https://doi.org/10.5194/egusphere-2025-4387
Preprint. Discussion started: 6 November 2025

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

Tarun Sadana[1], Jeroen C.J.H. Aerts[1,2], Dirk Eilander[1,2], Bruno Merz[3,4], Hans de Moel [1], Tim Busker[1], Veerle Bril[1] , Jens de Bruijn[1,5]

[1] Institute for Environmental Studies (IVM), Vrije Universiteit Amsterdam, Amsterdam, the Netherlands
[2] Deltares, Delft, the Netherlands
[3] GFZ Helmholtz Centre for Geosciences, Potsdam, Germany
[4] Institute for Environmental Sciences and Geography, University of Potsdam, Germany
[5] International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria

Correspondence: Tarun Sadana (t.sadana@vu.nl)

Keywords: hydrodynamic modelling, riverine flood, open-source, large-scale flood modelling

Abstract

We evaluate the performance of the Super-Fast INundation of CoastS (SFINCS) hydrodynamic model for simulating riverine floods, combined with a fully automated open-source data preprocessing pipeline. To do this, we assessed the simulated extent of 499 historic flood events against the satellite derived flood extents using the Critical Success Index (CSI) as a performance metric. We utilised simulated discharges from the Global Flood Awareness System (GloFAS) hydrological model and found that SFINCS performance improved with upstream basin size, with a global mean CSI of 0.42 for basins with large upstream area (>1,000 km²) and a CSI of 0.29 for basins with small upstream area (<50 km²). Our results illustrate the importance of accurate discharge data input to flood hazard simulations. When the (globally simulated) GloFAS data replaced with observed discharge data for ten events in the US, the CSI improved from 0.39 to 0.67. These results suggest that global hydrological model performance limits the accuracy of the flood hazard simulations. Our findings also showed a significant improvement in the CSI (from 0.37 to 0.57) when changing to a higher-resolution elevation input by contrasting a ~1m digital elevation model (DEM; 3DEP) with our default ~30m global DEM (FABDEM) in six U.S. events. Sensitivity analysis of bathymetric calculations revealed a systematic underestimation of the default 2-year return period estimated by GloFAS discharge, likely driven by underrepresentation of annual block maxima, which resulted in underestimated channel dimensions. All of these factors resulted in a loss of detail, which impacted model performance, especially in smaller headwater rivers. We recommend to improve the estimation of bathymetry, for instance by employing the "gradually varying solver" method or using data from the SWOT mission. Furthermore, incorporating additional validation data which ideally includes flood depth measurements can largely enhance our understanding of the model performance.

1. Introduction

Riverine floods pose a significant global challenge, which makes their assessment crucial for designing effective flood management strategies that reduce flood damage and fatalities (Merz et al., 2021). Over the past decades, flood inundation modelling has become an important tool for assessing flood hazard and supporting flood managers in prioritising interventions to reduce risk (Teng et al., 2017). These modelling efforts often involve using hydrological forcings in hydrodynamic models to simulate water movement. These models usually solve equations derived from physical laws of fluid motion to produce

https://doi.org/10.5194/egusphere-2025-4387
Preprint. Discussion started: 6 November 2025

flood extent and depth maps (e.g. Guo et al., 2021). However, other approaches exist, such as using machine learning (Nevo et al., 2022).

While local flood inundation studies have demonstrated promising results, large-scale or global inundation modelling still faces several challenges, such as topography and bathymetry data, computational intensity, and open accessibility of data (Wing et al., 2020; Wing et al., 2021; Dottori et al., 2022). For example, in recent years, several high-resolution global topography datasets have become available (Hawker et al., 2022; Abrams et al., 2020; Copernicus DEM, 2022), with increasing accuracy. Nevertheless, this data is too coarse to include river flood defences (Wing et al., 2019). Another significant challenge is the lack of open, high-resolution bathymetry data (Hawker et al., 2018). While river-width data has become available for rivers wider than 30 metres (Allen & Pavelsky, 2018), data on river depth is still missing.

Furthermore, the computational demand to run large-scale models is high (Leijnse et al., 2021). Two-dimensional hydrodynamic models, such as the state-of-the-art global LISFLOOD-FP model, offer high detail but at a substantial computational cost (e.g. Wing et al., 2021; Bates, 2023). The computational intensity at the global scale has led to different modelling simplifications (e.g. Winsemius et al., 2015; Van Ormondt et al., 2025) and the development of novel computational approaches using graphic processing unit (GPU) architecture (Shaw et al., 2021; Apel et al., 2024). Lastly, not all parts of the code for setting up global flood models are open-source, which limits comparability and reproducibility (Hall et al., 2022; Hoch and Trigg, 2019).

Due to these challenges, there is growing interest in open-source hydrological and hydrodynamic models (e.g. HEC-RAS; Zeiger and Hubart, 2021) that are easily applicable to data-scarce regions and computationally efficient (Kim et al., 2019). One such model is the Super-Fast INundation of CoastS (SFINCS) model (Leijnse et al. 2021), which has shown promise in coastal and compound flooding scenarios (Eilander et al., 2023b; Nederhoff et al., 2024) but can also be applied to fluvial and pluvial flooding. SFINCS achieves fast computational speeds by simplifying dynamic flow equations and using efficient spatial discretisation techniques with a subgrid (Leijnse et al., 2021). Another key advantage of the SFINCS model is that it is fully open-source, alongside (Python) packages to preprocess the data (Eilander et al., 2023a).

Although the SFINCS model has demonstrated high performance in modelling compound floods in coastal areas at both small and larger scales (e.g. Leijnse et al., 2021; Benito et al., 2024), its performance for riverine floods has not yet been investigated. Understanding how SFINCS performs for riverine floods, particularly at large scales and across different parts of the world, is important, especially due to the aforementioned challenges related to input data and model simplifications. Validation plays a key role in understanding model reliability, especially for flood hazard models applied at continental to global scales (Ward et al., 2013; Bates, 2023). However, global data for model validation, such as flood extent observations (e.g. Sampson et al., 2015), are scarce, with most large-scale studies lacking extensive validation against real flood events. Instead, these studies have often focused on producing flood hazard maps for several return periods (e.g. Dottori et al., 2022) and benchmarking these simulations against other national and regional engineering models (e.g. Wing et al, 2024). Studies conducting event-based validation have not been global in scale, or they have focused on a limited number of events (e.g. Wing et al., 2021, using 35 events in the US; Bernhofen et al., 2018, using five events in Nigeria and Mozambique). In the last decade, the use of satellite imagery has become an increasingly popular source for validating flood hazard models (e.g. Bernhofen et al., 2018; Dottori et al., 2016; Masafu and Williams, 2024; Landwehr et al., 2024). More validation of large-scale models can improve our understanding of how flood hazard models perform on a global scale and under varying environmental and climatic conditions. The limited validation at larger scales, along with the regional focus of many studies, restricts this understanding.

https://doi.org/10.5194/egusphere-2025-4387
Preprint. Discussion started: 6 November 2025

Therefore, the main goal of this paper is to validate the SFINCS hydrodynamic model using satellite-derived flood extent observations for a large set of global historical flood events (n = 499; Cloud to Street, 2022). The main novelty of our study lies in conducting a comprehensive validation and sensitivity analysis of SFINCS performance on a global scale to provide insights into its suitability for global flood hazard assessment. In addition, we publish an entirely open-source automated workflow that leverages only open-source data and inundation models. To achieve this, we couple SFINCS with different hydrological models, which provide forcing data at predefined inflow and headwater points for SFINCS. This event-based validation approach offers a unique opportunity to understand riverine flood behaviour better, as emphasised by Grimaldi et al. (2019).

2. Methodology

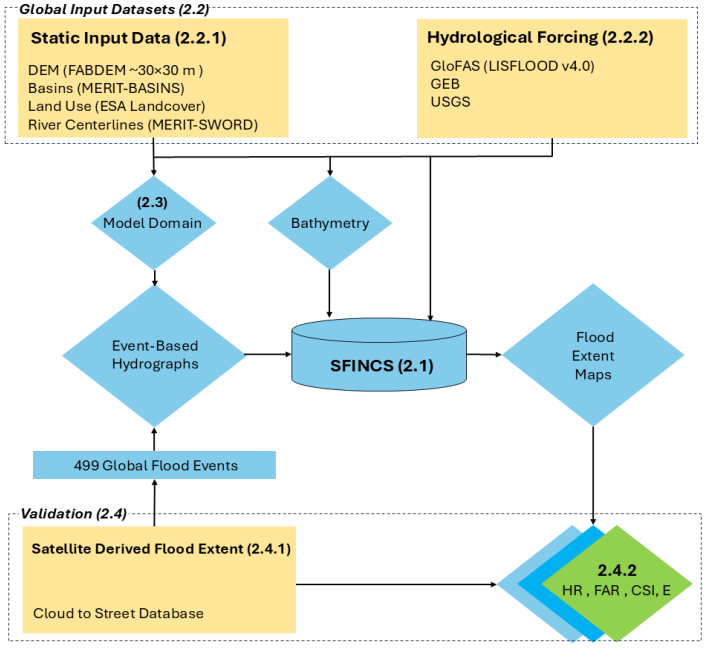


Figure 1. Automated framework for testing of the global hydrodynamic model SFINCS (section numbers are in brackets).

Figure 1 presents a fully automated global flood modelling framework integrating various data inputs and processing methods with validation steps. We set up this modelling framework for each of the 499 satellite-derived flood extents (2.4.1). The rectangular yellow boxes represent input datasets: the static input data (2.2.1) and the hydrological forcing data from two hydrological models and observed river gauges (i.e. the geographical, environmental, and behavioural [GEB] model, the Global Flood Awareness System [GloFAS], and U.S. Geological Survey [USGS] data; 2.2.2). The diamond-shaped boxes (Figure 1) represent the derived data (e.g. event-based hydrographs and bathymetry) made from input datasets. Event-based hydrographs were derived by combining the hydrological forcing data with the timings of the satellite-derived flood extents. The 499 flood events corresponded to the timings of floods within the 2000–2018 period of satellite-derived flood extent (2.4.1). The SFINCS model (2.1) generated flood extent maps, which were validated (2.4) by comparing SFINCS outputs against

https://doi.org/10.5194/egusphere-2025-4387
Preprint. Discussion started: 6 November 2025

satellite-derived flood extents from the Cloud to Street Database. The validation metrics included the hit rate (HR), false-alarm ratio (FAR), Critical Success Index (CSI), and error bias (E).

2.1 The hydrodynamic model SFINCS

SFINCS is a fully automated, 2D reduced-complexity hydrodynamic model that solves simplified Saint-Venant equations of mass and momentum (Leijnse et al., 2021). The model uses a localised inertial equation (LIE) formulation, which neglects advection, Coriolis force, and viscous effects, simplifying the full Saint-Venant equations (Bates et al., 2010). These simplifications significantly reduce computational costs while retaining sufficient floodplain accuracy (Leijnse et al., 2021). A comparison with full-physics models such as Delft3D (Deltares, 2022) revealed that SFINCS can achieve up to 100 times faster computation speed for equivalent domains with negligible loss in accuracy for sub-critical flow conditions (Leijnse et al., 2021). The model's reduced complexity enables rapid simulations, which makes it suitable for large-scale or global modelling applications like this study. For this study, we utilised SFINCS in subgrid mode to leverage its ability to integrate high-resolution elevation data while simulating flood dynamics efficiently across large river basins. We developed uncalibrated models by automating the model setup and execution process to ensure consistency and minimise manual intervention. While SFINCS was originally developed for coastal flooding applications, it has been adapted to model riverine flood events in this study. The fluvial setup focuses on river discharge as the primary forcing and incorporates several key additions to the coastal-oriented model configuration. These additions to enable fluvial flooding were as follows:

Bankfull width and depth estimation: These parameters determined the channel capacity of the rivers and thus indicate the moment after which inundation processes started. Calculating the bankfull discharge is explained in Section 2.2.3 and Appendix A1.

Hydrograph input and global model setup: We chose an automated event-based approach to assess the model's robustness and performance under varying conditions. Our simulations were structured event by event and assumed each observed flood event as a distinct and independent occurrence. The timings of the input hydrographs and model domain extent were derived from moderate resolution imaging spectroradiometer (MODIS) satellite observations, compiled into the Cloud to Street database (Section 2.4.1). This database includes a subset of 499 riverine flood events spanning five (sub)continents and 96 countries. To simulate floods on a global scale for each event, discharge data was obtained from global hydrological model LISFLOOD v4.0, as described in Section 2.2.2. Each event was processed and run separately, with all associated flooded basins (MERIT-BASINS) simulated in one model run. Notably, a single flood event can occur across multiple basins. We refer to these combined basins as the "model domain". This setup allowed us to handle large-scale flood modelling tasks efficiently. By applying the model across many events, we validated the model across a wide range of flood magnitudes in different geographical (e.g. catchment sizes) and climatological settings.

Headwater and river inflow points: Discharge source points were set up in the SFINCS model to simulate water coming in from rivers. These source points can be *river inflow points,* which simulated water entering the model domain from upstream rivers (e.g. rivers that cross the boundary of the model). Water can also enter the model coming from headwater rivers, which are streams that originate inside the model domain. The start or origin of these headwater rivers is called the *headwater point.* Accurate placement of both headwater and inflow points is essential for reliable flood modelling since the misalignment of these points can lead to the significant over- or underestimation of downstream flooding, particularly during catchment-wide events where multiple tributaries converge into a main river channel. Therefore, to ensure hydrological consistency between the hydrodynamic model and the hydrological input (which is at a much coarser resolution), headwater and inflow points are snapped to the grid cell that best matches the upstream contributing area of the hydrological model used. This

https://doi.org/10.5194/egusphere-2025-4387
Preprint. Discussion started: 6 November 2025

snapping was implemented using the HydroMT-SFINCS plugin (see Section 2.3: Data handling and pre-processing). Discharge input for inflow and headwater points was sourced from multiple datasets, as described in Section 2.2.2. To account for differences in resolution and small mismatches in the upstream area, we applied a tolerance of 5% deviation when snapping. This approach helped to ensure that discharge was applied at hydrologically consistent locations, which reduced the risk of errors in flood extent prediction due to poor spatial alignment between the hydrological and hydrodynamic models.

2.2 Global Input Datasets

2.2.1 Static Input Data

Static input data provided essential environmental and topographical information to the SFINCS model. The following datasets were used for the global automated setup which include:

Digital Elevation Model (DEM): The FABDEM V1-2 dataset (Hawker et al., 2022) was used to provide elevation data at a global scale. This dataset has a 30-metre resolution and is proven to be more accurate than existing global elevation datasets (Hawker et al., 2022).

Land-Use Data: Land-use patterns were derived from the ESA Worldcover 2021 dataset on a 10-metre resolution (Zanaga et al., 2022). The land-use data was used to define Manning's roughness coefficient for different land-use types.

River Network Data: River geometries were obtained from the MERIT-SWORD dataset v0.4 (https://zenodo.org/records/14675925), which is a dataset created by combining the SWOT River Database (SWORD; Altenau et al., 2021) and MERIT-BASINS (Lin et al., 2020). The MERIT-SWORD dataset transfers data, such as river width from SWORD rivers (30 m wide and greater), to corresponding MERIT-BASINS rivers by generating bidirectional links. Drainage areas smaller than 25 km², including non-channelised areas along the coast or certain endorheic regions (e.g. incomplete basins or hillslopes), were excluded to maintain focus on larger river systems in this dataset.

Basin Delineation: The MERIT-BASINS dataset (Lin et al., 2020) was used to define the boundaries of individual basins. Each individual basin in MERIT-BASINS carried a unique basin identifier and an upstream area attribute that directly corresponded to a specific river reach in the MERIT-SWORD network.

