EGUSPHERE-2025-4387

Title: Validation of the Open-Source Hydrodynamic Model SFINCS on Historical River Floods at the Global Scale

Author(s): Tarun Sadana et al.

MS No.: egusphere-2025-4387

MS type: Research article

Dear Dr. Wouter Buytaert,

We would like to express our gratitude for the constructive and helpful comments from you and the two reviewers. They have contributed to substantial improvements in our manuscript.

Please find below the review comments (in black) and replies (in purple). The section, figure and table numbers refer to the original pre-print manuscript version.

With kind regards, and on behalf of all co-authors,

Tarun Sadana

Black (Comment)

Purple (Reply)

Referee 1: Francesco Dottori

Dear Dr. Francesco Dottori,

We would like to thank you for your time and effort to review the manuscript. It is very nice to hear that you think the paper is interesting for researchers working with large-scale flood models.

Please find below the review comments (in black) and replies (in purple). The section, figure and table numbers refer to the original pre-print manuscript version.

With kind regards, and on behalf of all co-authors,

Tarun Sadana

Reply to Francesco Dottori

This paper describes a modelling framework for simulating riverine flooding, focusing in particular on its validation. The authors investigate the sensitivity of the model performance to different geographic conditions, input data and modelling parameters, leveraging a large validation set of flood events at global scale. This allows for drawing general conclusions that in my opinion are interesting for researchers and practitioners working with large-scale flood models.

Overall the paper is clear and well structured, with a detailed analysis of the results and in-depth discusssions of several aspects influencing the skill of the model. Also, I would like to commend the open-source structure of the modelling framework, which gives more relevance to the work.

I have some moderate/minor comments and suggestions that should be addressed to improve the quality of the manuscript before publication .

Main comments:

Thank you for these helpful comments. Please find our reply to the main comments below.

Have you investigated the performance of the modelling framework in different climate zones? Hydrological models, including GloFAS, are usually less skillful in arid and cold regions, and this can negatively affect results. Given the large number of case studies, I think this factor could be included in the analysis.

This is indeed an interesting suggestion. Based on your comment, we have performed an analysis on how the model performs in different climate zones. We propose to include this as a new part (b) in Section 3.1.1 (“Model Performance in Different Climate Zones”), which reads as follows:

"**Section 3.1.1 (b) Model Performance in Different Climate Zones:**

To investigate the influence of climatic variations on model performance, each basin simulation was classified according to the Köppen-Geiger climate classification (Beck et al., 2018) using the centroid of each basin. Individual climate subclasses were grouped into the main climate classes: A (Tropical), B (Arid), C (Temperate), D (Cold/Continental), and E (Polar). This analysis was conducted for individual basins instead of flood events because flood events can occur over very large spatial scales and across multiple climate zones. Thereby, individual basins are weighted equally, meaning that smaller basins contribute to the overall average CSI in the same way as larger basins.

The results show clear variation in model performance across climate zones (Table 4). Performance is generally higher in tropical and temperate climates, where CSI values are 0.43 and 0.38, respectively. Lower performance is observed in arid climates with an average CSI of 0.32, while cold/continental climates show further reduced performance with an average CSI of 0.30. The lowest performance is observed in polar climates with an average CSI of 0.19. It is worth mentioning that most of the events simulated are based in Tropical (n = 193) and Temperate (n =139) climate zones.

Table 4: Model performance grouped per main climate zone. The number of flood events per zone is shown in brackets.

Climate Zone	Description	Average CSI
A (n=196)	Tropical	0.43
B (n=103)	Arid	0.32
C (n=139)	Temperate	0.38
D (n=54)	Cold/Continental	0.30
E (n=7)	Polar	0.19

This pattern is consistent with known limitations of both global hydrological models and satellite-derived flood observations in cold and polar regions, where snow and ice dynamics, frozen soils, and limited river gauges reduce the reliability of both the forcing data (GloFAS discharge; Harrigan et al., 2020) and the flood observations (Tellman et al., 2021). Performance of large-scale hydrological models in arid regions is usually lower compared to temperate and tropical regions (Salinas et al., 2013), as flooding in arid regions is driven by short intense rainfalls that are difficult to capture by hydrological models (Smith et al., 2015; Harrigan et al., 2020). In addition, runoff generation is highly variable due to infiltration and transmission losses in arid and semi-arid river channels, making peak flows harder to estimate (Dottori et al., 2016). Overall, these results suggest that the modelling framework performs most reliably in tropical and temperate climates, and that performance in arid and cold/polar regions is much lower."

Another point is the influence of flood protection structures. To my knowledge, global DEMs are currently not able to capture the presence of embankments along the rivers, and this limitation can heavily affect flood estimates, as observed by Dottori et al (2022) in several European river basins. The authors quicky discuss this issue at the end of Section 3.3 but I think it deserves some more detail. For instance, is there a difference in performance between more protected regions (e.g. rivers on Europe and US) and less protected ones (e.g. in Africa, South America)?

