Multivariable Integrated Evaluation of Hydrodynamic Modeling: A Comparison of Performance considering different Baseline Topography Data

PRAKAT MODI[1], Menaka Revel[2], and Dai Yamazaki[2]

[1]The University of Tokyo
[2]University of Tokyo

November 24, 2022

Abstract

Continental-scale river hydrodynamic modeling is useful for understanding the global hydrological cycle, and model evaluation is essential for robust calibration and assessing model performance. Although many models have been robustly evaluated using several variables separately, methods for the integrated multivariable evaluation of models have yet to be established. Here, we propose an evaluation method using the overall basin skill score (OSK), based on considering the spatial distribution of different variables via a sub-basin approach. The OSK approach integrates multiple variables to overcome observation-related limitations, such as the distinct temporal and spatial dimensions and unit of measurement unique to each variable, thus judging model performance objectively at the sub-basin and basin scales. As a case study, the global river model, CaMa-Flood, was evaluated using three variables¾discharge, water surface elevation, and flooded area¾for the Amazon Basin, focusing on the impact of using different types of baseline topography data (SRTM and MERIT digital elevation models [DEMs]). CaMa-Flood with the MERIT DEM performed robustly well over a wide range of river depth parameters with a maximum OSK of 0.51 against 0.46 for the SRTM DEM. Single-variable evaluation for all three variables proved inadequate due to low sensitivity for river bathymetry, with good performance outcomes potentially arising for the wrong reasons. This study confirmed that model evaluation using this method enables a balanced evaluation of different variables and a robust estimation of the best parameter set. The proposed method proved useful for flexible, integrated multivariable model evaluation, with modifications allowed per the user's requirements.

Multivariable Integrated Evaluation of Hydrodynamic Modeling: A Comparison of Performance considering different Baseline Topography Data

Prakat Modi[1], Menaka Revel[2], and Dai Yamazaki[2]

[1] Department of Civil Engineering, The University of Tokyo, Tokyo, Japan

[2] Institute of Industrial Science, The University of Tokyo, Tokyo, Japan

Corresponding Author: Prakat Modi (modi@rainbow.iis.u-tokyo.ac.jp)

Key Points:

- Integrated Multivariable Evaluation of Hydrodynamic Model using subbasin approach.
- Performance metric combining variables with distinct temporal and spatial dimensions and measurement units.
- Multivariable evaluation suggested MERIT DEM performs consistently better in hydrodynamic modelling

Abstract

Continental-scale river hydrodynamic modeling is useful for understanding the global hydrological cycle, and model evaluation is essential for robust calibration and assessing model performance. Although many models have been robustly evaluated using several variables separately, methods for the integrated multivariable evaluation of models have yet to be established. Here, we propose an evaluation method using the overall basin skill score (OSK), based on considering the spatial distribution of different variables via a sub-basin approach. The OSK approach integrates multiple variables to overcome observation-related limitations, such as the distinct temporal and spatial dimensions and unit of measurement unique to each variable, thus judging model performance objectively at the sub-basin and basin scales. As a case study, the global river model, CaMa-Flood, was evaluated using three variables—discharge, water surface elevation, and flooded area—for the Amazon Basin, focusing on the impact of using different types of baseline topography data (SRTM and MERIT digital elevation models [DEMs]). CaMa-Flood with the MERIT DEM performed robustly well over a wide range of river depth parameters with a maximum OSK of 0.51 against 0.46 for the SRTM DEM. Single-variable evaluation for all three variables proved inadequate due to low sensitivity for river bathymetry, with good performance outcomes potentially arising for the wrong reasons. This study confirmed that model evaluation using this method enables a balanced evaluation of different variables and a robust estimation of the best parameter set. The proposed method proved useful for flexible, integrated multivariable model evaluation, with modifications allowed per the user's requirements.

Plain Language Summary

The hydrodynamic model is an import tool to understand the many natural phenomena related to the water cycle. Its accuracy depends on the many input data, like runoff and topography. Accurate representation of the natural system using the hydrodynamic model can be judged based on the models' performance. The model's performance depends on the accurate estimation of variables and judged based on the evaluation metric's score. Earlier, many studies focused on either a single variable or multiple variables considering separately for model evaluation. Here we proposed a method, integrating multiple variables by combining each variable's performance into a single overall basin skill score for comparing two topography data. The proposed method is the first of its kind, integrating variables with different temporal and spatial dimensions

and measurement units for performance evaluation. The method has a significant advantage of combining different variables for robust evaluation.

1 Introduction

Continental-scale river hydrodynamic modeling is essential for understanding the global hydrological cycle and supporting flood monitoring as well as water resource management with regard to water security and natural hazards (Siqueira et al., 2018). Hydrodynamic modeling processes depend on topographic data to replicate characteristics and processes within a landscape (Callow et al., 2007; Jarihani et al., 2015). Data from a digital elevation model (DEM) representing topography comprise one of the critical datasets required in many types of studies, such as those of runoff generation, lake water storage changes, river routing, and flood inundation modeling (Hawker, Bates, et al., 2018; Jung & Jasinski, 2015; Sampson et al., 2016; Yamazaki et al., 2014, 2017). DEMs provide key data that govern the accuracy of hydrological and hydrodynamic models (Bates et al., 1998; Bates et al., 2005; Baugh et al., 2013; Jarihani et al., 2015; Sanders, 2007). Highly accurate DEMs are needed for the better representation of river hydrodynamics and flood modeling.

Many regions of the world rely on spaceborne DEMs due to the lack of availability of highly accurate airborne DEMs. Advances in remote-sensing techniques helped with achieving more accurate spaceborne DEMs (Yamazaki et al., 2017). The Shuttle Radar Topography Mission (SRTM) DEM and Advance Spaceborne Thermal Emission and Reflection Radiometer Global DEM, which provide data covering the whole world, are examples of DEMs developed due to improvements in remote-sensing techniques (Jung

& Jasinski, 2015; Yamazaki et al., 2017). Spaceborne DEMs contain various non-negligible errors affecting vertical accuracy. Efforts have been made to correct these errors, such as the steps taken in producing the Multi-Error-Removed Improved-Terrain (MERIT) DEM (Yamazaki et al., 2017) and TanDEM-X 90 DEM (Rizzoli et al., 2017). The MERIT DEM is a highly accurate spaceborne DEM that removes major error components from existing DEMs. It was developed by removing the absolute bias, stripe noise, speckle noise, and tree height bias using multiple satellite datasets and filtering techniques (Yamazaki et al., 2017).