2.2.2 Hydrological Forcing

Discharge data required to force the SFINCS model at predefined discharge points (headwater and inflow points) was obtained from three sources, two hydrological models and a database with observed discharges. The primary global dataset was GloFAS (Grimaldi et al., 2022), which provided long-term global coverage discharge data suitable for simulating floods at a global scale. The second source was the GEB model (De Bruijn et al., 2022), which generates higher-resolution behaviourally informed discharge estimates that are particularly useful in basins where human interventions significantly influence flood dynamics. Finally, observed discharge records from the USGS database (USGS, 2023) were used as a benchmark to evaluate the performance of the model predictions in selected U.S. basins.

GloFAS (LISFLOOD v4.0): The GloFAS dataset (Alfieri et al., 2013; Grimaldi et al., 2022) is generated using the global hydrological model LISFLOOD-OS, which is a distributed, physically based rainfall-runoff model that uses the ERA5 reanalysis data as input to simulate global river discharge at ~5.5 km resolution at the equator. Because the timing of the peak discharges in GloFAS hydrographs can differ from the timing of the floods observed in the satellite-derived extent maps (Section 2.4), we captured GloFAS data 10 days before and 10 days after the timing of the flood event in the observed data. In this way, we are sure the flood event (i.e. discharge peak) was in the GloFAS dataset. The GloFAS dataset

https://doi.org/10.5194/egusphere-2025-4387
Preprint. Discussion started: 6 November 2025

was used as the default dataset for discharge in the global automated setup. The GloFAS dataset was selected because it offers long-term, global coverage of discharge data, which made it suitable for simulating floods on a large scale. The following mentioned datasets were used in the sensitivity analysis for this study (Section 2.5):

GEB Model: The GEB model (De Bruijn et al., 2022) integrates an agent-based model (ABM) and a hydrological model to simulate the flood and drought management decisions of farmers and urban households interacting with the hydrological system. Within GEB, the ABM is dynamically linked with the spatially distributed grid-based hydrological model CWatM at ~1km resolution at the equator (Burek et al., 2020). The GEB model incorporates not only surface and groundwater hydrological processes but also the impacts of human activities like water consumption and reservoir operations. We used the daily discharge estimates from the GEB model as input to SFINCS during the sensitivity analysis (Section 2.5). The GEB model was selected because it captures socio-hydrological interactions and provides high-resolution, behaviourally informed discharge estimates, which are particularly useful in basins where human interventions significantly influence flood dynamics, such as the Krishna Basin in India, where it was applied in this study (see Section 2.5).

USGS Discharge Observations: The USGS (2023) provides real-time streamflow data, which we used as a benchmark to evaluate how well the global hydrological model's predicted discharge aligned with actual observations in a selection of U.S. basins (Section 2.5). We limited the benchmark to U.S. basins due to the availability of high-quality, long-term, and consistent discharge records from the USGS, which have been globally recognised for their reliability.

2.2.3 Bathymetry

Bankfull discharge represents the level at which a river fills its channel without overflowing and is a key threshold in determining the channel dimensions of a river. We followed the approach outlined by Sampson et al. (2015), using bankfull discharge as the primary indicator for estimating the river's cross-sectional shape. To calculate the bankfull discharge, we used the 2-year return period calculated from the GloFAS hydrological model discharge (Section 2.2.2). The 2-year return period was selected because it closely corresponds to the bankfull flow for many rivers, representing a typical flow regime that occurs frequently enough to shape the channel's morphology over time (e.g. Edwards et al., 2019). However, recent studies have shown that this threshold can vary significantly across different (sub)basins (Roy and Sinha, 2016). Therefore, we included a sensitivity analysis to assess the impact of this assumption (Section 2.5). This allows for modelling of channel dimensions based on recurrent hydrological conditions. We then assumed a rectangular cross-sectional shape for the river channels, which was burned into the DEM (for more detailed information, see Appendix A1).

To estimate the bankfull width, we primarily used values from the MERIT-SWORD database, which integrates observations from multiple global rivers and satellite-related datasets (see Section 2.2.1). For the river sections whose values were not represented in MERIT-SWORD, we applied a power-law relationship between bankfull discharge and channel width, as proposed by Leopold and Maddock (1953). This empirical relationship reflects how larger discharges are accommodated by wider channels, while smaller discharges are carried by narrower channels. This approach ensured consistency in width estimation where observational data was lacking and serves as a gap-filling strategy to ensure complete channel representation across the river network.

The bankfull depth was then calculated using Manning's open channel flow equation. This equation accounts for both riverbed roughness and slope. Importantly, the equation also incorporates the bankfull width either observed (e.g. from MERIT-SWORD) or estimated using the power-law relationship ensuring that depth estimates were consistent with the defined channel geometry.

In the sensitivity analysis, we also use the GEB model outputs and the USGS discharge time series to estimate the 2-year return period and the related bathymetry (Section 2.5).

https://doi.org/10.5194/egusphere-2025-4387
Preprint. Discussion started: 6 November 2025

2.3 Determining the Model Domains

Basin delineation was guided by observed flood events, which are identified from satellite-derived flood maps (Cloud to Street database; Section 2.4.1). These events helped determine which basins were affected and how they were grouped together (here referred as 'basin clusters" (see Figure 2) during each flood. The following criteria were applied during this step to delineate the basins:

- *Delineating basin clusters:* We delineated clusters of hydrologically connected basins for each event to capture the structure and connectivity of flooded areas within a catchment. First, we identified which HydroBASINS Level 8 basins were flooded. Flooded basins touching the coastal areas were excluded from the analysis; only inland basins were considered. Then, we used these flooded HydroBASINS as a guide to select MERIT-BASINS, which ensured the simulation areas were neither too large nor too small. As a result, the model domain regions were not larger than HydroBASINS Level 8, which kept the SFINCS simulations efficient. Next, we focused on converting the MERIT river network into a directed graph based on downstream flow connectivity (Figure 2B). Using the river network graph, we built a network of basins that were hydrologically connected. Next, we identified flooded basins by overlaying the observed flooded pixels from satellite imagery with MERIT-BASINS boundaries. Any basin that intersected with observed flood pixels (based on the aforementioned criteria) was marked as flooded. We also included a user-defined number of downstream basins (default: 1 basin) into the simulated area. Doing so helps to reduce the 'boundary effects' problem that could happen at the edges of the simulation when water could not flow properly out of the model area. Including downstream areas aimed to help prevent distortions near the outflow edge and maintain flood flow dynamics in the hydrodynamic simulations. Subsequently, we analysed the resulting set of basins (flooded MERIT-BASINS + 1 downstream basin) and group them together in a cluster, as illustrated by the subsets shown in Figure 2 (e.g. Subsets 1–3). Each cluster represented a distinct set of hydrologically linked basins affected by the same flood event. Subsequently, we ran the hydrodynamic simulations (SFINCS) for each cluster separately using snapped (see Section 2.1) discharge points with simulated discharge as input into the model domain. This approach ensured that the full spatial structure of each flood event was respected, enabling more realistic hydrodynamic modelling across complex river networks.
- *Flooded pixel area (>1% model domain area)*: The flooded pixel area after the delineation should cover more than 1% of the basin cluster area. This approach ensured that basins without a substantial flood were excluded to omit irrelevant model runs and shorten run times.

https://doi.org/10.5194/egusphere-2025-4387
Preprint. Discussion started: 6 November 2025

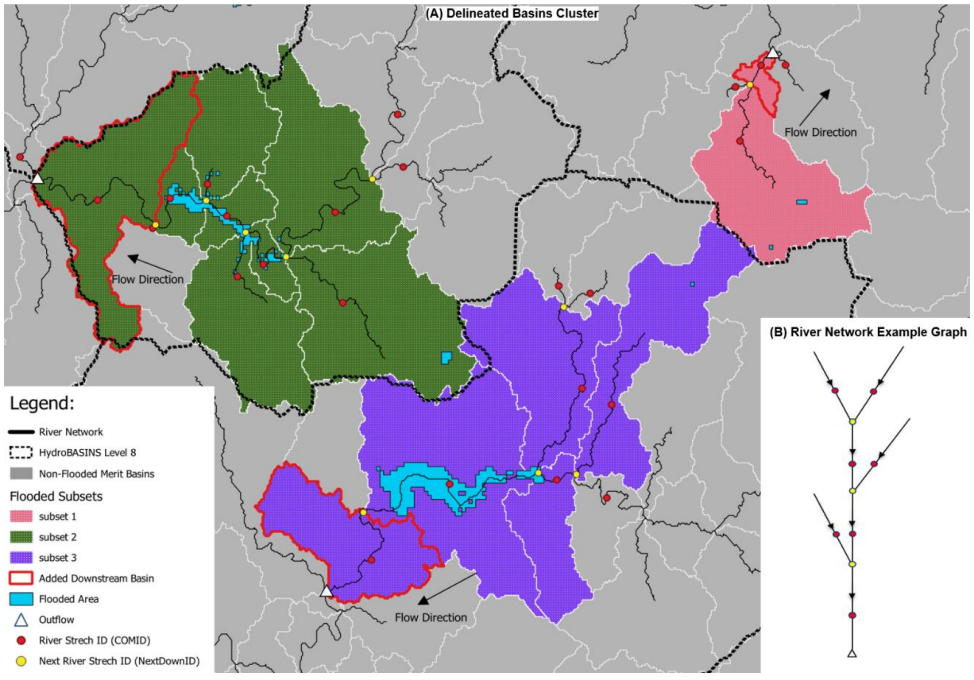


Figure 2: (A) An example of a delineated basin cluster for flood event observed between 22 December 2013 and 4 January 2014 in the United States, showing how individual basins affected by the flood were grouped based on hydrological connectivity. (B) An example river network graph used to represent downstream flow connectivity between basins, as derived from the MERIT river network.

Data handling and preprocessing: We used the HydroMT package (Eilander et al., 2023a) to set up a separate model for each basin cluster across all events. HydroMT handled the automatic data processing and configuration of the SFINCS hydrodynamic model while integrating multiple data sources (e.g. hydrographs, DEMs, and land-use maps). This automation reduced potential errors in manual data processing and enabled the rapid testing of different scenarios. The modular setup of the model allows for further iterative enhancements to the model, ensuring it can effectively adapt and improve as new data is integrated over time. To run both HydroMT and SFINCS for all clusters and events, we further automated our approach using Snakemake (Mölder et al., 2020), a Python-based workflow management system to manage parallel simulations efficiently. Appendix Section A2 explains the use of Snakemake and the data handling of our study in more detail.

2.4 Validation

2.4.1 Satellite-Derived Flood Extent

A validation procedure was designed to validate the simulated flood extents predicted by the SFINCS model. We evaluated whether these simulated flood extents aligned with satellite-derived flood events provided by the Cloud to Street (2022), also known as the Global Flood Database. This database was selected because it offered several key advantages for this type of analysis. Firstly, the database provided global coverage, which ensured that flood events from diverse regions around the world were represented. Secondly, it organised all satellitederived flood events into clear categories of flood drivers. Thirdly, the data was openly accessible and downloadable, which streamlines integration into research

https://doi.org/10.5194/egusphere-2025-4387
Preprint. Discussion started: 6 November 2025

studies. Importantly, all the flood maps were consolidated under a single dataset, which eliminated the need to combine multiple data sources.

The database was developed using data from NASA's MODIS satellite which captures daily flood extents documented by the Dartmouth Flood Observatory (DFO) with a spatial resolution of 250m. MODIS provides consistent global coverage with a daily revisit cycle, making it well-suited for large-scale flood monitoring. The full dataset included 913 observed flood events from 2000 to 2018 (Tellman et al., 2021), which represented the maximum observed surface water extent over the event duration. However, it is important to note that these maximum flood extents may be underestimated due to limitations such as cloud cover, vegetation, or narrow channels that may obscure floodwaters in the satellite imagery (Tellman et al., 2021). To prepare the dataset for validation, we selected events (n = 499) from the full database by applying the following filters:

- *Flood type in observations (heavy rain only):* The global flood database lists the main flood drivers, which we filtered only to include floods caused by "Heavy Rain", which excluded other flood drivers such as tropical storm surge, dam breaks.
- *Flood duration (<30 days):* Floods lasting more than 30 days were excluded because the satellite-observed flood images in the analysis are composite images created from multiple days of observations. During longer flood events, the flood extent may change over time, potentially leading to inaccurate or mixed representations of the flood. By focusing on shorter floods, we ensured that the composite images more accurately reflected the flood's maximum extent, avoiding the temporal variations that can occur during longer-lasting floods. This approach improved the reliability of the flood extent data used for model validation and analysis (Cloud to Street, 2022).
- *Permanent water-body masking:* Permanent water bodies were masked from the observed flood dataset by removing permanent water pixels extracted from the JRC Global Surface Water Mapping Layer v1.1 (Pekel et al., 2016). This procedure excluded pixels that were always flooded (e.g. the rivers themselves and large parts of reservoirs) to ensure that only genuinely flooded areas which are not typically inundated were included in the analysis. These permanent water features are always present and do not reflect new inundation caused by flooding events. This helped in accurately assessing flood extents and refining the validation of hydrodynamic models based on observed flood extent (Fleischmann et al., 2019).

2.4.2 Validation Metrics

We computed the maximum flood extent across all time steps to generate a single inundation map representing the peak of the event for every SFINCS simulation. The observation data only showed if a pixel was flooded or dry (without depth data), so we also changed the SFINCS output to a pixel with only two states: either flooded or dry. In a postprocessing procedure, a minimum depth of 0.05 m was used to classify a pixel as flooded (Wing et al., 2024). Note that we resampled the high-resolution SFINCS outputs of 30 m to 250 m using mode resampling to match the resolution of observed flood extents. This approach meant that a 250 m SFINCS output pixel was classified as flooded if the majority (i.e. more than one-half) of the underlying SFINCS 30 m cells were flooded. This method ensured consistency with the binary "flooded or dry" classification of MODIS data while reflecting the majority condition in each area. Notably, the permanent water bodies (already 250 m), identified from band 5 of the satellite data, were also masked from the SFINCS simulated flood event after it has been resampled to 250 m. Importantly, the additional downstream basin included in the model domain to avoid boundary effects was not considered when applying the validation criteria. This approach ensured that the validation only assessed the basins directly affected by flooding to provide a fair comparison between the simulated and observed flood extents.

https://doi.org/10.5194/egusphere-2025-4387
Preprint. Discussion started: 6 November 2025

Table 1 shows the validation metrics used to compare SFINCS results with the observations. We validated our maps for each flood event by calculating the contingency table values (i.e. Hit, Miss, False Alarm, and Correct Negative) over the full extent (i.e. all basins) of the event.

Table 1. Contingency table used to calculate the SFINCS validation metrics.