We agree that lacking infrastructure representation (e.g., levees, embankments) is a limitation of global DEMs. Following this suggestion, we have performed an exploratory analysis using the FLOPROS dataset (Scussolini et al., 2016) and the modelled floods. FLOPROS provides protection levels in terms of return periods which can be assigned to a basin cluster. We assigned the protection level to each basin cluster based on its centroid. We analyzed model performance across four key regions (Europe, North America, Africa, and South America), representing relatively highly protected and less protected regions. However, this analysis is exploratory and subject to several limitations, including the coarse spatial resolution of FLOPROS, uneven data availability (some regions have missing data), and confounding factors such as basin size and hydrological forcing. Therefore, to maintain focus and clarity in the main manuscript, we propose to include this analysis in the Appendix rather than as a separate main result section.

We present these results in a new paragraph in Appendices A3. The newly proposed section reads as follows:

"Section A3: Effect of Protection Standards on Model Performance:

A key limitation of global flood modelling frameworks is the inability of global DEMs to capture flood protection structures such as embankments and levees along river channels (Dottori et al., 2022). In this study, the modelled framework makes use of FABDEM (Hawker et al., 2022) and the floodplains are simulated using the subgrid approach. While subgrid methods have been shown to improve the representation of floodplains (Reshma et al., 2024), they still rely on the underlying DEM and therefore do not specifically take into account flood protection structures. To explore the effect of the protection standards on the model performance we make use of the FLOPROS dataset (Scussolini et al., 2016), which provides regional protection levels in terms of return period. Using this dataset, we assign flood protection standards (in return period years) at sub-national administrative level. As flood events can occur over a large area, we look at each basin cluster separately and assign values from the FLOPROS database. Some events occurring predominantly in low-income regions or smaller countries could not be assigned a protection level and were excluded from this analysis. The analysis was conducted in 4 key regions (North America, Europe, Africa, and South America). The results show consistent differences between regions, and the regions with higher protection standards (Europe and North America) show relatively higher bias values (1.12 and 1.26, respectively), indicating a tendency of the model to overestimate flood extent. In contrast, less well protected regions (South America and Africa) show lower bias values (0.96 and 0.89, respectively), suggesting more balanced or slightly underpredicted flood extents.

These differences can be explained by both the absence of flood protection structures in the DEM and regional variations in hydrological forcing. In highly engineered regions such as Europe and North America, levees and embankments are not represented in global DEMs (Dottori et al., 2022), leading to overestimation of flood extent. In contrast, in less protected regions, floodplain processes are closer to natural conditions, which can reduce this source of overprediction.

In addition, differences in hydrological model performance may also contribute to the observed patterns. Alfieri et al. (2013) showed that GloFAS discharge simulations perform well in river systems such as the Amazon and Mississippi but exhibit limitations in semi-arid regions (e.g. the Niger), where processes such as evaporation and infiltration are not fully captured. This can affect the accuracy of simulated floods and may contribute to regional differences in model performance observed in this study.

Table 5: Model Performance grouped per region.

Region	**FLOPROS protection range (return period years)**	**CSI**	**Bias**
Europe	2-10000	0.37	1.12
North America	2-500	0.41	1.26
South America	6.5-357	0.38	0.96
Africa	2-142	0.36	0.89

Minor comments:

Please add the scale on maps where appropriate (e.g. Figure 6 and 9)

Thank you for pointing this out. We revised Figures 6 and 9 to include scale bars in the revised manuscript to improve the readability of the maps.

Lines 47 "While local flood inundation studies have demonstrated promising results..." I would rather say that the use of local-scale flood models is a consolidated approach to map flood hazard and assess flood risk.

We agree with the reviewer and have revised the sentence to better reflect the role of local-scale flood models. The sentence is revised as following:

> "While the use of local-scale hydrodynamic flood models is a well-established approach for mapping flood hazard and assessing flood risk, large-scale or global inundation modelling still faces several challenges, such as limitations in topography and bathymetry data, computational intensity, and the open accessibility of datasets (Wing et al., 2020; Wing et al., 2021; Dottori et al., 2022)."

L 83-84: "Studies conducting event-based validation have not been global in scale, or they have focused on a limited number of events". I would include in the list the work by Risling et al (2024), as one of the few examples of validation studies involving multiple global flood models. The outcomes of the study (e.g. Figure 2) might be also relevant when discussing the performance of the SFINCS modelling framework.

Thank you for this suggestion. The study by Risling et al (2024) has been added to the in-text references. The revised sentence is as following:

> "Studies conducting event-based validation have not been deployed on a global level or have focused on a limited number of events (e.g. Risling et al., 2024, validating eight flood events across large river basins on four continents; Wing et al., 2021, using 35 events in the United States; Bernhofen et al., 2018, using five events in Nigeria and Mozambique)."

Section 3.2.3: My understanding is that all simulations described in this section are based on GloFAS discharge, is it correct?

The simulations performed in this section are indeed based on GloFAS discharge. All boundary conditions including discharge forcing were kept unchanged. Only the DEM input has been changed and the results are evaluated based on this change. We have clarified this in the text as follows:

"Section 3.2.3: The influence of the DEM input

> This section explores how different DEMs influence the performance of the SFINCS model in simulating flood extents. All simulations in this section are forced using GloFAS discharge, while only the DEM input is varied."