Many studies have demonstrated the advantages of the MERIT DEM over the SRTM and other DEMs (Hawker, Bates, et al., 2018; Hawker, Rougier, et al., 2018; Hawker et al., 2019; Liu et al., 2019; Yamazaki et al., 2017). The comparisons were mostly restricted to vertical accuracy or spatial error assessment using reference altimetry data. There has been little research on evaluating the MERIT DEM using river hydrodynamic model simulations. A few studies have examined the effects of DEMs, including MERIT, in flood inundation modeling (Archer et al., 2018; Hawker, Bates, et al., 2018; Hawker, Rougier, et al., 2018), but the evaluations were limited to flood extent or the water surface elevation (WSE). The hydrodynamic performance can be affected by uncertainty in the runoff or other model parameters, such as Manning's coefficient, river channel bathymetry, and channel width. In accordance with the theory of equifinality, models based on many parameter sets can perform acceptably well, but they "might be right for the wrong reasons" (Beven, 2006; Kirchner, 2006). The acceptable performance of a hydrodynamic model is more likely for single-variable rather than multivariable evaluation. In addition, evaluation based on fewer variables makes it difficult to trace errors (García-Díez et al., 2015). There are many possible solutions to improve the process representation and reduce uncertainty in model predictions (Meyer Oliveira et al., 2021). One of the easiest and most efficient solutions is the use of a complementary dataset to evaluate the model. Insight regarding the performance of different DEMs from considering multiple uncertainties while using river hydrodynamic models can help to reduce the errors for better prediction of flood events.

The robustness and proper representation of the natural system can be confirmed by evaluating the model across various simulated hydrological variables (Stisen et al., 2018). The availability of remote-sensing data with fair temporal and spatial resolution is advantageous for hydrodynamic model evaluation using multiple variables. Many studies (Meyer Oliveira et al., 2021; Paiva et al., 2013; Patro et al., 2009) performed robust evaluation of hydrological-hydrodynamic and hydrodynamic models using multiple variables. The techniques involved the use of many possible objective functions, including the root mean square error (RMSE), Nash-Sutcliffe efficiency coefficient (NSE), and coefficient of determination (R^2). The use of multiple objective functions makes the evaluation cumbersome, and it is difficult to describe the overall combined performance due to all variables. Selection of the best model parameters using these techniques is complex, making objective model evaluation difficult. The integration of these metrics for evaluating river hydrodynamic models is impossible due to different measurement units and ranges. For example, the NSE is unitless and ranges from $-\infty$ to 1, whereas the RMSE has the same unit as the variable of interest and ranges from 0 to ∞. In addition, the spatial dimension and measurement units of observed variables, such as discharge (Q), Water Surface Elevation (WSE), and flood extent, vary. Q and WSE are recorded as point observations with measurement units of m^3/s and m, respectively, whereas flood extent involves two-dimensional observations with a measurement unit of m^2. A methodology for evaluating river hydrodynamic models that includes integrating various observations with different temporal and spatial dimensions has not been established.

This study proposes an integrated multivariable evaluation of river hydrodynamic models for subjective assessment. The proposed multivariable integrated evaluation technique was applied for catchment-based macro-scale floodplain (CaMa-Flood) river hydrodynamic model (Yamazaki et al., 2011) simulations considering two different DEMs, the MERIT DEM and SRTM DEM, as a case study. The evaluation was performed using a range of parameters to cover the various possible uncertainties and errors. An integrated metric approach considering observations of multiple variables (Q, WSE, and flooded area) at the sub-basin scale was adopted. The overall basin skill score (OSK) was calculated using the sub-basin skill score to represent the performance of CaMa-Flood with the two DEMs. The OSK helps to rank the simulations to compare the best set of parameters for the CaMa-Flood river hydrodynamic model considering multiple variables. The proposed evaluation technique will help with determining the improvement in predictions of river hydrodynamics by a representative model due to improvement of a DEM, e.g., the MERIT DEM.

2 Methodology and Data Description

2.1 Study Framework

Simulations with a global river hydrodynamic model, CaMa-Flood, at a resolution of 0.1° (~10 km at the equator) were performed using the SRTM DEM and MERIT DEM, and model performance for multiple variables was evaluated. The framework of the study is presented in **Error! Reference source not found.**. The CaMa-Flood model (see

Section 2.3) used runoff forcing data (see Section 2.5) as the input for the simulations. Simulations were performed using two different DEMs, the SRTM DEM and MERIT DEM (see Section 2.4), for various parameters (see Section 2.7), and uncertainties were included. The sub-basin approach was adopted for skill score computation using three different variables. Q, WSE, and the flooded area (FA) (see Section 2.6) were taken as observations for each sub-basin for metric computation (see Section 2.8). The overall model performance was evaluated based on the OSK, which was calculated by averaging the sub-basin skill scores for a DEM that were estimated for a parameter using an evaluation metric. The best parameter set, as assessed by comparing the OSK, was used to compare the model performance of CaMa-Flood between the SRTM and MERIT DEMs (see Section 3). The integrated multivariable evaluation technique was applied to Amazon River Basin data to compare the performance of the CaMa-Flood river hydrodynamic model with two different DEMs. The average normalized NSE (NNSE) value was calculated for Q, WSE, and FA observations for each sub-basin, considering mainstream observations for Q and WSE and a 50-km buffer around the main stream for FA. The OSK was calculated by averaging the sub-basin skill scores, representing the performance of CaMa-Flood for a given DEM and river depth parameter.