	Flooded Observation Pixel	**Dry Observation Pixel**
Flooded Model Pixel	Hit	False alarm (FA)
Dry Model Pixel	Miss	Correct negative (CN)

Hit Rate (HR; Eq. 1): The HR measures the proportion of the flood area observed that was successfully predicted by the model, balancing misses (*Miss*) and hits (*Hit*). The HR can range from 0 to 1. A higher HR indicates better model accuracy in terms of correctly predicting the flood extent.

$$Hit\ Rate = \frac{Hits}{(Hits + Misses)} \ (Eq. 1)$$

False Alarm Ratio (FAR; Eq. 2): The FAR represents the proportion of predicted flood areas that did not flood in satellite-derived flood extents. The FAR can range from 0 to 1. A low FAR suggests that the model is conservative in its flood predictions.

$$FAR = \frac{FA}{(FA + Hits)} \ (Eq.\ 2)$$

Critical Success Index (CSI; Eq. 3): The CSI balances the HR and FAR in Equations 1 and 2 to provide a comprehensive metric of the overall model performance. The CSI can range from 0 to 1, with 1 indicating a perfect model. The CSI accounts for correct predictions (i.e. *Hits*) and wrong predictions (i.e. *Misses* and *FAs*), offering a more nuanced view of model accuracy.

$$CSI = \frac{Hits}{(Hits + FA + Misses)} \ (Eq.\ 3)$$

Error Bias (E; Eq. 4): E is a metric used to determine whether a model is overpredicting or underpredicting the occurrence of an event, such as flooding. A value of $E = 1$ indicates no bias, meaning the model's predictions are balanced between overprediction and underprediction. If E falls between 0 and 1, the model tends to underpredict, while values greater than 1 suggest the model overpredicts the event.

$$E = \frac{FA}{Misses} \ (Eq.\ 4)$$

2.5 Sensitivity Analysis

The sensitivity analysis focused on how variations in key parameters affected the SFINCS model's flood predictions. For this study, we focused on three critical factors: river depth estimation, elevation data, and discharge data. These parameters were chosen because they directly influenced flood extent predictions, and variations in their accuracy can have significant implications for model performance in large-scale flood modelling. The rectangular cross-section used in SFINCS is a widely adopted representation in hydrodynamic modelling, consistent with prior studies such as Neal et al. (2021), and the river roughness value (0.02) derived from widely accepted Manning's roughness values. For the sensitivity analysis, we focused on two geographical regions: the US and India. We selected the flood events from the Global Flood Database using the DFO's "Severity Level 2" classification (Tellman et al., 2021), which identifies floods with high impacts (recurrence interval >100 years). The severity level

https://doi.org/10.5194/egusphere-2025-4387
Preprint. Discussion started: 6 November 2025

was a criterion for choosing representative events for the sensitivity analysis. This selection identified ten U.S. events and 11 events in India for our sensitivity runs.

River depth estimation: To test the effect of our bathymetry estimation method, we used different return periods [1.5-year, 2-year (default in global setup) and 2.5-year] following Andreadis et al. (2013) to estimate bankfull discharge (see Appendix A1) and assess their impact on flood dynamics. We used GloFAS discharge to estimate bankfull discharge for rivers by applying different return periods. For some rivers, bankfull discharge might correspond to a higher or lower return period, so using the same return period for all rivers might not give accurate estimates of river bathymetry and flood extents (Roy and Sinha, 2016; Rad et al., 2024). Therefore, testing multiple return periods provided a better understanding of the river's bathymetry.

Elevation data: Different DEM resolutions (FABDEM 30 m vs 3D Elevation Product 3DEP 1 m) were compared to assess the impact of topographic resolution on flood modelling. Higher-resolution DEMs often provide more accurate flood predictions, particularly in areas with complex terrain (Jiang et al., 2022).

Discharge data: The model's sensitivity to discharge inputs was tested by forcing the model with USGS observation discharge instead of simulated discharge from GloFAS. Thus, the actual event forcing and river channel size estimates were updated based on the observed discharge. Accurate river discharge input is crucial for hydrodynamic modelling, as it directly influences flood predictions (Zhou et al., 2022). This comparison between modelled (GloFAS, default input) and observed (USGS) discharge allowed for disentangling the source of the modelling errors from either the inflow or the processes within SFINCS. USGS River discharge data were filtered for locations containing data for more than 30 years to ensure that return periods and thus bathymetry was more accurately predicted.

Furthermore, to ensure robustness and avoid reliance on a single hydrological model, we also forced SFINCS with the GEB model discharge (Section 2.2.2). This comparison was designed to test the sensitivity of SFINCS to hydrological input characteristics since GloFAS and the GEB model fundamentally differed in model resolution. The GEB model discharges were specifically tested in the Krishna Basin in India (11 riverine events), a large and significant river basin known for its significance due to its size being nearly 8% of the total geographical area of India and extensive agricultural and hydrological use. The GEB discharge outputs were used to drive the flood simulations and derive river channel dimensions. The resulting inundation extents were compared against the Cloud to Street database to assess model accuracy. This dual-model comparison between GloFAS and GEB provided insights into variability and reliability across different hydrological sources with different resolutions, thereby strengthening the analysis of SFINCS flood simulations.

3. Results and Discussion

Here, we present the outcomes of the validation by first describing the general results at the global and continental scales (Section 3.1). To better understand spatial differences in model performance, we also focus specifically on the continental United States (Section 3.1.2). Next, we explore how sensitive our results are to three key inputs: hydrological forcing, bathymetry, and DEM (Section 3.2). Finally, we compare our findings with other similar studies, discuss the model's limitations, and identify areas for future improvement (Section 3.3).

https://doi.org/10.5194/egusphere-2025-4387
Preprint. Discussion started: 6 November 2025

3.1 Global Performance Results

We set up the SFINCS model for 499 riverine flood events worldwide in our validation dataset (Section 2.4.1). These simulations were created using discharge from the global hydrological model GloFAS and validated against satellite-derived flood extents. To calculate the global performance metrics shown in Table 1, we aggregated the results across all simulated events by summing the total number of pixels in each classification category (i.e. Hit, Miss, and FAs) across all events (Section 2.4.2). This analysis gave more weight to larger flood events (with more pixels) by summing all pixels across all events, which avoided biasing the results by treating small and large events equally when calculating the mean.

Table 2 shows that the model achieved an HR of 0.58 for simulated discharge, meaning that 58% of the observed flooded pixels were correctly simulated. The CSI, which balances both Misses and FAs, had a global mean value of 0.39. E was 1.24, which indicated an overestimation of flood extent across events.

Table 2. Event-based performance metrics over 499 events all over the globe, as calculated by comparing SFINCS simulations to the satellite derived flood extents.

	Hit Rate (HR)	False-Alarm Ratio (FAR)	Critical Success Index (CSI)	Error Bias (E)
Global Mean simulated discharge *(All Basins, 499 Events)*	0.58	0.49	0.39	1.24

3.1.1 Effect of Upstream Basin Area on Performance

Table 3 shows considerable variety in the performance metrics across basins with different upstream area sizes. We classified basins into six different upstream area size classes and compared them using the same model runs, hydrological inputs, and observational data as in the above-described global analysis. Most flood events that we simulated in this exercise involved multiple MERIT-BASINS (i.e. sub-basins) that were hydrologically connected. For this analysis, we split the simulated flood event into individual MERIT-BASINS and assigned an upstream area to each basin by determining the contributing drainage area at the subbasin outlet, as defined by Lin et al. (2020). Notably, the total number of basins varied significantly across the upstream area classes. Basins with a large upstream area (≥1,000 km²) were the most represented group (n = 8,834) as compared to basins with a medium-sized upstream area (500 to 1,000 km²; n = 1,533).

Table 3. Performance of the SFINCS flood simulations, classified by basin upstream area (Clipped from the same model runs)

Basin Upstream Area	Hit Rate (HR)	False-Alarm Ratio (FA)	Critical Success Index (CSI)	Error Bias (E)

https://doi.org/10.5194/egusphere-2025-4387
Preprint. Discussion started: 6 November 2025

≥1,000 km² (*n* = 8,834)	0.62	0.41	0.42	1.22
500–1,000 km² (*n* = 1,533)	0.49	0.44	0.33	0.84
100–500 km² (*n* = 6,034)	0.46	0.51	0.31	0.91
50–100 km² (*n* = 4,887)	0.44	0.51	0.30	0.83
<50 km² (*n* = 5,304)	0.42	0.46	0.29	0.61

The analysis showed a clear trend of increasing model performance with increasing upstream basin area. Specifically, the average CSI rose from 0.29 in the basins with a small upstream area (<50 km²) to 0.42 in the basins with a large upstream area (≥1,000 km²). This result aligned with prior studies that have either excluded basins with a small upstream area altogether or reported lower accuracy for them (e.g. Wing et al., 2017; Bernhofen et al., 2018). Three main reasons exist for this scale-dependent behaviour.

1) Firstly, smaller upstream basins often suffer from poorly defined or incomplete river networks (Figure 3a) that result in missing discharge forcing at these locations and can introduce substantial structural errors in the flood extent simulations. For instance, global hydrography datasets like MERIT-SWORD used in this study typically include only river channels with drainage areas larger than 25 km². As a result, key hydrological pathways may be missing or truncated in basins with very small upstream areas, which leads to a significant underestimation of the inundated area. However, even upstream areas exceeding this 25 km² threshold can still contain errors or missing tributaries, which affect the accuracy of the simulated flood extents. This limitation is illustrated in Figure 3a, where a small upstream catchment shows a river stretch with an upstream area of 40.5 km². However, a smaller contributing river upstream is absent from the dataset, which resulted in missed flooding (red pixels) despite clear observational evidence.

2) Secondly, larger upstream basins benefit from the spatial averaging of hydrological processes. As seen in Figures 3b and 3c, mid-sized to large-sized upstream basins integrate runoff contributions from diverse upstream areas and multiple tributaries, which dampens localised anomalies and smooths the overall hydrological signal (Bernhofen et al., 2018; Salinas et al., 2013). These outcomes lead to more predictable and consistent flood responses, especially in models forced by coarsely resolved (~5.5km resolution) inputs like GloFAS (Harrigan et al., 2020).

3) Thirdly, hydrological forcing becomes more uncertain in small basins, where local rainfall-runoff processes dominate and are poorly captured by global-scale models (Smith et al., 2014). Coarse meteorological input data and simplified rainfall-runoff representations are less effective at resolving the fine-scale variability that drives flooding in these areas (Harrigan et al., 2020). Consequently, smaller upstream basins tend to show lower HRs compared to larger ones. Table 3 supports this interpretation. The model performed best in large basins (≥1,000 km²), where it achieved the highest HR (0.62) and lowest FAR (0.41). In contrast, basins under 50 km² showed an HR of 0.42, a FAR of 0.46, and a CSI of just 0.29. Interestingly, these smallest basins also exhibited the lowest error bias (E = 0.61), which suggests that underprediction was more common than overprediction. This underprediction can potentially be reduced by also including a pluvial setup which accounts for direct rainfall-driven flooding within these smaller basins.

https://doi.org/10.5194/egusphere-2025-4387
Preprint. Discussion started: 6 November 2025

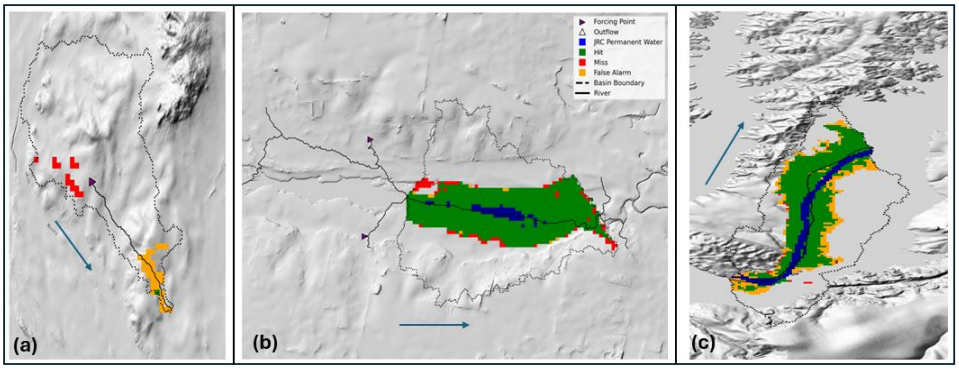


Figure 3. Some example flood events illustrating that with increasing basin sizes, the number of hits increases as well: (a) small upstream basin (40.5 km²) with a river channel (black) that does not connect to a potential flood zone in the north, which results in misses (red area); (b) mid-sized basin 500–1,000km²; (c) large-sized basin of >1,000 km². Flow direction is shown with a blue arrow.

These results highlight the effect of upstream basin area in global flood modelling and the need to better understand the processes in small upstream catchments in large-scale studies. Analysing them separately rather than excluding them, as done in some previous efforts (e.g. Wing et al., 2017), can offer valuable insights into improving model performance in these challenging regions.

3.1.2 Model Performance Across Continental U.S. Basins

Figure 4 shows the spatial distribution of CSI values across all basins within the U.S., where 52 flood events recorded in the Cloud to Street database were analysed. Note that in Figure 4, each MERIT-BASIN was validated individually, as explained in Section 3.1.1. When calculating the mean CSI over the continental US, we considered all pixels of the 52 events modelled using GloFAS discharge. The results showed that the average model performance was slightly higher in the US (CSI = 0.41, including the small basins <50 km²) as compared to the global average (CSI = 0.39). The colour gradient signifies the range of CSI values, with red representing lower scores (CSI ≤ 0.2), orange representing the mid-range (0.2 ≤ CSI ≤ 0.4), and light green (0.4 ≤ CSI ≤ 0.6) and dark green indicating higher scores (CSI ≥ 0.6). Distinct spatial patterns emerged with clusters of bad-performing basins (CSI ≤ 0.2) in two parts of the US (Figures 4a and 4b).

Firstly, we see that in Figure 4a the region of Florida performed consistently low, with many individual basins showing CSI values below 0.2. Florida is a low-lying, flat region, which makes flood simulations particularly challenging (Hawker et al., 2022). Furthermore, we observed that all 45 individual MERIT-BASINS with poor performance corresponded to a single flood event (i.e. DFO_3544), which yielded an average CSI of 0.28 across the affected basins. Potential uncertainty also existed in the DFO_3544 flood type validation (notably, dfo_validation_type = 0, which indicated this event's primary confirmation source as undefined or missing). This means that the event, while classified under "Heavy Rain," could also have been influenced by other flood drivers. The low scores partly resulted from both regional and event-specific limitations, such as DEM errors or challenges in detecting floods with optical remote sensing in flat, low-relief terrain (Tellman et al., 2021).

Secondly, model performance was also generally poor in the Northern Continental U.S. near the Great Lakes (Figure 4b), with multiple basins exhibiting CSI values below 0.3. Notably, only two events (i.e. DFO_2606 and DFO_2412) contributed to the majority of low CSI values in this region. For the dominant event in this region (i.e. DFO_2606), the HR was 0.73, while E = 1.36 and FAR = 0.62 indicated that the low CSI values were driven primarily by overprediction of the flood extent. Further

https://doi.org/10.5194/egusphere-2025-4387
Preprint. Discussion started: 6 November 2025

investigation into DFO_2606 revealed that it corresponded to a severe winter storm that affected the Northeastern US between 5 and 6 January 2005. According to the National Weather Service (2005a), the storm brought heavy snow and freezing rain, followed by power outages and ice damage. Subsequent river flooding was reported in the vicinity of the Great Lakes. However, MODIS satellite data used in the Cloud to Street dataset only captured a brief cloud-free window on 9 January 2005, several days after the event's peak. The MODIS imagery, published by NASA (National Weather Service 2005b), shows swollen rivers like the Ohio, Wabash, and White rivers. However, flooding is difficult to detect in snow-affected, cloudy, or densely vegetated areas (Tellman et al., 2021). Moreover, since MODIS flood extents reflect surface water conditions only during cloud-free acquisitions, the full flood footprint could have been underestimated or misaligned with actual peak inundation.

The third cluster represented good performance (Figure 4c), near the Mississippi River. These basins had an average CSI of 0.43. Notably, the cluster was largely associated with a single event (i.e. DFO_4337), which could suggest this event had good agreement between the observations and simulations. Many of these basins also had large upstream contributing areas, which, as discussed in Section 3.1.1, tend to improve simulation accuracy due to more defined river networks and better-integrated hydrological processes stemming from the hydrological model.

