the two forcing datasets at individual stations and to identify potential spatial patterns in their differences that may not be apparent from the regional averages.

Planned revision 09: We will add a station-level spatial map showing the differences in simulated trends between the ERA5-Land- and ISIMIP3a-driven simulations for all GRDC stations included in the trend analysis. The new analysis will complement the regional averages in Figure 4 and provide a more complete assessment of the spatial variability and consistency of trend estimates across the observational network.