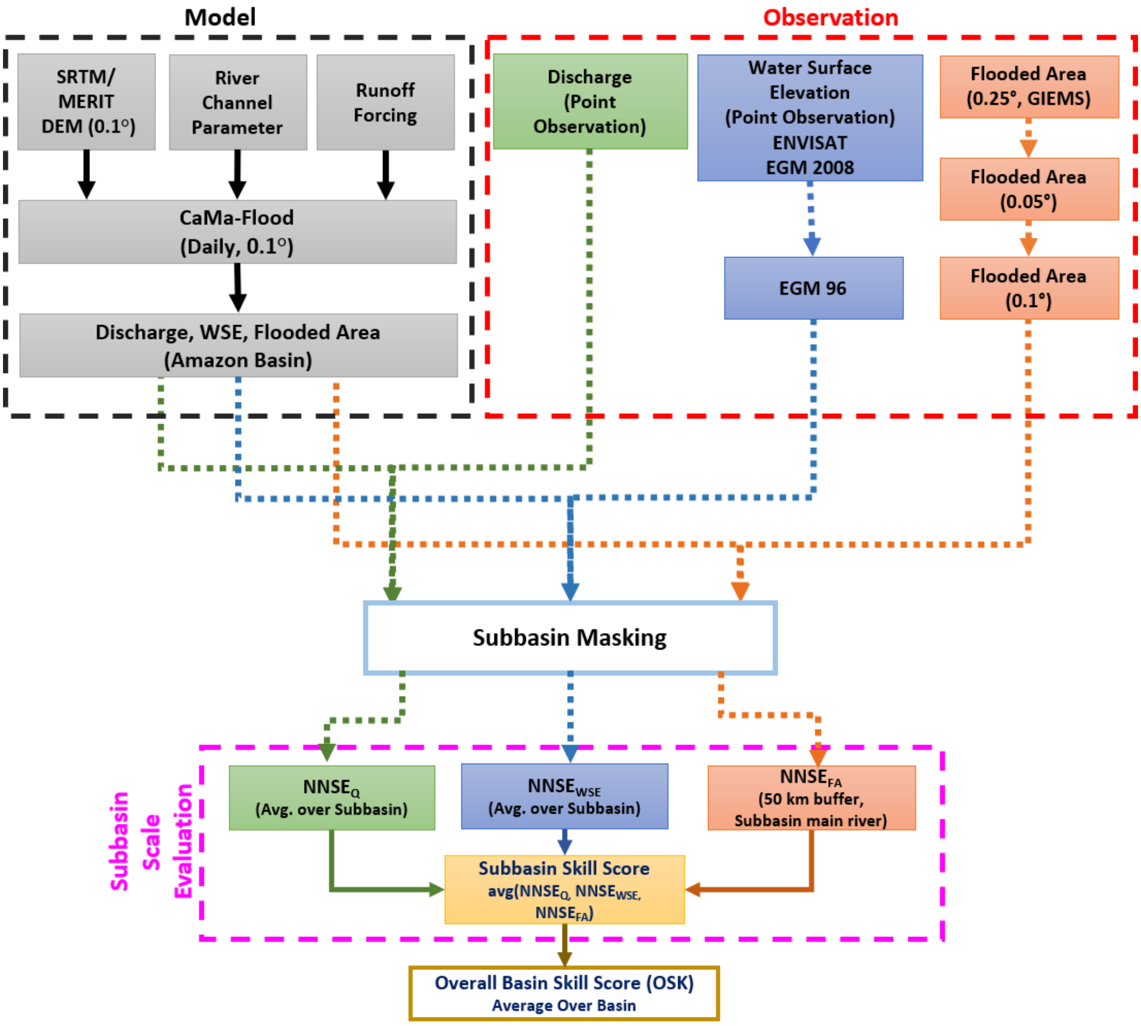


Figure 1: Study framework showing the flowchart of calculation of the OSK from sub-basin skill scores calculated using the NNSE values associated with Q, WSE, and FA.

2.2 Study Area

The Amazon Basin (Figure 4**Error! Reference source not found.**) was selected as the study region to evaluate the MERIT and SRTM DEMs using CaMa-Flood. The Amazon Basin has the largest drainage basin in the world, covering an area of approximately 7,050,000 km^2, with the largest river discharge (annual average of 200,000 m^3/s at the river mouth), and accounts for 20% (one-fifth) of the total runoff discharge into the world's oceans (Richey et al., 1989).

The Amazon region is characterized by complex river hydraulics. The low river slopes cause backwater effects to control part of the river (Meade et al., 1991; Paiva et al., 2013). The vast floodplains along the Amazon main stem significantly impact the hydrodynamics of the middle and lower reaches. Seasonally flooded areas are found on the Amazon plains (Hess et al., 2003; Papa et al., 2010). The main river channel exchanges a 5%–30% annual discharge with the surrounding areas (Alsdorf et al., 2010; Richey et al., 1989). These characteristics make the Amazon Basin a good case study for hydrodynamic model evaluation.

Quantification of hydrological model events for a large basin is often difficult due to a lack of adequate data (Chen et al., 2010). However, large numbers of observations are available for the Amazon region. Many *in situ* gauges are installed along the main stem and major tributaries (Yamazaki et al., 2012), and satellite observation data are useful for monitoring large-scale floods and droughts (Chen et al., 2010).

2.3 Hydrodynamic Model

In this study, we used CaMa-Flood (Yamazaki et al., 2011, 2012, 2013), which is a river routing model with a global distribution. It uses the one-dimensional St. Venant local inertial equation (Bates et al., 2010), making simulations of continental-scale river hydrodynamics computationally efficient. The model simulates river and floodplain hydrodynamics (e.g., river Q, WSE, inundated area, and surface water storage) by routing the input runoff generated by a land surface model to a predefined river network map. The river networks and sub-grid parameters were created by applying the FLOW upscaling method (Yamazaki et al., 2009) at 0.1° (~10 km) resolution, using a flow direction map and DEMs. The SRTM DEM combined with HydroSHEDS (Lehner et al.,

2008) flow directions and MERIT Hydro (Yamazaki et al., 2019) combined with a hydrologically adjusted version of the MERIT DEM were used as input to apply the FLOW upscaling method. River networks were discretized into unit catchments with sub-grid topographic parameters for river channels and floodplains. The water balance equation determines the water storage in each catchment as a prognostic variable, where the catchment is delineated by a DEM. The local inertial equation is used to calculate river Q, whereas the water level and FA are identified using water storage levels in each unit catchment based on sub-grid topographic information.