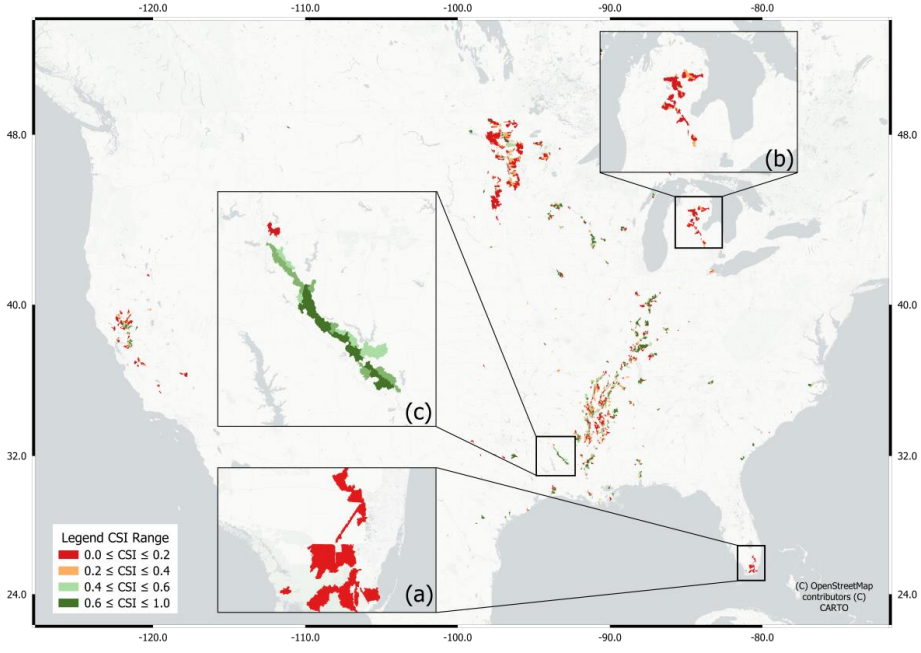


Figure 4. Spatial map with CSI performance numbers for the US. The red colours indicate clusters of low-performing basins; the light and dark green colours represent better-performing basins. The panels zoom in to the following regions: (a) Florida, (b) the Great Lakes, and (c) Mississippi. Base map © OpenStreetMap contributors, rendered with Carto, licensed under ODbL.

3.1.3 Other Factors Influencing Performance

https://doi.org/10.5194/egusphere-2025-4387
Preprint. Discussion started: 6 November 2025

We identified several other factors that considerably affected model performance related to 1) resampling, 2) snapping and channel width in headwater catchments, 3) observed water bodies, and 4) coarse resolution and inflow points.

Resampling: Figure 5a shows that in regions with a good agreement between the SFINCS simulation and the satellite observations (Event DFO_4075; CSI = 0.55), the resampling of a higher-resolution SFINCS output (30 m) to match the observations (250 m) introduced rough edges where the simulation could not match the boundaries of the observed inundation extent. These rough edges produced misses (red areas in Figure 5a) around the boundaries, even when the true flood extent lay just inside the 250 m cell.

Snapping and channel width in headwater catchments: Figure 5b (Event DFO_1996; CSI = 0.48) showed inconsistencies (i.e. FAs) occurring in the headwater rivers. These inconsistencies could have arisen from several factors, such as the snapping procedure (Section 2.1). For example, in headwater locations, the coarse GloFAS grid (~5 km) often misaligned with the finer SFINCS grid, so even with a 5–10 % snapping tolerance (see Section 2.1), small discharge mismatches were introduced into headwater tributaries, which in these narrow channels could produce localised FAs (yellow pixels in Figure 5b) in areas that did not flood. Furthermore, we did not have width observations, so headwater river widths were estimated from the power-law equation (Section 2.2). This approach can sometimes under- or overestimate channel widths in smaller headwater rivers. If the equation gives too-narrow widths, conveyance capacity is underrepresented, which forces excess water onto adjacent floodplains (causing false alarms) (Dey et al., 2022). Conversely, if the equation gives too-wide widths, the simulated flood may be unrealistically confined to the channel and lead to misses.

Observed water bodies: Figure 5c depicts a region (Event DFO_4140; CSI = 0.178) where permanent water was not well represented in the underlying observation layer of the Cloud to Street dataset (i.e. band 5). Although both the model and the satellite-derived flood extent relied on the JRC Global Surface Water dataset as a permanent water mask, the Cloud to Street product provided a 250 m resampled version (Tellman et al., 2021). When rivers meander or curve and are narrower than a MODIS pixel, the resampling process causes permanent open-water bodies to disappear. As a result, channels that are correctly simulated as inundated by SFINCS are not flagged as water in the observation layer and are instead counted as FAs during validation. This effect was particularly visible in low-order rivers where the channel was below 250 m wide. In these cases, the issue was not that the model overpredicted flooding but that the validation mask failed to represent permanent water bodies.

Coarse resolution and inflow points: In Figure 5d, the region's (Event DFO_4451; CSI=0.177) lower performance was due to only one discharge point (i.e. inflow point) carrying water into the model domain. The other discharge points (i.e. headwater points) remained dry (see Appendix A, Fig. A1), which was likely caused by the coarse resolution (~5.5km2) of the GloFAS input discharge data. Thus, the model failed to capture peak flows in smaller upstream tributaries (Alfieri et al., 2013; Grimaldi et al., 2024), which created a significant gap in the simulated flood where the satellite observations clearly showed inundation.

https://doi.org/10.5194/egusphere-2025-4387
Preprint. Discussion started: 6 November 2025

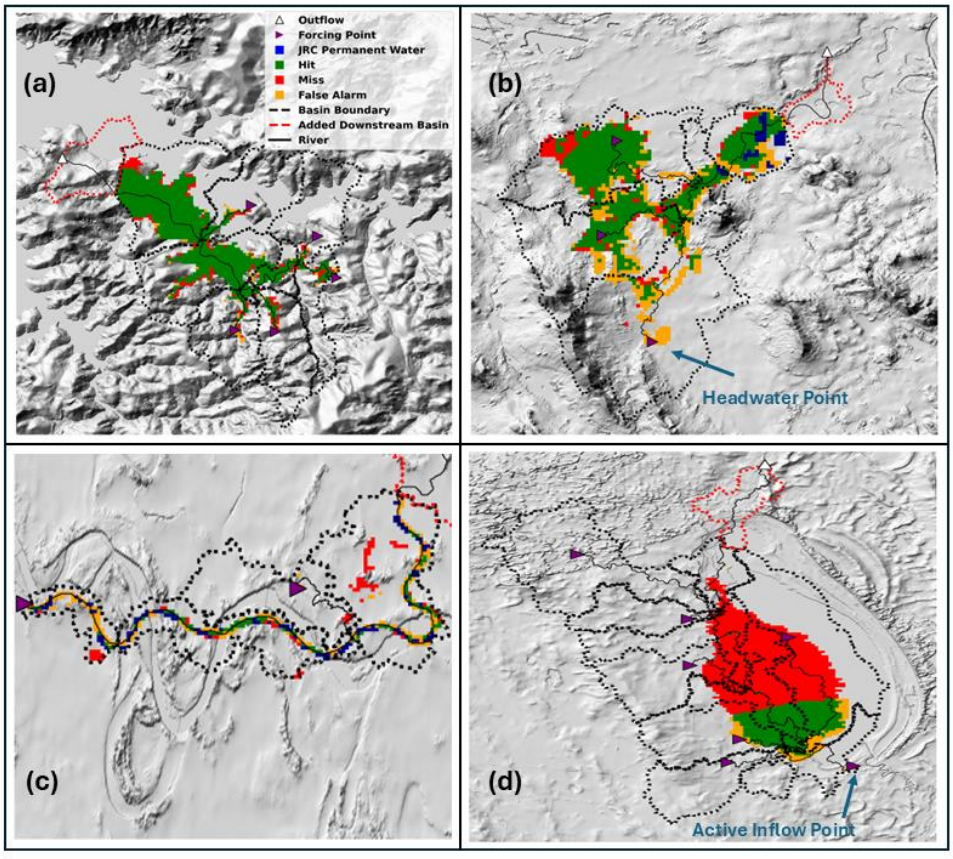


Figure 5. Model simulations across diverse regions illustrating the effect of (a) resampling with some misses on the rough edges of the simulated flood, (b) snapping and channel width, (c) missing observed water bodies, and (d) the influence of coarse resolution of discharge data on missing discharge into inflow points (Appendix A, Fig. A1).

3.2 Sensitivity Analysis

3.2.1 The influence of Hydrological Forcing

We conducted a two-part sensitivity analysis in two global regions (i.e. India and the US) to evaluate the role of hydrological forcing in our global flood modelling framework. We investigated (a) the differences in model performance when using two different global hydrological models (i.e. GloFAS and GEB) for 11 Events in India and (b) the influence of modelled GloFAS discharges compared to observed discharges for ten basins in the US.

(a) Comparing the influence of GloFAS discharge with GEB discharge

https://doi.org/10.5194/egusphere-2025-4387
Preprint. Discussion started: 6 November 2025

To compare the influence of using two different global hydrological models (GLOFAS and GEB) as input to SFINCS, we focused on the Krishna Basin in India. Here, we simulated floods for 11 discrete events with both models. Bathymetric calculations were done separately, using both GloFAS and GEB long-term discharges (>40 years) to compare all aspects of the two models. Using the GloFAS model discharge as forcing for SFINCS resulted in an average CSI of 0.34, while forcing with the GEB model led to an average CSI of 0.38.

One critical distinction between the two models lies in their spatial resolution. The GloFAS operates at a coarser grid resolution (~5.5 km), while the GEB provides discharge data at a finer grid resolution (30″; ~1 km). An example basin cluster within the Krishna basin is shown in Figure 6 (Event 3551). The hydrographs on the right (Figure 6) show the discharge forcings from both hydrological models at six main discharge points of the SFINCS simulation.

Notably, the GEB hydrographs exhibited a more complete flood hydrograph shape, especially toward the later stages of the event (Figure 6). In contrast, the GloFAS hydrographs of some of the inflow points did not capture a high peak flow, suggesting that peak flows could be underrepresented. This issue directly affected the resulting flood maps (Figures 6a and 6b, left panels), where simulations using the GEB input captured more hits than those driven by the GloFAS.

The improved performance of the GEB-SFINCS chain may be attributed to its higher spatial resolution, which allowed for a more accurate representation of smaller headwater rivers and their channel dimensions. In contrast, GloFAS lacks the ability to resolve these smaller streams directly, so bathymetric calculations relied on the nearest available grid point, which might not have aligned with the true upstream location. With the GEB, however, the finer resolution enabled matching bathymetry calculations to the correct upstream grid pixel, which resulted in more accurate low flows and a better representation of channel dimensions in smaller rivers. By resolving finer spatial scales, higher-resolution models like the GEB reduce uncertainties associated with inflow dynamics and provide more reliable input for hydrodynamic models like SFINCS. Moreover, the improved representation of human behaviour and reservoir management in the GEB can lead to more realistic streamflow estimates, which in turn enhances the accuracy of the flood simulations. This outcome underscores the importance of investing in higher-resolution global datasets to improve flood modelling accuracy in river systems.

https://doi.org/10.5194/egusphere-2025-4387
Preprint. Discussion started: 6 November 2025

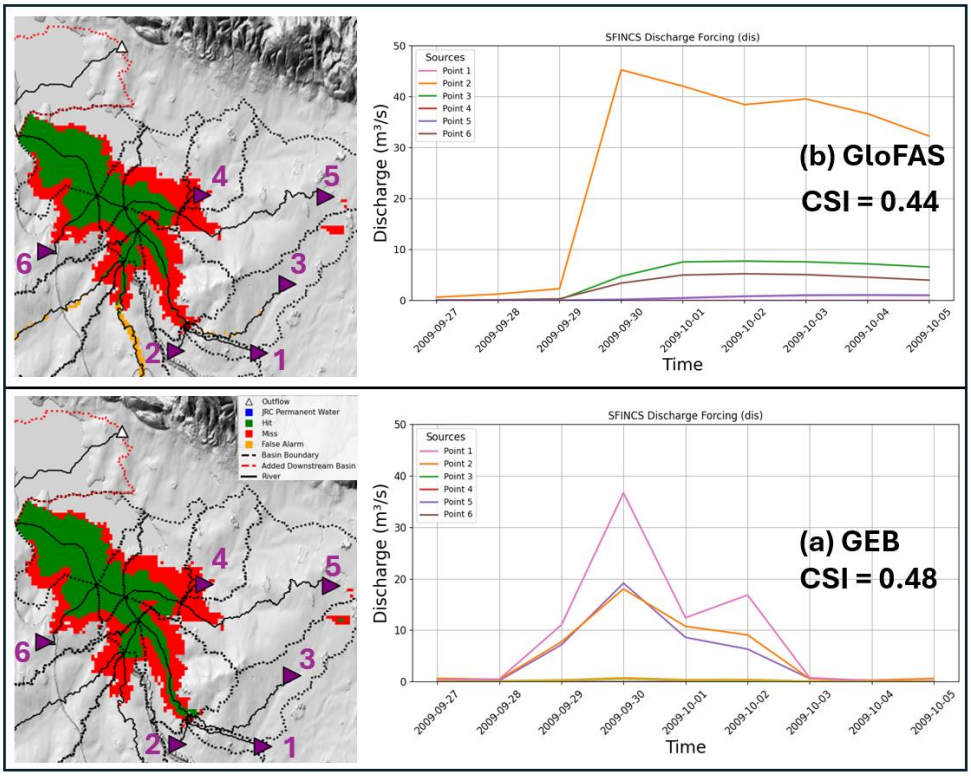


Figure 6. The sensitivity of SFINCS performance to different hydrological forcings by (a) the GEB model and (b) the GloFAS model is shown for a zoomed-in region in the Krishna Basin.

(b) Comparing GloFAS discharge with observed discharge in the US

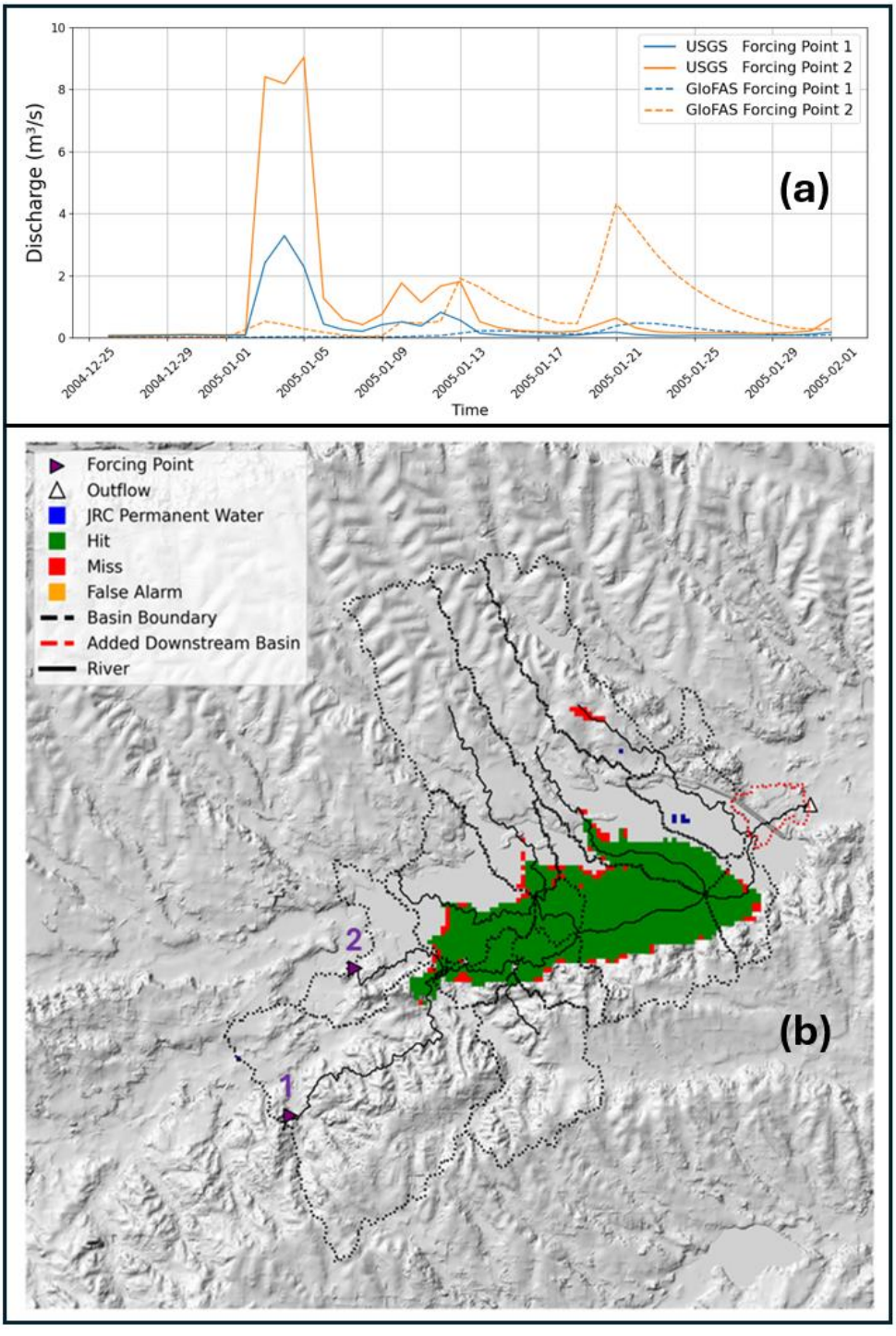
This section describes our assessment of the performance of our modelling setup by comparing modelled discharge from GloFAS with observed discharge data from the USGS (2024). The analysis evaluated model performance using event-based hydrographs from the GloFAS and USGS. Because this analysis could only be applied in regions with long-term discharge records (e.g. for bathymetry estimation), only ten distinct basin clusters remained suitable.