L 777-778: If I remember well, DEM for this study was also based on higher-resolution DTMs, can authors provide here more references?

We have clarified this by stating that in the study by Wing et al. (2017), the floodplains are represented by a DEM derived from 1 arc sec (~30m) USGS National Elevation Dataset (NED). We have revised the text as follows:

> "In contrast, studies that make use of observed discharge consistently achieved much higher performance. Wing et al. (2017) used the USGS National Elevation Dataset (1 arc sec or ~30m) and evaluated LISFLOOD against high water marks in 35 discrete U.S. events using gauge records and reported a CSI of 0.87 "

Section 4: are you planning to use the SFINCS to produce global-scale flood maps, similarly to what was done by JRC (see for instance the latest update by Baugh et al., 2024)?

The possibility of producing large-scale flood maps using SFINCS, similar to the work conducted by the JRC (Baugh et al., 2024), is indeed an important aspect. We are now exploring the opportunities for upscaling the GEB modelling framework. This is quite a complex task since it also involves upscaling other models such as the agent based model.

Can you include a link to the Global Flood Database and specify whether flood extent data are freely accessible?

The flood extent maps from the Global Flood Database (Tellman et al., 2021) are freely accessible. The dataset can be accessed through the project website (https://global-flood-database.cloudtostreet.ai) as well as through Google Earth Engine (https://developers.google.com/earth-engine/datasets/catalog/GLOBAL_FLOOD_DB_MODIS_EVENTS_V1).
We will include the links mentioned here in the Data Availability Statement. Thank you for pointing out this was missing in the previous manuscript version.

L1240:"Hence, the results were generated in a timely manner". This is rather vague. Can you provide a broad estimate of running times based on the temporal and spatial extent of the events?

We agree with the reviewer that the statement is vague. The simulations were analyzed and a broad estimate was calculated based on the spatial and temporal extent of all the flood events simulated in this study. The following sentence is added to the section "A2.2 Snakemake".

> "For example, events were split and executed in parallel across multiple compute nodes equipped with AMD EPYC 9354P processors (64 CPU cores, 96 GB RAM per node). Based on Snakemake log files, the average wall-clock time for a single basin simulation was approximately 7 minutes for a ~1,000 km² domain and a typical simulated flood period of ~30 days.
>
> Each flood event consists of multiple basin cluster simulations (on average ~30 per event), resulting in a large number of individual simulations across the full dataset. These simulations were executed in parallel; however, the overall workflow also includes sequential steps per event (e.g. model setup, forcing update, simulation, and post-processing), as well as I/O and scheduling overhead. As a result, the total elapsed wall-clock time for completing all simulations was approximately 44 days."

References

Risling, A., Lindersson, S. & Brandimarte, L. A comparison of global flood models using Sentinel-1 and a change detection approach. Nat Hazards 120, 11133–11152 (2024). https://doi.org/10.1007/s11069-024-06629-7

Baugh, C.; Colonese, J.; D'Angelo, C.; Dottori, F.; Neal, J.; Prudhomme, C.; Salamon, P. (2024): Global river flood hazard maps. European Commission, Joint Research Centre (JRC) [Dataset] PID: http://data.europa.eu/89h/jrc-floods-floodmapgl_rp50y-tif

References:

- Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., & Wood, E. F. (2018). Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific Data*, *5*(1), 180214. https://doi.org/10.1038/sdata.2018.214
- Tellman, B., Sullivan, J.A., Kuhn, C. et al. (2021) Satellite imaging reveals increased proportion of population exposed to floods. Nature 596, 80–86 (2021). https://doi.org/10.1038/s41586-021-03695-w
- Harrigan, S., Zsoter, E., Alfieri, L., Prudhomme, C., Salamon, P., Wetterhall, F., Barnard, C., Cloke, H., & Pappenberger, F. 990 (2020). GloFAS-ERA5 operational global river discharge reanalysis 1979–present. Earth System Science Data, 12(3), 2043–2060. https://doi.org/10.5194/essd-12-2043-2020
- Smith, A., Sampson, C., & Bates, P. (2014). Regional flood frequency analysis at the global scale. Water Resources Research, 51(1), 539–553. https://doi.org/10.1002/2014wr015814
- Dottori, F., Salamon, P., Bianchi, A., Alfieri, L., Hirpa, F. A., & Feyen, L. (2016). Development and evaluation of a framework for global flood hazard mapping. Advances in Water Resources, 94, 87–102. https://doi.org/10.1016/j.advwatres.2016.05.002
- Scussolini, P., Aerts, J. C. J. H., Jongman, B., Bouwer, L. M., Winsemius, H. C., De Moel, H., & Ward, P. J. (2016). FLOPROS: an evolving global database of flood protection standards. *Natural Hazards and Earth System Sciences*, *16*(5), 1049–1061. https://doi.org/10.5194/nhess-16-1049-2016
- Reshma, R., Devi, N. N., & Kuiry, S. N. (2024). Real-time urban flood modeling: exploring the sub-grid approach for accurate simulation and hazard analysis. *Natural Hazards*, *120*(11), 9609–9647. https://doi.org/10.1007/s11069-024-06579-0