2.4 Digital Elevation Models

The DEM is one of the most important inputs for physical-based models. In this study, the SRTM and MERIT DEMs were compared using a river hydrodynamic model. SRTM is an international research effort in which DEMs are obtained on a near-global scale from 56°S to 60°N to generate the most complete high-resolution digital topographic database of the Earth. The SRTM DEM (Farr et al., 2007) uses synthetic aperture radar interferometry data to produce the highest resolution digital topographic map of the Earth. It has a resolution of 1 arc-second (i.e., 30 m at the equator) with 15 m of vertical accuracy.

The MERIT DEM (Yamazaki et al., 2017) is a highly accurate global DEM with a 3 arc-second resolution (~90 m). The SRTM3 DEM (Farr et al., 2007) and AW3D-30 m DEM (Tadono et al., 2015) were used as baseline DEMs to produce the MERIT DEM after the removal of multiple error components (absolute bias, stripe noise, speckle noise, and tree height bias) from existing spaceborne DEMs. Improved DEMs (without biases and errors) can represent the altitude and slope more accurately, as illustrated in

Figure 2(b) and (c). For a river channel delineated using a DEM, any error in altitude in the DEM will directly affect the channel depth; hence, accurate WSE values may not be obtained even if Q can be estimated correctly. As shown in Figure 2, a biased DEM cannot provide accurate results, even if Q values are correct, and the WSE data will be affected due to errors in elevation and slope or vice versa. The MERIT DEM is expected to analyze multiple variables simultaneously in an improved manner, thus enhancing the overall performance of the river hydrodynamic model.

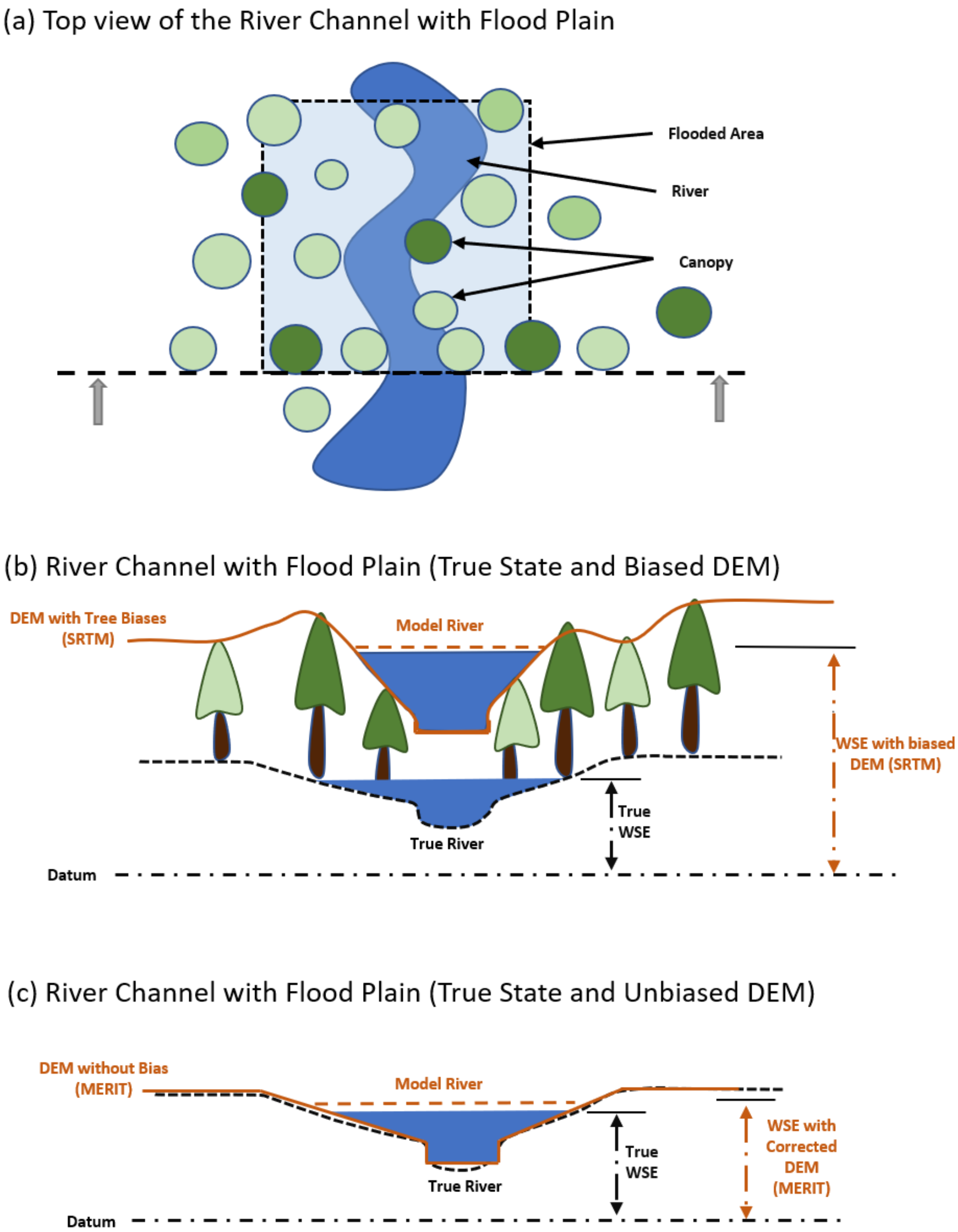


Figure 2: Biased and corrected DEM showing the effects of trees bias on Q, WSE, and FA estimations.

2.5 Runoff Forcing Data

CaMa-Flood uses runoff as an input for simulations. We used eartH2Observe (Dutra et al., 2017) runoff data produced by the land surface hydrological model Hydrology Tiled ECMWF Scheme for Surface Exchanges over Land (HTESSEL) forced with the

WATCH Forcing Data methodology applied to ERA-Interim dataset (WFDEI; Weedon et al., 2014) on weather boundary conditions (Balsamo et al., 2009). The combination of HTESSEL runoff data and CaMa-Flood produced better results (Zhou et al., 2020). The runoff data resolution is 0.25°, distributed to each unit catchment according to the areal proportion of the unit catchment in the corresponding grid.