The results revealed that the model using observed USGS discharge data considerably outperformed the one using GloFAS discharge data by achieving an average CSI of 0.67 compared to the average global mean of 0.39 (Table 2). This significant difference shows the importance of accurate discharge data for enhancing model accuracy while highlighting the limitations of global hydrological models. Figures 7a and 7b provide insights into one such high-performing basin forced with USGS observations (CSI = 0.78), while the same basin forced with GloFAS discharge yielded a lower CSI of 0.56. Nonetheless, a considerable number of misses occurred, even when observed discharge data were used. These misses can be explained by the resampling causing the rough edges. Furthermore, in the northeastern part of the basin (Figure 7b), a side stream showed underprediction (i.e. misses), which was likely due to the absence of a river gauge in that tributary. Thus, to better represent this basin, an additional discharge input should be provided to reflect the flood extent accurately. A disadvantage of working with observational river gauges is the potential for gaps in spatial coverage, such as the side stream shown in Figure 7b, which can limit model performance.

https://doi.org/10.5194/egusphere-2025-4387
Preprint. Discussion started: 6 November 2025

Figure 7a shows the hydrograph during the flood event for both modelled (i.e. GloFAS) and observed (i.e. USGS) discharge data. The GloFAS hydrograph failed to capture the peak discharge during a flood event, which contributed to its lower performance (CSI = 0.56) compared to the USGS-based forcing. The inability to capture discharge peaks decreased the precision of the simulated flood extents during high-flow conditions. Harrigan et al. (2020) showed that river discharge is negatively biased in 64% of the basins globally in the GloFAS archive. Moreover, they reported that GloFAS-ERA5 discharge skill, measured by the Kling–Gupta efficiency (KGE), is strongly catchment-size-dependent: median KGE is only 0.21 for basins < 10,000 km² (e.g. Figure 7b basin has a catchment size of ~500 km²), which rises to 0.56 for basins > 50,000 km². This failure to capture discharge peaks in smaller basins has a more immediate and direct impact on flood extent simulations and can reduce the performance of the flood simulations.

https://doi.org/10.5194/egusphere-2025-4387
Preprint. Discussion started: 6 November 2025



695

https://doi.org/10.5194/egusphere-2025-4387
Preprint. Discussion started: 6 November 2025

Figure 7: (a) Hydrograph analysis during the flood event. The dotted line represents the discharge from the GloFAS model, and the solid line represents the observations obtained from USGS. (b) Results from a high-performing U.S. basin (CSI = 0.78) forced with observed discharge input from the USGS. Forcing points 1 and 2 correspond to the hydrograph shown in panel (a).

3.2.2 Bathymetry

A fixed 2-year return period for bathymetric calculations is commonly used as a proxy for bankfull discharge, but this approach may not be universally applicable (Andreadis et al., 2013). Adjusting this parameter based on the characteristics of specific river systems could improve the accuracy of bathymetric estimates and better reflect actual flood behaviour (Roy and Sinha, 2016). Thus, this section presents our evaluation of how model performance was affected when different return periods were used to derive bankfull discharge in the SFINCS fluvial setup. We compared GloFAS-derived discharge with observed USGS river gauge data for ten U.S. basins to assess the accuracy of bathymetry estimates.

The bankfull discharge was calculated via long-term yearly maxima peaks extracted from the GloFAS hydrological model and was evaluated against observed data from USGS river gauges. The results showed that the modelled bankfull discharge (Qbf) was consistently underestimated across all tested return periods.

Figure 8 shows the simulated Qbf (2-year RP) against USGS-derived Qbf on a logarithmic scale. All points lay below the 1:1 line, which indicated that the GloFAS-based RP were systematically lower, particularly in the lower quartile of values, where smaller bankfull discharges were more strongly biased. This pattern indicates that GloFAS underrepresents the frequency and magnitude of the low-flow discharges that influence bankfull conditions. As a result, bathymetry derived from GloFAS tends to yield narrower and shallower channels, which reduces modelled conveyance capacity and potentially produces overestimated flood extents (i.e. higher FAs). These results imply that (i) using the 2-year default alone may not capture observed bankfull behaviour in the current fluvial setup and (ii) alternative approaches (e.g. gauge-based bankfull estimates, where available) can improve bathymetric realism (Zarrabi et al, 2025; Rad et al, 2024). Appendix Figures A3 and A4 present results for the 1.5-year and 2.5-year return periods, respectively.

https://doi.org/10.5194/egusphere-2025-4387
Preprint. Discussion started: 6 November 2025

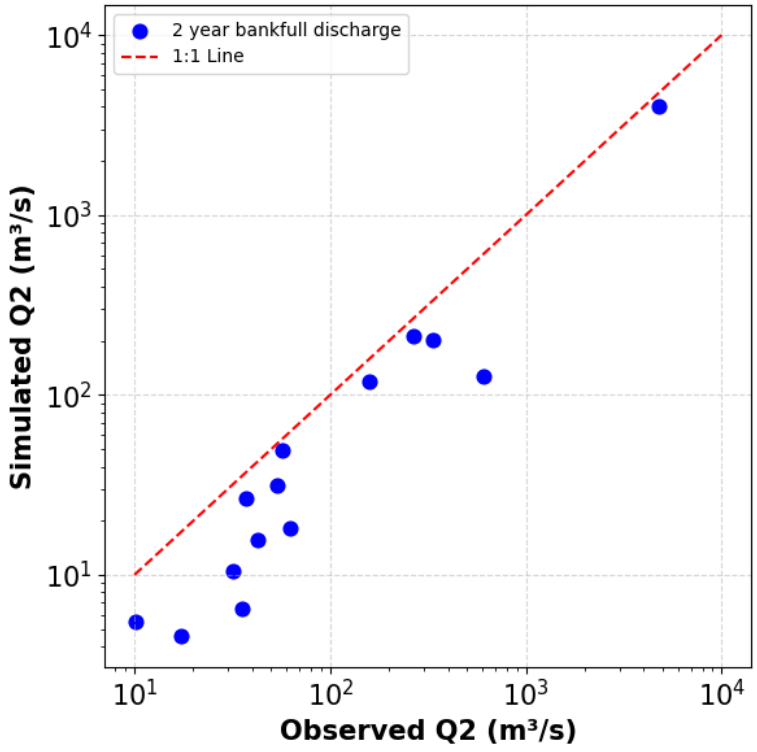


Figure 8. Observed vs modelled 2-year Qbf, for river gauges in ten U.S basins. All points fall below the 1:1 line, which indicates a consistent low bias in the bathymetry.

3.2.3 The influence of the DEM input

This section explores how different DEMs were found to influence the performance of the SFINCS model in simulating flood extents. We focused on two open-source DEMs with different spatial resolutions: FABDEM (30 m; Hawker et al., 2022) and 3D Elevation Program (3DEP; 1 m; USGS, 2015). The resolution of the SFINCS model was kept the same between the two different inputs, although we increased the level of detail (i.e. hypsometry levels, $n = 20$, default 10) stored in the subgrid tables when using the 3DEP DEM. The same ten sub-basins in the United States were selected for this analysis based on the availability of 3DEP. However, vertical inconsistencies were observed in the original 3DEP dataset, where merging adjacent tiles sometimes introduced elevation differences of up to 3 metres. As a result, only six sub-basins were used in the final performance comparison. When using FABDEM, the average CSI was 0.37 across the six sub-basins. The mean CSI increased to 0.57 when using 3DEP DEM, which highlights the positive effect of higher-resolution elevation data on model accuracy. This improvement likely resulted from the finer spatial resolution of 3DEP and higher vertical accuracy. Moreover, 3DEP's 1m LiDAR-derived DEM achieved a root mean square error of 0.53 m (USGS, 2022) compared to FABDEM's mean absolute vertical error reductions from 1.61 m to 1.12 m in built-up areas and from 5.15 m to 2.88 m in forests (Hawker et al., 2022).

Figure 9 illustrates the model's performance with these two DEMs in one of the six basins. Figure 9b shows the FABDEM dataset results, where the CSI was 0.46 for this particular basin. In this case, the lower-resolution DEM resulted in poorly defined river channels, especially in side streams, which prevented water from flowing accurately to downstream areas of the basin. In the upstream part of this

https://doi.org/10.5194/egusphere-2025-4387
Preprint. Discussion started: 6 November 2025

basin, FAs were seen where water accumulated but could not reach the downstream areas. In contrast, when the 3DEP DEM was used (Figure 9a), the model's performance improved significantly. A CSI of 0.59 showed a much better flow representation (i.e. fewer FAs in the upstream part), as the higher-resolution DEM allowed water to reach the downstream parts of the catchment. The increased detail in the terrain allowed for a more accurate representation of floodplains and channels (Jiang et al., 2022), which improved the HR while reducing the number of misses (red pixels) and FAs (yellow pixels).

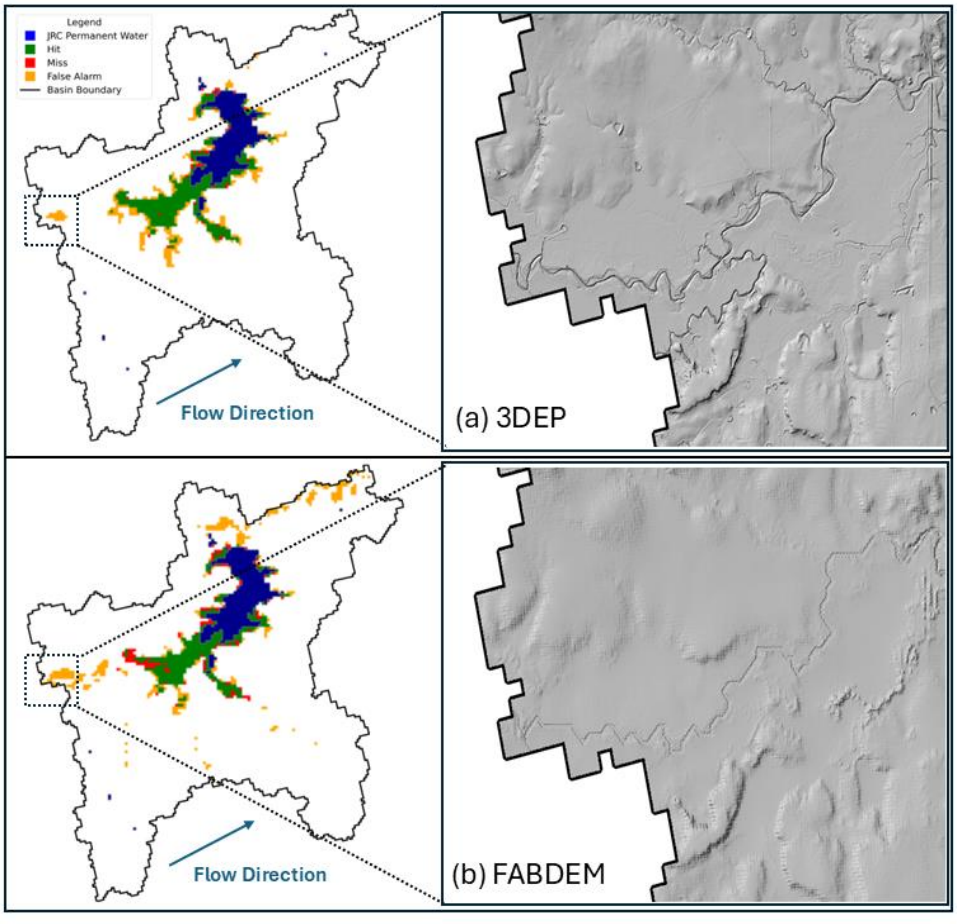


Figure 9. Influence of two DEMs on SFINCS performance: (a) 3DEP DEM of 1 m resolution and (b) FABDEM of 30 m resolution.

3.3 Cross-comparison and limitations

The global mean CSI of 0.39 of this study falls within the broad range reported by other large-scale flood modelling studies, but a clearer picture emerges when we distinguish between those studies using only modelled discharge and those incorporating observed streamflow. Studies driven exclusively by modelled discharges have tended to report lower skill. For example, in a global model intercomparison, Bernhofen et al. (2018) compared six global flood models against three historic African flood events

https://doi.org/10.5194/egusphere-2025-4387
Preprint. Discussion started: 6 November 2025

and found mean CSI values of ~0.4–0.7 for a 25-year flood. Likewise, Mester et al. (2021) used ten different global hydrological models to force the hydrodynamic model CAMAFlood and reported CSI values between 0.3 and 0.5 across eight historic flood events. Similarly, Dottori et al. (2022) used ERA5-driven discharges in LISFLOOD for ten flood events in Europe and obtained CSI values of 0.11–0.28 for high probability small events (<30 yr) and 0.45–0.56 for large extreme events (>500 yr). The SFINCS simulations using GloFAS discharge performed in this study (average CSI = 0.39) align closely with these findings. In contrast, studies that make use of observed discharge data consistently achieved much higher performance. Wing et al. (2017) evaluated LISFLOOD against high water marks in 35 discrete U.S. events using gauge records and reported a CSI of 0.87. Similarly, Hawker et al. (2023) combined field-measured discharges with remote-sensing extents in three Vietnamese case studies and obtained CSI values between 0.37 and 0.62. In this study, when SFINCS was forced with observed USGS discharges, the average CSI was 0.67 (HR = 0.86; FAR = 0.21), which shows the benefit of using accurate discharge inputs in hydrodynamic modelling. While such accuracy is currently best achieved with stream gauge observations, hydrological models capable of producing similarly accurate discharge estimates can deliver comparable results, especially given the sparse spatial coverage of gauges in many regions. Beyond discharge forcing, differences in DEM and bathymetric inputs also play an important role. The sensitivity analyses undertaken in this study show that using higher-resolution elevation data (e.g. 3DEP vs FABDEM) and changing the bankfull return period to local conditions using observed streamflow data can significantly enhance model accuracy.

However, some differences and similarities exist in the setup and findings of our study compared to other large-scale modelling studies.

Number of events: Our study is unique in the high number (n = 499) of events covered, which is much higher than in existing studies covering approximately three to 35 events. Moreover, our study includes small upstream basins (<50 km^2) to address the critical need to understand local upstream floods better. Hence, our model's performance has been assessed across a more diverse hydrological and geographical setting. Notably, we simulated multiple events with different timings and intensities in the same basin, which can be quite challenging to explain performance (Wing et al., 2021).

Hydrological forcing: The accuracy of hydrodynamic flood models (e.g. SFINCS) depends on the quality and characteristics of the hydrological drivers used as input. Previous studies have shown that the choice of hydrological forcing can strongly influence simulated flood extents. For example, a study by Mester et al. (2021) assessed the sensitivity of hydrodynamic models to hydrological forcings using ten different global hydrological models and eight case study areas with observed flood extent values. These basins were selected based on their relatively large size, as it was assumed that the relatively coarse hydrological models would not perform well for smaller sub-basins. Mester et al. found that the agreement between simulated and observed flood extents varied significantly across models and climate forcings. Moreover, Wing et al. (2021) found a CSI of 0.87 but used only observed hydrological forcing for their hydrodynamic model. It is well-known that global hydrological models perform poorly in smaller upstream basins due to their coarse resolution (e.g. Salinas et al., 2013; Mester et al., 2021), which could partly explain the poorer performance of the SFINCS model in these areas. We saw higher performance with SFINCS when it was supplied with observed discharges (CSI = 0.57). The inability of the GloFAS hydrological model to capture the discharge peaks during a flood can significantly impact performance (Figures 5d and 7a). This uncertainty in event discharge significantly impacts flood extent and likely explains much of the reduced CSI (0.39 globally; 0.57 when using USGS observations).

In our current setup, we saw that the bathymetry and river channel dimensions were underestimated (Section 3.2.2), which resulted in higher FAs because channel conveyance capacity was not represented correctly. The SWOT satellite mission provides high-accuracy measurements of water surface elevation, river width, and slope for rivers greater than 100 m in width (Neal et al., 2021; Larnier et al., 2020). Incorporating this dataset may improve flood modelling accuracy, particularly for river systems that have no data. In addition, using the "gradually varying solver" method to estimate river channel

https://doi.org/10.5194/egusphere-2025-4387
Preprint. Discussion started: 6 November 2025

capacity, can replace the traditional Manning's equation. This method can also significantly improve bathymetry and flood inundation simulations (Wing et al., 2024). Neal et al. (2021) demonstrated the benefits of this approach in localised studies, and its application at the global scale may yield more accurate flood predictions.

Buffer: Another difference in our approach is the exclusion of permanent water bodies from the observed and simulated flood extent (250 m resolution), with no additional buffer applied around the rivers. This approach contrasts with methods used in other large-scale validation studies, such as Wing et al. (2017) and Dottori et al. (2022), which applied spatial buffers around the rivers to better capture floodplains and address benchmark data limitations. For example, Wing et al. (2017) applied a fixed buffer (i.e. ~1 km around the rivers) to include areas potentially missed by benchmark data, particularly in small tributaries. Dottori et al. (2022) employed variable buffer zones (i.e. ~5 km and ~10 km) tailored to account for diverse floodplain morphologies and the variable extent and density of the mapped river network. In our validation, we excluded permanent water using satellite-derived flood extent, so correctly simulated inundation in meandering and narrow channels (<250 m wide) was instead counted as an FA, even though these areas should be masked as permanent water. This misclassification inflated the FA ratio while decreasing the HR, which contributed to a lower average global CSI.

Uncertainties in validation data: As Tellman et al. (2021) noted, MODIS often fails to capture floods in rapid flash-flood events or under dense canopy cover, which leads to underestimation of true flood extents and contributes to lower CSI values. Additionally, cloud and snow cover can further obscure floodwaters, which adds to observational uncertainty. Some northern latitudes can have errors greater than 65% in the flood detection algorithm due to the low sun angle on dark soil, which causes low reflectance that mimics water.

The comprehensive validation dataset relies on 250 m MODIS pixels, which limits the detection of narrow or small-scale inundation features and can misclassify flood zone delineation (Landwehr et al., 2024). Moreover, most binary pattern matching metrics are sensitive to the proportion of flooded area, meaning that large-scale floods are favoured while smaller-scale floods are less accurately reflected in the validation scores (Landwehr et al., 2024). Binary flood extent masks and class-based metrics (e.g. the CSI) do not account for important factors such as flood depth and its influence on impacts and model accuracy (Stephens et al., 2014). As a result, some of the poorer performance in smaller basins likely reflects limitations in the reference dataset and validation framework rather than shortcomings in the hydrodynamic model.

Flood protection: Although studies on global flood protection standards exist (e.g. Scussolini et al., 2016), incorporating more accurate data on flood protection standards into future global flood models should be considered. Previous research, including studies by Mester et al. (2021), has shown that flood protection can influence model performance by increasing variability, though it does not necessarily change the maximum performance scores. Further exploration into this area may provide valuable insights into how flood protection measures impact flood risk across regions.

4. Conclusions and Recommendations

Our study evaluated the performance of the SFINCS model in simulating 499 riverine floods globally. Our findings show that the model can simulate riverine flood extents globally, with a mean CSI of 0.39 using GloFAS modelled discharges as input forcing. However, performance improved considerably when using observed discharge inputs. It reached a CSI of 0.67 across ten U.S. events, which highlights the considerable value of in-situ hydrological observations for model accuracy. Furthermore, using a

https://doi.org/10.5194/egusphere-2025-4387
Preprint. Discussion started: 6 November 2025

higher-resolution DEM (3DEP, ~1 m) improved the mean CSI from 0.37 (FABDEM, ~30 m) to 0.57. Conversely, a lower-resolution DEM resulted in poorly defined river channels, especially in side streams, which prevented water from flowing accurately to downstream areas.

The model accuracy substantially improved for larger upstream basin sizes. Specifically, simulations in basins with very large upstream areas (≥1,000 km²) achieved an average CSI of 0.42, whereas those with small upstream areas (<50 km²) had an average CSI of 0.29. Sensitivity analyses further showed that model accuracy was sensitive to the quality of input data. For example, the GEB hydrological model outperformed GloFAS in selected regions, likely due to its finer spatial resolution (~1 km²) and more realistic hydrographs.

Bathymetric calculations revealed a systematic low bias in the default return period estimates used in the global analysis, which are critical for defining realistic channel geometry. This underestimation propagated into narrower and shallower channel representations that reduced conveyance capacity while likely inflating flood extents during high-flow conditions.

These findings suggest that the SFINCS model is highly suitable for modelling river floods. Nevertheless, model accuracy can be improved through targeted enhancements in both hydrological and topographic inputs. Future research can explore methods to regionalise or dynamically calibrate bankfull return periods using observed discharge records where available (e.g. USGS or GRDC) rather than applying a fixed global default. A combined approach that uses long-term gauge data with emerging global remote sensing products (e.g. the SWOT dataset) can allow for more accurate bathymetric representation while reducing systematic biases in large-scale flood simulations.

Our cross-study comparison also confirmed that methodological choices (e.g. DEM resolution), bathymetric assumptions, and masking permanent water affect accuracy and comparability. For example, our decision not to apply spatial buffers or include permanent water bodies may have contributed to more conservative performance metrics, but the choice reflected a stricter and more objective comparison with satellite-derived flood extents. Future research may build on this decision by standardising validation approaches and testing the effect of different masking strategies, such as permanent water removal from reservoirs and the bankfull width of the rivers. The dataset used for validation, 250 m resolution MODIS-based flood maps, cannot detect narrow or meandering channels, particularly in smaller basins. Future large-scale research can prioritise integrating higher-resolution observational datasets such as Sentinel-1 Synthetic Aperture Radar (~10 m), which can capture finer floodplain dynamics and provide real-time images multiple times during the day.

While the CSI provides good insight into the model's performance, it does not capture the vertical accuracy of flood simulations. As Wing et al. (2021) emphasised, incorporating flood depth can offer a more comprehensive evaluation, particularly for identifying errors in floodplain dynamics and understanding biases in inundation extent. Additionally, the upscaling of SFINCS results from 30 m to 250 m resolution (to match our observed data) caused a loss of details that further impacted model performance. The findings of this study echo research that calls for better open-source observed flood validation datasets, such as the Global Flood Monitoring System offered by Copernicus (Bates, 2023). We recommend the development of standardised, large-scale validation frameworks including agreed masking protocols, and using both simulated and observed discharges so that future studies can be directly compared.

Author contributions: TS performed the hydraulic analyses of the models and wrote the paper. TS, VB, and JdB developed and ran the models. JA, TB, and JdB acquired funds and coordinated the project. All authors aided in the conceptualisation of the analysis and contributed to writing and commenting on drafts.

https://doi.org/10.5194/egusphere-2025-4387
Preprint. Discussion started: 6 November 2025

Competing interests: The authors declare that they have no conflict of interest.

Acknowledgements: The authors thank IIASA for their support in GEB simulations and the U.S. Geological Survey for providing open access to data. We kindly acknowledge the IT for Research (ITvO) ADA Linux computational cluster at VU Amsterdam.

Financial support: This project has been funded by the ERC COASTMOVE project nr. 8888442 and the EU CLIMAAX project grant nr. 101093864.

References

Abrams, M., Crippen, R., & Fujisada, H. (2020). ASTER Global Digital Elevation Model (GDEM) and ASTER Global Water Body Dataset (ASTWBD). Remote Sensing, 12(7), 1156. https://doi.org/10.3390/rs12071156

Allen, G. H., & Pavelsky, T. M. (2018). Global River Widths from Landsat (GRWL) Database [Dataset]. In Zenodo (CERN European Organization for Nuclear Research). https://doi.org/10.5281/zenodo.1297434

Altenau, E. H., Pavelsky, T. M., Durand, M. T., Yang, X., De Moraes Frasson, R. P., & Bendezu, L. (2021). The Surface Water and Ocean Topography (SWOT) Mission River Database (SWORD): a global river network for satellite data products. Water Resources Research, 57(7). https://doi.org/10.1029/2021wr030054

Andreadis, K. M., Schumann, G. J., and Pavelsky, T. (2013). A simple global river bankfull width and depth database. Water Resources Research, 49(10), 7164–7168. https://doi.org/10.1002/wrcr.20440

Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., and Pappenberger, F. (2013) GloFAS – global ensemble streamflow forecasting and flood early warning, Hydrol. Earth Syst. Sci., 17, 1161–1175, https://doi.org/10.5194/hess-17-1161-2013, 2013

Apel, H., Benisch, J., Helm, B., Vorogushyn, S., & Merz, B. (2024). Fast urban inundation simulation with RIM2D for flood risk assessment and forecasting. Frontiers in Water, 6. https://doi.org/10.3389/frwa.2024.1310182

Bates, P. D., Horritt, M. S., & Fewtrell, T. J. (2010). A simple inertial formulation of the shallow water equations for efficient two-dimensional flood inundation modelling. Journal of Hydrology, 387(1–2), 33–45. https://doi.org/10.1016/j.jhydrol.2010.03.027

Bates, P. (2023) Fundamental limits to flood inundation modelling. Nature Water 1, 566–567 (2023). https://doi.org/10.1038/s44221-023-00106-4

Benito, I., Jeroen C.J.H. Aerts, Dirk Eilander, Philip J. Ward, and Sanne Muis (2024) Stochastic coastal flood risk modelling for the east coast of Africa. NPJ Natural Hazards (in press)

Bernhofen, M., Whyman Ch., Mark A Trigg, P Andrew Sleigh, Andrew M Smith, Christopher C Sampson, Dai Yamazaki, Philip J Ward, Roberto Rudari, Florian Pappenberger, Francesco Dottori, Peter Salamon, Hessel C Winsemius (2018) A first collective validation of global fluvial flood models for major floods in Nigeria and Mozambique, environmental Research Letters, doi.org/10.1088/1748-9326/aae014

Burek, P., Satoh, Y., Kahil, T., Tang, T., Greve, P., Smilovic, M., Guillaumot, L., Zhao, F., and Wada, Y. (2020) Development of the Community Water Model (CWatM v1.04) – a high-resolution hydrological model for global and regional assessment of integrated water resources management, Geosci. Model Dev., 13, 3267–3298, https://doi.org/10.5194/gmd-13-3267-2020, 2020

Cloud to Street, Microsoft, Radiant Earth Foundation. (2022). A global flood events and cloud cover dataset (Version 1.0). [Date Accessed: October 1th 2024]. Radiant MLHub. https://doi.org/10.34911/rdnt.oz32gz

Copernicus DEM. (2022). [Dataset]. https://doi.org/10.5270/esa-c5d3d65

De Bruijn, J. A., Smilovic, M., Burek, P., Guillaumot, L., Wada, Y., & Aerts, J. C. J. H. (2023). GEB v0.1: a large-scale agent-based socio-hydrological model – simulating 10 million individual farming households in a fully distributed hydrological model. Geoscientific Model Development, 16(9), 2437–2454. https://doi.org/10.5194/gmd-16-2437-2023

Dey, S., Saksena, S., Winter, D., Merwade, V., & McMillan, S. (2022). Incorporating network scale river bathymetry to improve characterization of fluvial processes in flood modeling. Water Resources Research, 58(11). https://doi.org/10.1029/2020wr029521

Deltares: D-Flow Flexible Mesh. Computational Cores and User Interface, User Manual, Deltares, https://www.deltares.nl/en/software/delft3d-flexible-mesh-suite/ (last access: 16 January 2025), 2022.

Dey, S., Siddharth Saksena, Venkatesh Merwade (2019) Assessing the effect of different bathymetric models on hydraulic simulation of rivers in data sparse regions, Journal of Hydrology, 575, pp838-851, https://doi.org/10.1016/j.jhydrol.2019.05.085.

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

Dottori, F., Alfieri, L., Bianchi, A., Skøien, J. O., & Salamon, P. (2022). A new dataset of river flood hazard maps for Europe and the Mediterranean Basin. *Earth System Science Data*, *14*(4), 1549–1569. https://doi.org/10.5194/essd-14-1549-2022

Edwards, P.J., Edward A. Watson, Frederica Wood (2019) Toward a Better Understanding of Recurrence Intervals, Bankfull, and Their Importance. Journal of Contemporary Water Research & Education. https://doi.org/10.1111/j.1936-704X.2019.03300.x

https://doi.org/10.5194/egusphere-2025-4387
Preprint. Discussion started: 6 November 2025

Eilander, D., Boisgontier, H., Bouaziz, L. J. E., Buitink, J., Couasnon, A., Dalmijn, B., Hegnauer, M., De Jong, T., Loos, S., Marth, I., & Van Verseveld, W. (2023a). HydroMT: Automated and reproducible model building and analysis. The Journal of Open Source Software, 8(83), 4897. https://doi.org/10.21105/joss.04897

Eilander, D., Couasnon, A., Leijnse, T., Ikeuchi, H., Yamazaki, D., Muis, S., Dullaart, J., Haag, A., Winsemius, H., & Ward, P. (2023b). A globally applicable framework for compound flood hazard modeling. *Natural Hazards and Earth System Sciences*, *23*(2), 823–846. https://doi.org/10.5194/nhess-23-823-2023

Fleischmann, A., Collischonn, W., Paiva, R., & Tucci, C. E. (2019). Modeling the role of reservoirs versus floodplains on large-scale river hydrodynamics. Natural Hazards, 99(2), 1075–1104. https://doi.org/10.1007/s11069-019-03797-9

Grimaldi, S., G. J.-P. Schumann, A. Shokri, J. P. Walker, V. R. N. Pauwels (2019) Challenges, Opportunities, and Pitfalls for Global Coupled Hydrologic-Hydraulic Modeling of Floods. Water Resources Research. doi.org/10.1029/2018WR024289

Grimaldi, S., Salamon, P., Disperati, J., Zsoter, E., Russo, C., Ramos, A., Carton De Wiart, C., Barnard, C., Hansford, E., Gomes, G. and Prudhomme, C., GloFAS v4.0 hydrological reanalysis, European Commission, 2022, JRC131349.