2.6 Observation Data

2.6.1 Discharge

Daily Q data for the observation locations shown in Figure 4 were used to evaluate the Q simulations. The Brazilian Agency for Water Resources (ANA), Peruvian and Bolivian National Meteorology and Hydrology Services (Servicio Nacional de Meteorología e Hidrología), and Hydrology, Biogeochemistry and Geodynamic of the Amazon Basin (HYBAM) program (http://www.ore-hybam.org) provided daily-scale data for the 1999–2009 period. The data for 2001–2009 were considered for integrated multivariable evaluation.

2.6.2 Water Surface Elevation

ENVISAT satellite altimetry data (Santos da Silva et al., 2010, http://hydroweb.theia-land.fr/) were used to evaluate the WSE. The ENVISAT satellite has a 35-day repeat orbit and an intertrack distance of 80 km. Data from 2002 to 2010 were used for metric calculation. The ENVISAT altimetry data referenced EGM 2008. Preprocessing was performed using the program provided by the National Geospatial-Intelligence Agency (http://earth-info.nga.mil) to convert the data to EGM96 geoid format. The processed EGM96 geoid-referenced data were compared with the CaMa-Flood simulations.

2.6.3 Flood Extent

A multi-satellite monthly global inundation extent dataset with a spatial resolution of approximately 25 km × 25 km, available from 1993 to 2004 (Papa et al., 2010), was used for flood extent comparison. These data were derived from multiple satellite observations comprising passive (Special Sensor Microwave Imager) and active (ERS scatterometer) microwaves along with visible and near-infrared imagery (Advanced Very High-resolution Radiometer). Multi-satellite data can capture inundation under the vegetation canopy and were therefore used to evaluate CaMa-Flood simulations of the Amazon Basin. Data were provided for an equal-area grid of 0.25° × 0.25° at the equator, where each pixel has a surface area of 773 km^2. They were converted into the Cartesian coordinate system for comparison with CaMa-Flood simulations. The data with the modified coordinate system were downscaled from 0.25° (~25 km) to 0.05° (~5 km) by dividing the values equally over 25 pixels. Later, they were upscaled from 0.05° (~5 km) to 0.1° (~10 km) by summing the values over 2 × 2 pixels. The final 0.1° downscaled observation data were used for comparison with the CaMa-Flood simulations.

2.7 Model Parameterization

River hydrodynamic models have many sources of uncertainty. These model uncertainties were taken into consideration in the model evaluation by assessing a range of parameters. The CaMa-Flood model has three river channel parameters: channel width, channel depth, and Manning's roughness coefficient. A constant value of 0.03 is given as Manning's coefficient in the model. Channel width was derived using satellite data (Yamazaki et al., 2014), whereas river depth was estimated using the power law equation as follows:

$$H = aQ^b \tag{1}$$

Where H is the river depth, Q is annual average discharge, a is coefficient b exponent of the power law given by constant value

River depth is the most uncertain of the three parameters as the uncertainty in the Manning's coefficient is small, with the usual range for rivers lying between 0.02 and 0.04 (Brêda et al., 2019; Chow, 1959), whereas an empirical equation is used to calculate depth. Depth is perturbed to obtain a range of parameters for each simulation to encompass the model's uncertainties. There are many possible ways to change river depth by changing the value of the constant in the power law equation. Here, the river depth was varied by varying the depth at the river mouth and changing the gradient across the basin, as illustrated in Figure 3. The exponent "b" varied from 0.35 to 0.85 with an interval of 0.05 and the coefficient "a" of the power law equation to achieve a fixed depth value at the river mouth.

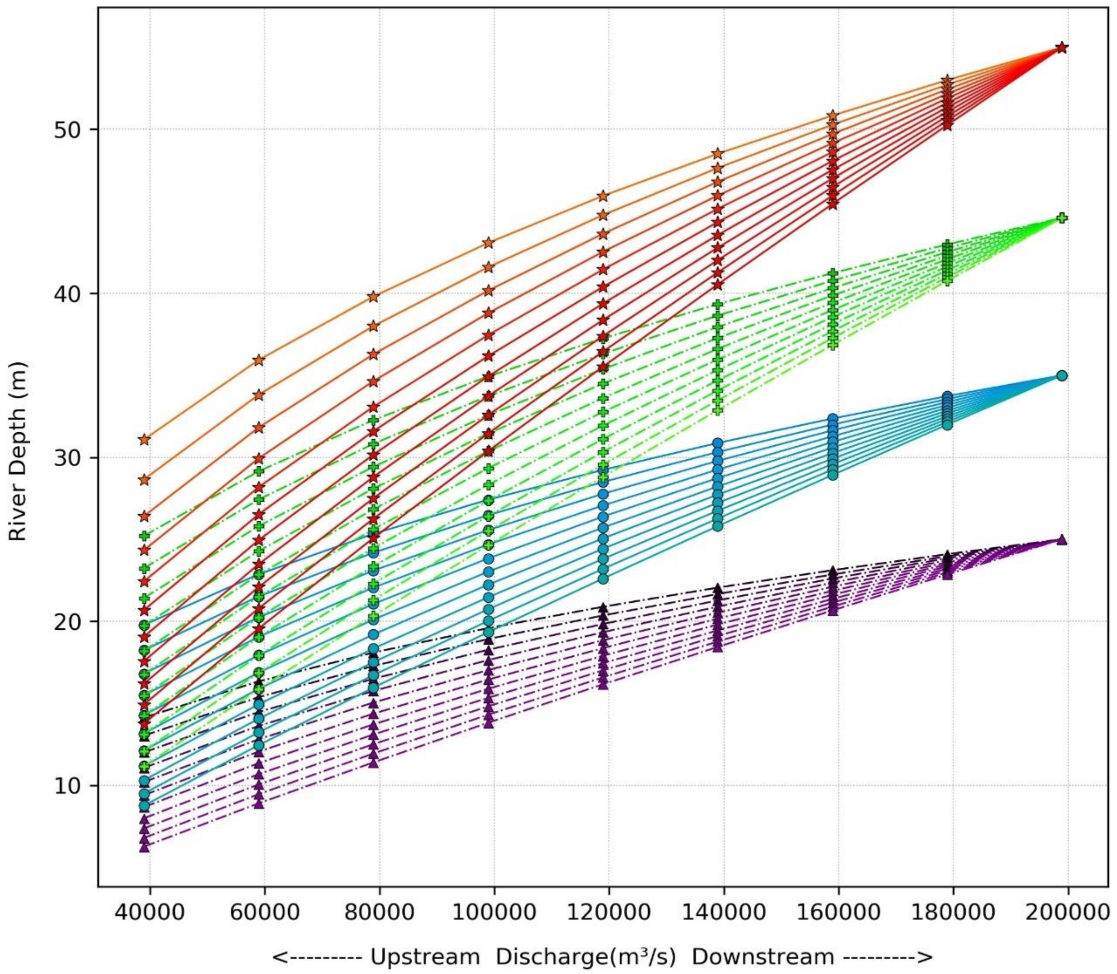


Figure 3: River depth vs. Q plot representing the power law equation ($H = aQ^b$). Lines were drawn considering the depth at the river mouth (the same color represents the same depth at the river mouth), and the gradient along the river was changed by adjusting the range of exponent "b" and coefficient "a" accordingly.