Guo, K., M. Guan, D. Yu (2021) Urban surface water flood modelling–a comprehensive review of current models and future challenges. Hydrol. Earth Syst. Sci., 25 (5) (2021), pp. 2843-2860

Hall, C. A., Saia, S. M., Popp, A. L., Dogulu, N., Schymanski, S. J., Drost, N., van Emmerik, T., and Hut, R. (2022) A hydrologist's guide to open science, Hydrol. Earth Syst. Sci., 26, 647–664, https://doi.org/10.5194/hess-26-647-2022, 2022.

Harrigan, S., Zsoter, E., Alfieri, L., Prudhomme, C., Salamon, P., Wetterhall, F., Barnard, C., Cloke, H., & Pappenberger, F. (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

Hawker, L., Rougier, J., Neal, J. C., Bates, P. D., Archer, L., and Yamazaki, D. (2018) Implications of Simulating Global Digital Elevation Models for Flood Inundation Studies, Water Resour. Res., 54, 7910–7928, https://doi.org/10.1029/2018WR023279, 2018a.

Hawker, L., Jeffrey Neal, James Savage, Thomas Kirkpatrick, Rachel Lord, Yanos Zylberberg, Andre Groeger, Truong Dang, Thuy, Sean Fox, Felix Agyemang , Pham Khanh Nam (2023) Assessing the next generation of Global Flood Models in the Central Highlands of Vietnam. NHESS, https://doi.org/10.5194/nhess-2023-93

Hoch, J. M., Eilander, D., Ikeuchi, H., Baart, F., & Winsemius, H. C. (2019). Integrating large-scale hydrology and hydrodynamics for nested flood hazard modelling from the mountains to the coast. Natural Hazards and Earth System Sciences. https://doi.org/10.5194/nhess-2019-75

Jiang, W. Jingshan Yu, Qianyang Wang, Qimeng Yue (2022) Understanding the effects of digital elevation model resolution and building treatment for urban flood modelling. Journal of Hydrology SSN 2214-5818, https://doi.org/10.1016/j.ejrh.2022.101122.

Kim, DE., Gourbesville, P. & Liong, SY. Overcoming data scarcity in flood hazard assessment using remote sensing and artificial neural network. Smart Water 4, 2 (2019). https://doi.org/10.1186/s40713-018-0014-5

Landwehr, T., Dasgupta, A., & Waske, B. (2024). Towards robust validation strategies for EO flood maps. Remote Sensing of Environment, 315, 114439. https://doi.org/10.1016/j.rse.2024.114439

Larnier, K., Monnier, J., Garambois, P. A., & Verley, J. (2020). River discharge and bathymetry estimation from SWOT altimetry measurements. Inverse Problems in Science and Engineering, 29(6), 759–789. https://doi.org/10.1080/17415977.2020.1803858

Leijnse, T., Van Ormondt, M., Nederhoff, K., & Van Dongeren, A. (2021). Modeling compound flooding in coastal systems using a computationally efficient reduced-physics solver: Including fluvial, pluvial, tidal, wind- and wave-driven processes. *Coastal Engineering*, *163*, 103796. https://doi.org/10.1016/j.coastaleng.2020.103796

Leopold, L.B. and Maddock, T. (1953) The Hydraulic Geometry of Stream Channels and Some Physiographic Implications. USGS Professional Paper No. 252, 1-57.

Lin, P., Pan, M., Beck, H. E., Yang, Y., Yamazaki, D., Frasson, R., David, C. H., Durand, M., Pavelsky, T. M., Allen, G. H., Gleason, C. J., & Wood, E. F. (2019). Global reconstruction of naturalized river flows at 2.94 million reaches. Water Resources Research, 55(8), 6499–6516. https://doi.org/10.1029/2019wr025287

Liu, K., Song, C., Wang, J., Ke, L., Zhu, Y., Zhu, J., Ma, R., & Luo, Z. (2020). Remote Sensing-Based modeling of the bathymetry and water storage for Channel-Type reservoirs worldwide. Water Resources Research, 56(11). https://doi.org/10.1029/2020wr027147

Masafu, Ch., and Williams, R (2024) Satellite Video Remote Sensing for Flood Model Validation. Water Resources Research . https://doi.org/10.1029/2023WR034545

Mester, B., Sven Norman Willner, Katja Frieler,1 Jacob Schewe (2021) Evaluation of river flood extent simulated with multiple global hydrological models and climate forcings. Environmental Research Letters, 16-9, DOI 10.1088/1748-9326/ac188d

Messager, M. L., Lehner, B., Grill, G., Nedeva, I., & Schmitt, O. (2016). Estimating the volume and age of water stored in global lakes using a geo-statistical approach. Nature Communications, 7(1). https://doi.org/10.1038/ncomms13603

Merz, B., Bloschl, G., Vorogushyn, S., Dottori, F., Aerts, J.C.J.H., Bates, P., Bertola, M., Kemter, M., Kreibich, H., Lall, U. and Macdonald, E. (2021) Causes, impacts and changes of extreme river floods, Nature reviews Earth and Environment, ISSN 2662-138X, 2, 2021, p. 592-609, JRC120240.

Mölder, F., Jablonski, K. P., Letcher, B., Hall, M. B., Tomkins-Tinch, C. H., Sochat, V., Forster, J., Lee, S., Twardziok, S. O., Kanitz, A., Wilm, A., Holtgrewe, M., Nahnsen, S., & Köster, J. (2020). Sustainable data analysis with Snakemake. Zenodo (CERN European Organization for Nuclear Research). https://doi.org/10.5281/zenodo.4063463

Namgyal, T., Dev Anand Thakur, Rishi D.S, Mohit Prakash Mohanty (2023) Are open-source hydrodynamic models efficient in quantifying flood risks over mountainous terrains? An exhaustive analysis over the Hindu-Kush-Himalayan region.

https://doi.org/10.5194/egusphere-2025-4387
Preprint. Discussion started: 6 November 2025

Science of The Total Environment, Volume 897, 2023, 165357, ISSN 0048-9697, https://doi.org/10.1016/j.scitotenv.2023.165357.

National Weather Service. (2005a, January 6). Winter storm summary for January 05, 2005 to January 6, 2005 event. https://www.weather.gov/phi/01052005wss

National Weather Service. (2005b, January 6). January 5–6, 2005 Ice Storm and Flooding. https://www.weather.gov/iln/20050105

Nederhoff, K., Crosby, S. C., Van Arendonk, N. R., Grossman, E. E., Tehranirad, B., Leijnse, T., Klessens, W., & Barnard, P. L. (2024). Dynamic modeling of coastal compound flooding hazards due to tides, extratropical storms, waves, and Sea-Level rise: a case study in the Salish Sea, Washington (USA). *Water*, *16*(2), 346. https://doi.org/10.3390/w16020346

Neal, J., Laurence Hawker, James Savage, Michael Durand, Paul Bates, Christopher Sampson (2021) Estimating River Channel Bathymetry in Large Scale Flood Inundation Models, Water Resources Research. https://doi.org/10.1029/2020WR028301

Nevo, S., Morin, E., Rosenthal, A. G., Metzger, A., Barshai, C., Weitzner, D., Voloshin, D., Kratzert, F., Elidan, G., Dror, G., Begelman, G., Nearing, G., Shalev, G., Noga, H., Shavitt, I., Yuklea, L., Royz, M., Giladi, N., Levi, N. P., . . . Matias, Y. (2022). Flood forecasting with machine learning models in an operational framework. Hydrology and Earth System Sciences, 26(15), 4013–4032. https://doi.org/10.5194/hess-26-4013-2022

Pekel, J., Cottam, A., Gorelick, N., & Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes. Nature, 540(7633), 418–422. https://doi.org/10.1038/nature20584

Rad, A. M., Johnson, J. M., Ghahremani, Z., Coll, J., & Frazier, N. (2024). Enhancing River Channel Dimension estimation: A machine learning approach leveraging the national water model, hydrographic networks, and landscape characteristics. Journal of Geophysical Research Machine Learning and Computation, 1(4). https://doi.org/10.1029/2024jh000173

Roy, N. G., & Sinha, R. (2016). Linking hydrology and sediment dynamics of large alluvial rivers to landscape diversity in the Ganga dispersal system, India. Earth Surface Processes and Landforms, 42(7), 1078–1091. https://doi.org/10.1002/esp.4074

Salinas, J. L., Laaha, G., Rogger, M., Parajka, J., Viglione, A., Sivapalan, M., & Blöschl, G. (2013). Comparative assessment of predictions in ungauged basins – Part 2: Flood and low flow studies. Hydrology and Earth System Sciences, 17(7), 2637–2652. https://doi.org/10.5194/hess-17-2637-2013

Sampson, C., Smith, A., Bates, P., Neal, J., Alfieri, L., & Freer, J. E. (2015). A high-resolution global flood hazard model. *Water Resources Research*, *51*(9), 7358–7381. https://doi.org/10.1002/2015wr016954

Scussolini, P., Aerts, J. C. J. H., Jongman, B., Bouwer, L. M., Winsemius, H. C., de Moel, H., and Ward, P. J (2016) FLOPROS: an evolving global database of flood protection standards, Nat. Hazards Earth Syst. Sci., 16, 1049–1061, https://doi.org/10.5194/nhess-16-1049-2016, 2016.

Sebastian, A., Bader, D.J., Nederhoff, C.M. Leijnse, T., Bricker, J., Aarninkhof, A. (2021) Hindcast of pluvial, fluvial, and coastal flood damage in Houston, Texas during Hurricane Harvey (2017) using SFINCS. Nat Hazards 109, 2343–2362 (2021). https://doi-org.vu-nl.idm.oclc.org/10.1007/s11069-021-04922-3

Shaw, J., Georges Kesserwani, Jeffrey Neal, Paul Bates, and Mohammad Kazem Sharifian (2021) LISFLOOD-FP 8.0: the new discontinuous Galerkin shallow-water solver for multi-core CPUs and GPUs. Geosci. Model Dev., 14, 3577–3602, 2021. https://doi.org/10.5194/gmd-14-3577-2021

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

Sutanudjaja, E. H., van Beek, R., Wanders, N., Wada, Y., Bosmans, J. H. C., Drost, N., van der Ent, R. J., de Graaf, I. E. M., Hoch, J. M., de Jong, K., Karssenberg, D., López López, P., Peßenteiner, S., Schmitz, O., Straatsma, M. W., Vannametee, E., Wisser, D., and Bierkens, M. F. P. (2018) PCR-GLOBWB 2: a 5 arcmin global hydrological and water resources model, Geosci. Model Dev., 11, 2429–2453, https://doi.org/10.5194/gmd-11-2429-2018, 2018.

Scussolini, P., Aerts, J. C. J. H., Jongman, B., Bouwer, L. M., Winsemius, H. C., de Moel, H., and Ward, P. J.: FLOPROS: an evolving global database of flood protection standards, Nat. Hazards Earth Syst. Sci., 16, 1049–1061, https://doi.org/10.5194/nhess-16-1049-2016, 2016.

Teng, J., A.J. Jakeman, J. Vaze, B.F.W. Croke, D. Dutta, S. Kim (2017) Flood inundation modelling: A review of methods, recent advances and uncertainty analysis, Environmental Modelling & Software, 90, pp201-216, https://doi.org/10.1016/j.envsoft.2017.01.006.

Ward, Ph., Brenden Jongman, Frederiek Sperna Weiland, Arno Bouwman, Rens van Beek, Marc F P Bierkens, Willem Ligtvoet and Hessel C Winsemius (2013) Assessing flood risk at the global scale: model setup, results, and sensitivity. Environ. Res. Lett. 8 044019, DOI 10.1088/1748-9326/8/4/044019

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

Teng, J., A.J. Jakeman, J. Vaze, B.F. Croke, D. Dutta, S.J.E.M. Kim (2017) Flood inundation modelling: a review of methods, recent advances and uncertainty analysis Environ. Model. Softw., 90 (2017), pp. 201-216

USGS (2015) 3D Elevation Program (3DEP) 1-meter Digital Elevation Model. U.S. Department of the Interior. https://www.usgs.gov/3d-elevation-program/about-3dep-products-services

USGS (2023) USGS current water data for the nation. https://waterdata.usgs.gov/nwis/rt

USGS (2022.) What is the vertical accuracy of 3D Elevation Program (3DEP) DEMs? U.S. Geological Survey. Accessed 28 July 2025. https://www.usgs.gov/faqs/what-vertical-accuracy-3d-elevation-program-3dep-dems

Van Ormondt, M., Leijnse, T., De Goede, R., Nederhoff, K., & Van Dongeren, A. (2025). Subgrid corrections for the linear inertial equations of a compound flood model – a case study using SFINCS 2.1.1 Dollerup release. Geoscientific Model Development, 18(3), 843–861. https://doi.org/10.5194/gmd-18-843-2025

https://doi.org/10.5194/egusphere-2025-4387
Preprint. Discussion started: 6 November 2025

Wilkerson, G. (2008) Improved Bankfull Discharge Prediction Using 2-Year Recurrence-Period Discharge. J. of the American Water Resources Association. https://doi.org/10.1111/j.1752-1688.2007.00151.x

Wing, O. E. J., Bates, P. D., Sampson, C. C., Smith, A. M., Johnson, K. A., & Erickson, T. A. (2017). Validation of a 30 m resolution flood hazard model of the conterminous United States. Water Resources Research, 53(9), 7968–7986. https://doi.org/10.1002/2017wr020917

Wing, O. E. J., Bates, P. D., Neal, J. C., Sampson, C. C., Smith, A. M., Quinn, N., Shustikova, I., Domeneghetti, A., Gilles, D. W., Goska, R., and Krajewski, W. F. (2019) A New Automated Method for Improved Flood Defense Representation in Large-Scale Hydraulic Models, Water Resour. Res., 55, 11007–11034, https://doi.org/10.1029/2019WR025957, 2019.

Wing, O., Quinn, N., Bates, P., Neal, J., Smith, A., Sampson, C., Coxon, G., Yamazaki, D., Sutanudjaja, E. H., & Alfieri, L. (2020). Toward global stochastic river flood modeling. *Water Resources Research*, *56*(8). https://doi.org/10.1029/2020wr027692

Wing, O., Smith, A., Marston, M. L., Porter, J. R., Amodeo, M., Sampson, C., & Bates, P. (2021). Simulating historical flood events at the continental scale: observational validation of a large-scale hydrodynamic model. *Natural Hazards and Earth System Sciences*, *21*(2), 559 –575. https://doi.org/10.5194/nhess-21-559-2021

Wing, O., Paul D. Bates, Niall D. Quinn, James T. S. Savage, Peter F. Uhe, Anthony Cooper, Thomas P. Collings, Nans Addor, Natalie S. Lord, Simbi Hatchard, Jannis M. Hoch, Joe Bates, Izzy Probyn, Sam Himsworth, Josué Rodríguez González, Malcolm P. Brine, Hamish Wilkinson, Christopher C. Sampson, Andrew M. Smith, Jeffrey C. Neal, Ivan D. Haigh (2024) A 30 m Global Flood Inundation Model for Any Climate Scenario. Water Resources Research. https://doi.org/10.1029/2023WR036460

Winsemius, H. C., Aerts, J. C. J. H., Van Beek, L. P. H., Bierkens, M. F. P., Bouwman, A., Jongman, B., Kwadijk, J. C. J., Ligtvoet, W., Lucas, P. L., Van Vuuren, D. P., & Ward, P. J. (2015). Global drivers of future river flood risk. Nature Climate Change, 6(4), 381–385. https://doi.org/10.1038/nclimate2893

Yamazaki, D., Ikeshima, D., Sosa, J., Bates, P. D., Allen, G. H., & Pavelsky, T. M. (2019). MERIT Hydro: A High-Resolution Global Hydrography Map based on latest Topography Dataset. Water Resources Research, 55(6), 5053–5073. https://doi.org/10.1029/2019wr024873

Zarrabi, R., McDermott, R., Erfani, S. M. H., & Cohen, S. (2025). Bankfull and Mean-Flow channel geometry estimation through machine learning algorithms across the CONtiguous United States (CONUS). Water Resources Research, 61(2). https://doi.org/10.1029/2024wr037997

Zeiger, S.J., and Hubbart, J.A. (2021) Measuring and modeling event-based environmental flows: an assessment of HEC-RAS 2D rain-on-grid simulations J. Environ. Manag., 285 (2021), p. 112125

Zhou, X., Revel, M., Modi, P., Shiozawa, T., & Yamazaki, D. (2022). Correction of river bathymetry parameters using the Stage–Discharge Rating curve. Water Resources Research, 58(4). https://doi.org/10.1029/2021wr031226

https://doi.org/10.5194/egusphere-2025-4387
Preprint. Discussion started: 6 November 2025

Appendices

A1 Bathymetry

Bathymetry, the measurement of the depth and width of rivers, is one of the most challenging variables to estimate on a global scale due to the lack of comprehensive, high-resolution data, particularly in remote or unsurveyed regions (Dey et al., 2019). Accurate representation of river bathymetry is critical for hydrodynamic modelling, as it directly influences the conveyance capacity of rivers and channels, which affects flood inundation predictions (Wing et al., 2024). We simulated bathymetry with the following steps.