2.8 Sub-basin Skill Evaluation

The three variables used in the model evaluation study were Q, WSE, and FA. All three variables have different dimensions, i.e., Q and WSE are recorded as point observations, whereas FA is recorded as areal observations. In addition, these variables were not observed at the same location, as shown in Figure 4(a)-(b). The sub-basin approach was adopted to overcome the difficulties related to observation locations and dimensions.

The basin was decomposed into similar sub-basins, and the skill score was calculated considering all three variables for each sub-basin. In this study, the Amazon Basin was decomposed into 25 sub-basins as shown in Figure 4**Error! Reference source not found.**(c). The sub-basins were produced using a sub-basin minimum area threshold value of 1.5×10^{11} m^2 and a minimum percentage of each sub-basin's contribution to the confluence point of 1%. Tributaries with contribution areas to the main stream between two sub-basins (i.e., interbasin) larger than a threshold value of 1.5×10^{11} m^2 were also treated as separate sub-basins. River pixels having an upstream area larger than the input threshold of 1.5×10^{11} m^2 treated as the main stream river for each subbasin.

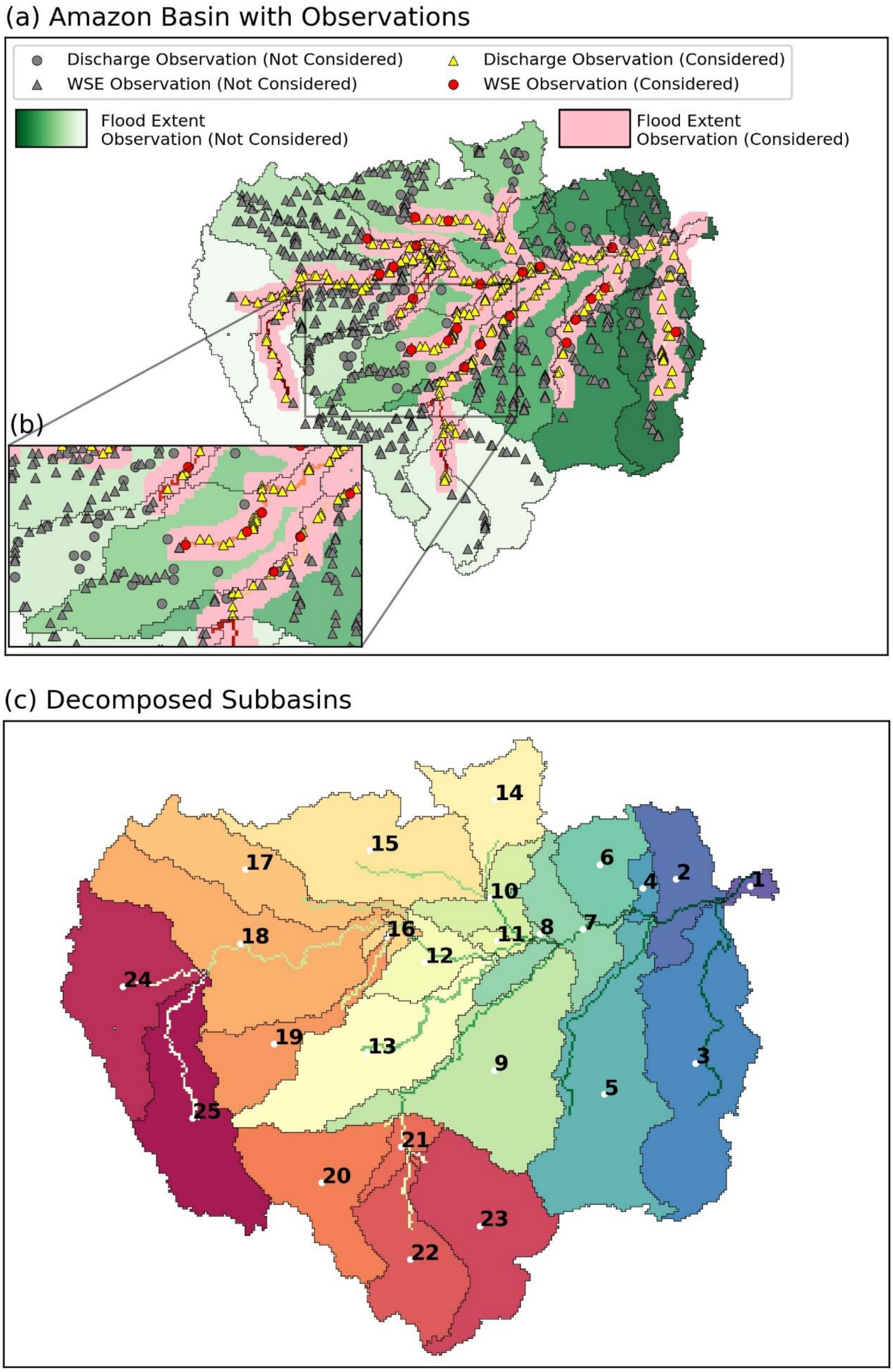

(a) Amazon Basin with Observations
Discharge Observation (Not Considered)
WSE Observation (Not Considered)
Discharge Observation (Considered)
WSE Observation (Considered)
Flood Extent Observation (Not Considered)
Flood Extent Observation (Considered)
(b)
(c) Decomposed Subbasins
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

Figure 4: (a) Map of the Amazon Basin with Q, WSE, FA observation locations indicated (gray circles, triangles, and green shaded areas, respectively), (b) The zoomed-in area shows sub-basin 13, with red circles and yellow triangles indicating the considered observations of Q and WSE, respectively, along the sub-basin main river channel and the pink shaded area indicating the FA considered (50 km on each side of the main river). (c) Decomposed sub-basins with identification numbers.

The skill score of each sub-basin was calculated considering the Q and WSE point observations for the main stream of the sub-basin to exclude impacts from minor streams and take large-scale hydrodynamics into consideration, as shown in Figure 4(b). For FA, permanent water bodies were excluded from the analysis by considering a 100-km-wide buffer zone (50-km buffer on each side) along the main river channel (orange-red thin line), as shown in Figure 4(a)-(b). The floodplain width of Amazon is approximately 30 times of main river channel width (Paiva et al., 2011), and hence 50-km buffer on each side of the river channel selected to represent the farthest flood region influenced by the main river. The buffer region helps with determining the impacts of main stream parameters on flood extent while neglecting flooding from small streams and permanent water bodies in the model evaluation.

Most model evaluations using multiple variables are performed with multiple objective functions, which usually have different scales and units. This makes multivariable evaluation burdensome and inappropriate for large basins. A simple model evaluation metric for measuring the quality of a simulation using multiple variables was developed, namely, performance was evaluated in terms of the OSK, a single value with no units.

This metric was developed using the NNSE, as calculated in Eq. 2, and derived using the NSE (Nossent & Bauwens, 2012), as shown in Eq. 3.

$$NNSE = \frac{1}{2-NSE} \tag{2}$$

$$NSE = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(O_i - \bar{O})^2} \tag{3}$$

where O_i and P_i are the observed and modeled values at time i, respectively, and $\bar{O}$ is the mean of observations.

As the model was evaluated for different variables with different dimensional units, the NNSE makes it easier to assess the overall performance across all variables. This evaluation method is not overly sensitive to a model consistently deviating from mean observation values, and the metric ranges from 0 to 1. The metric does not have any units and can be compared or combined for many different variables, as well as arithmetically averaged to obtain a unique value for spatially distributed values. The sub-basin skill score was calculated for the Amazon sub-basins for which observations were available for all three variables. A sub-basin without observations for any of the three variables, Q, WSE, and FA, was excluded from the calculation of OSK. The OSK was calculated as follows:

$$F_{OSK} = \frac{1}{K}\sum_{i=1}^{K}\left(W_Q \times \frac{1}{m}\sum_{j=1}^{m} NNSE_Q + W_{WSE} \times \frac{1}{n}\sum_{j=1}^{n} NNSE_{WSE} + W_{FA} \times NNSE_{FA}\right) \tag{4}$$

where m and n are the number of observations for Q and WSE, respectively, in each sub-basin for the main stream. W_Q, W_{WSE}, and W_{FA} are the weights given to each variable, where $W_Q + W_{WSE} + W_{FA} = 1$ and K is the number of sub-basins (here, it is 25) used in the OSK calculations. For the current assessment, we weighted each variable equally, i.e., $W_Q = W_{WSE} = W_{FA} = 1/3$.

3 Results

Integrated multivariable evaluation of Amazon Basin modeling was applied. Figure 5 shows the contour plot of OSKs for different parameters defined using the coefficient “a” and exponent “b” from Eq. 1. The OSK (Figure 5) indicates the performance of the CaMa-Flood model for the respective parameter and DEMs with equal weighting for each variable for each sub-basin. As observed for most parameters, the CaMa-Flood hydrodynamic model performed better with the MERIT DEM than with the SRTM DEM. The maximum OSK values of 0.46 and 0.51 for the SRTM and MERIT DEMs indicated the best set of parameters for multivariable evaluation. Here, CaMa-Flood performed best with the SRTM DEM for a = 0.349 and b = 0.35, and with the MERIT DEM for a = 0.144 and b = 0.45. In the simulations performed with wide ranges of parameter values, the robust performance of CaMa-Flood was confirmed, and the model was more accurate with the MERIT DEM than with the SRTM DEM (Figure 5).

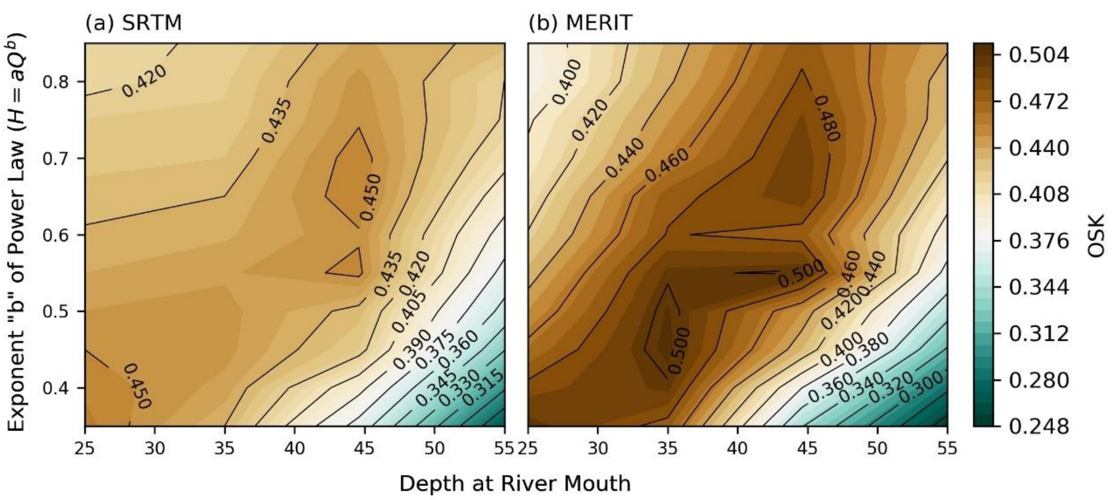


Figure 5: OSK contour plot with equal weighting of all three variables (Q, WSE, and FA) for (a) the SRTM DEM and (b) the MERIT DEM.