Step 1. Matching River Centerline and Discharge Data

The process began by clipping the river centerline vector dataset (MERIT-SWORD) to the model domain (flooded MERIT-BASINS + additional downstream basin) with a small buffer. Then, we matched the individual river segments to the global hydrological model GloFAS using upstream area data. The matching used the origin points of each river segment, which were snapped to a particular upstream area pixel in the GloFAS dataset using the HydroMT package. When snapping points to the discharge grid, this package considers a 5% relative error tolerance and a 50 km² absolute error cap. The upstream area was a key indicator of river flow and helped to identify the centerline segments corresponding to different river sections. This step ensured that discharge calculations were correctly linked to the appropriate locations on the river. Once the river segments were matched, we extracted discharge values focused on the yearly discharge maxima from 1979 to 2024. These maxima formed the basis for our return-period analysis via the block-maxima method, which selected the single largest discharge value each year (45 years).

Step 2. Utilising PyExtremes for Distribution Analysis of Discharge Data

Next, we used the *PyExtremes* Python package (https://georgebv.github.io/pyextremes/) to fit the yearly discharge maxima to an appropriate distribution. PyExtremes helped to determine the best fit for the maximum discharge values, which might follow different distributions for extreme value analysis. The distribution selection was based on the characteristics of the discharge data (from GloFAS) to ensure that we accurately captured the probability of various discharge levels. Using these fitted distributions, we calculated discharge values for different return periods. We used a 2-year return period discharge, a typical proxy for bankfull discharge (e.g. Wilkerson, 2008). Bankfull discharge refers to the flow level at which the river is filled to the top of its banks without overflowing, a key indicator for estimating the river's cross-sectional shape. To compute return periods, the package ranks extreme values, calculates exceedance probabilities, and derives return periods as multiples of a specified return period size (typically one year). This systematic approach allowed it to assign empirical return periods to the extreme values extracted from our discharge data.

Step 3. Estimating Bankfull Width and Depth

Only global river segments 30 m wide and greater are represented in the SWORD database (Altenau et al., 2021). Thus, the river centerline vector dataset MERIT-SWORD is missing river-width values for river segments smaller than 30 m (https://zenodo.org/records/14675925). To estimate the bankfull width of the rivers (<30m), we applied a power-law relationship that linked discharge to channel width, based on empirical studies (Leopold and Maddock, 1953). The power-law formula allowed us to predict the width based on the bankfull discharge calculated earlier.

$$W = a \cdot Q^b$$

https://doi.org/10.5194/egusphere-2025-4387
Preprint. Discussion started: 6 November 2025

Where:

- *W is* the bankfull width,
- Q is the bankfull discharge (from the 2-year return period),
- a & b are empirically derived constants

Once the width was estimated, we proceeded to calculate the bankfull depth for all the river segments in the model domain. Depth was estimated by applying Manning's equation for open channel flow, which relates the river's flow velocity, roughness, and channel geometry (including slope and depth) to its discharge. Manning's equation is expressed as follows:

$$Q = \frac{1}{(n)} \cdot AR^{\{\left(\frac{2}{3}\right)} \cdot S^{\{\left(\frac{1}{2}\right)}$$

Where:

- Q is the discharge,
- n is the Manning's roughness coefficient,
- A is the cross-sectional area,
- R is the hydraulic radius (cross-sectional area divided by wetted perimeter),
- S is the slope of the river.

A2: Hydro-MT and Snakemake

A2.1 Hydro-MT (Eilander et al., 2023a)

HydroMT is an open-source Python package designed to streamline the process of building and configuring water system models (e.g. SFINCS). The framework handles various spatial data types, including gridded raster data like DEMs and vector data like shapefiles. It automates essential steps in the preprocessing, setup, and postprocessing of hydrodynamic models, so it is a versatile tool for hydrological modelling.

The HydroMT-SFINCS sub-package, specifically tailored for the SFINCS model, extends the core HydroMT functionalities to address the specific needs of SFINCS users. One key function of HydroMT-SFINCS is the configuration of boundary conditions, which includes integrating inflow and outflow boundary conditions for rivers, setting up forcing conditions (e.g. hydrographs and precipitation), and defining external water level boundaries. The sub-package also generates a mask for active and inactive cells based on the basin boundary. Thus, we ensured that our calculations were carried out only in relevant flood-prone areas while excluding regions unaffected by flooding.

HydroMT-SFINCS is also adept at handling and converting input data from various formats (e.g. NetCDF, GeoTIFF, and shapefiles) into the structure required by the model, which includes preparing static input layers (e.g. DEMs, land-use maps, and basin boundaries) in a unified format. The tool ensures spatial consistency across different datasets through automated reprojection, resampling, and cropping to minimise potential errors that can arise from manual data processing.

Regarding postprocessing, HydroMT-SFINCS automated the interpolation of SFINCS output, which was stored as NetCDF files containing water levels for each subgrid cell. For each time step, these water

https://doi.org/10.5194/egusphere-2025-4387
Preprint. Discussion started: 6 November 2025

levels were interpolated onto the DEM data to produce water depths, which were then used to generate the maximum flood extent. By automating the pre- and postprocessing steps, HydroMT-SFINCS minimised the risk of errors and significantly accelerated the model setup and output generation process. This approach enabled the rapid testing of various model configurations and ensured reproducibility, which is especially important for large-scale and scenario-based modelling studies.

A2.2 Snakemake (Mölder et al., 2020)

Snakemake is a workflow management system designed to handle complex data analysis pipelines in a reproducible and scalable manner. It is particularly useful in scientific computing, where numerous tasks and steps must be executed in a specific order to ensure the efficient processing of large datasets. Snakemake uses a simple syntax to define rules for data processing, where each rule specifies input files, output files, and the command to execute.

In our workflow, Snakemake was used to automate and streamline the processing of all flood modelling tasks, from preprocessing to validation. By organising each task into separate rules, Snakemake enabled the reproducibility of the entire process. The flexibility of Snakemake allowed us to easily parallelise tasks, for example, by running multiple model domains (i.e. clusters) for the same event simultaneously on different computing nodes. This ability expedited the analysis while making it easier to handle large-scale simulations. Hence, the results were generated in a timely manner. By integrating Snakemake into our modelling pipeline, we ensured that each step of the analysis, from data preprocessing to final flood map and validation, was reproducible and easily adjustable for future adjustments or expansions of the study.

Additional Figures:

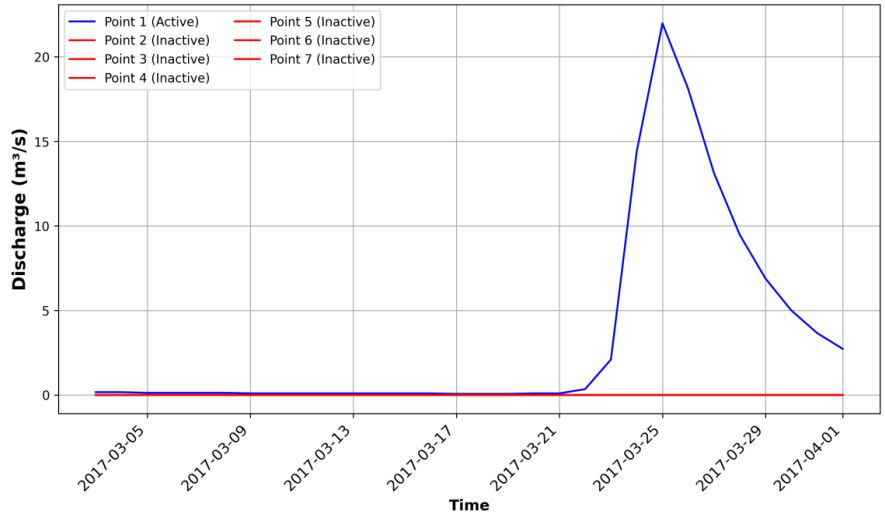


Figure A1: Section 3.1.1 figure 5d supporting argument figure

https://doi.org/10.5194/egusphere-2025-4387
Preprint. Discussion started: 6 November 2025

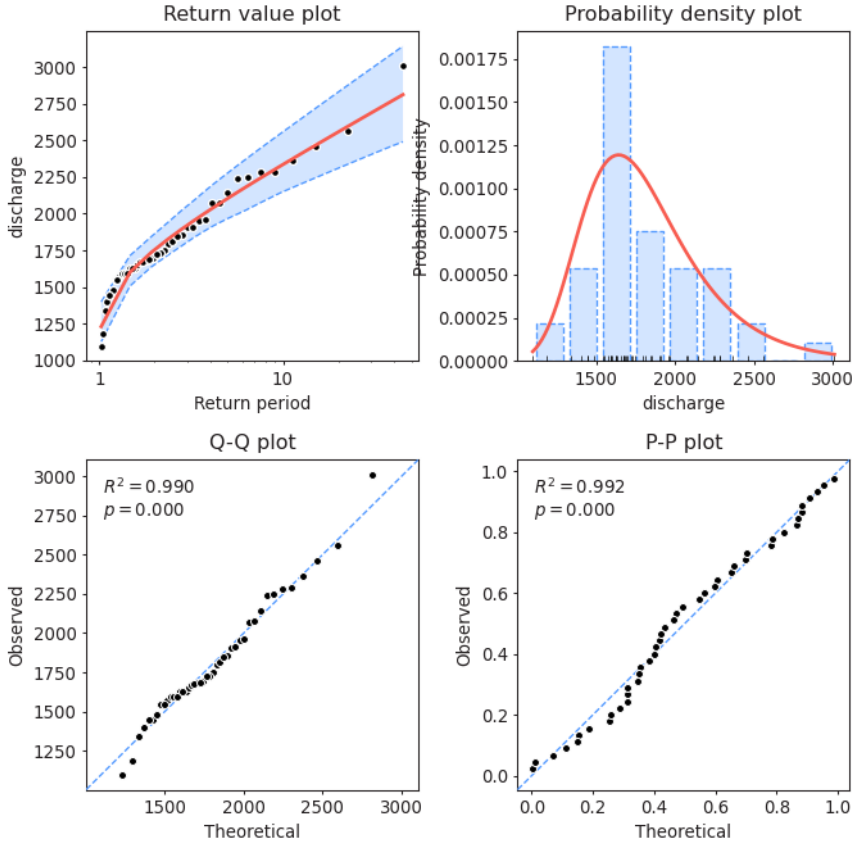


Figure A2: Example return period calculations “PyExtremes”, which were conducted for every river segment in the MERIT-BASINS dataset

https://doi.org/10.5194/egusphere-2025-4387
Preprint. Discussion started: 6 November 2025

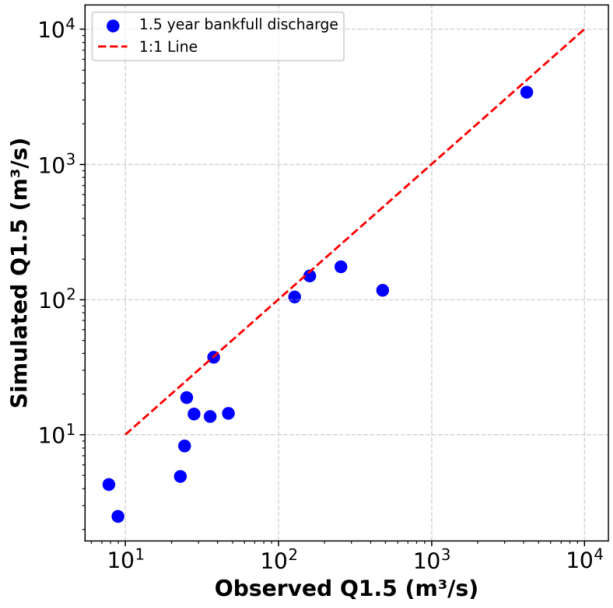


Figure A3. Observed vs modeled 1.5-year Qbf, for river gauges in ten U.S basins.

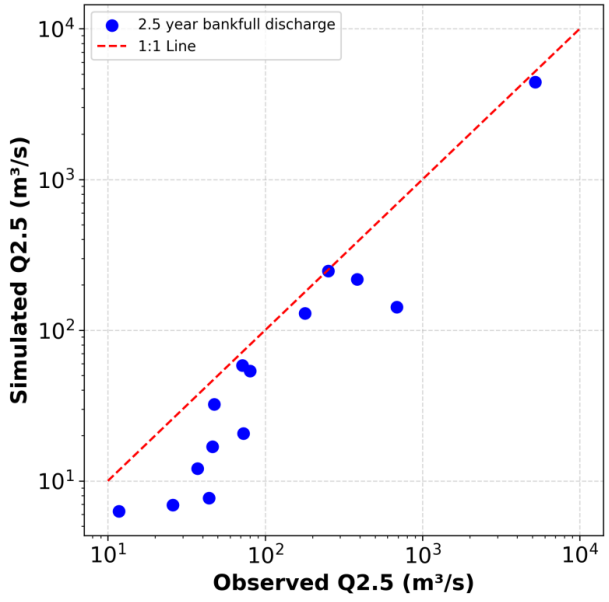


Figure A4. Observed vs modeled 2.5-year Qbf, for river gauges in ten U.S basins.

https://doi.org/10.5194/egusphere-2025-4387
Preprint. Discussion started: 6 November 2025

Data Availability statement:

The version 2.2.0 of the SFINCS model used in this study can be found on docker (https://hub.docker.com/r/deltares/sfincs-cpu/tags, last access: July 25 2025). The framework, list of modelled events and python scripts, are accessible through Zenodo (https://doi.org/10.5281/zenodo.16759099).

1. USGS Discharge:
   https://waterdata.usgs.gov/nwis
2. FABDEM: V1-2
   https://data.bris.ac.uk/data/dataset/s5hqmjcdj8yo2ibzi9b4ew3sn
3. ESA Landcover
   https://worldcover2021.esa.int/
4. MERIT-BASINS Dataset:
   https://www.reachhydro.org/home/params/merit-basins
5. HydroBasins:
   HydroBASINS (hydrosheds.org)
6. GloFAS: v4.0
   https://cds.climate.copernicus.eu/cdsapp#!/dataset/cems-glofas-historical?tab=doc
7. GloFAS upstream area: v4.0
   https://cds.climate.copernicus.eu/cdsapp#!/dataset/10.24381/cds.ff1aef77?tab=overview
8. MERIT-SWORD River Vector Dataset:
   https://zenodo.org/records/14675925
9. GEB Model:

   https://github.com/GEB-model