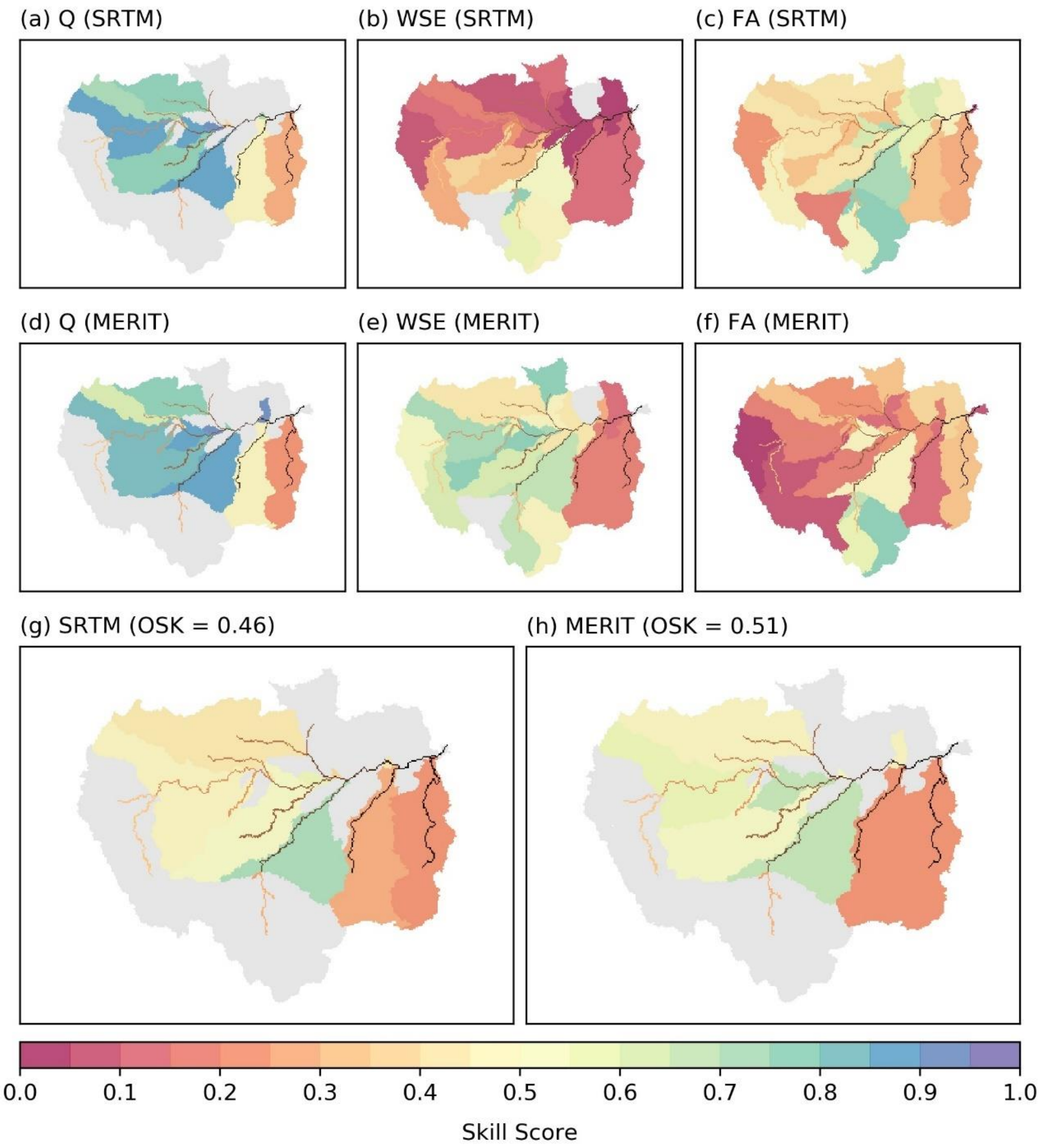


Figure 6: Sub-basin skill scores for (a) Q, (b) WSE, and (c) FA (or flood extent) with the SRTM DEM and for (d) Q, (e) WSE, and (f) FA with the MERIT DEM, and skill scores representing combined multiple variables with the (g) SRTM and (h) MERIT DEMs for the best set of parameters.

The sub-basin skill scores for each variable are shown separately in Figure 6(a)–(f) for the SRTM DEM with a = 0.349 and b = 0.35, corresponding to a depth at the river mouth of 25 m, and for the MERIT DEM with a = 0.45 and b = 0.144, corresponding to a depth at the river mouth of 35 m. Here, a and b represent the coefficient and exponent of Eq. 1, respectively, and the parameters corresponding to the best OSK in Figure 5. The OSK for the Amazon Basin was obtained by taking the arithmetic average of the skill score of each sub-basin. Considerable improvement was observed for WSE predictions as it is directly affected by topography. The gray color (Figure 6) indicates sub-basins for which observations of the variables of interest were unavailable. Among the three variables, the FA predictive performance was poorer than that for the other two variables with both the MERIT and SRTM DEMs. This may be due to many reasons, including poor spatial (~25 km) and temporal resolution (monthly average) of the FA observation data, and the indirect estimations of flood extent (i.e., FA) in the Global Inundation Extent from Multi-Satellites dataset are associated with more uncertainty compared to the directly measured Q and WSE observations. The simplicity of the downscaling method adopted (see Section **Error! Reference source not found.**) may not represent the complexity of the true values, and uncertainties may be present. The sub-basin skill scores and OSKs for combined multiple variables are shown in Figure 6(g)–(h). A significant improvement was observed with the MERIT DEM compared to the SRTM DEM. In addition, a major improvement occurred for upstream sub-basins with the MERIT DEM compared to the SRTM DEM. This suggests that the MERIT DEM can more accurately represent small streams and topography than can the SRTM DEM, especially for upstream basins.

Plots of simulated and observed Q over time for sub-basin 18 and sub-basin 3 are shown in Figure 7(a) and (b) (see Figure 4(c)) for the best set of parameters. These sub-basins were selected to represent basins located at upstream and downstream, respectively, with good and poor predictive performance based on the sub-basin skill score value (calculated using the NNSE value for Q, WSE, and FA). Sub-basin 18 had sub-basin skill scores of 0.47 and 0.57, whereas sub-basin 3 has sub-basin skill scores of 0.21 and 0.24 for the SRTM and MERIT DEM, respectively. As shown in Figure 7(a) andFigure 7(b), which represent sub-basin 18 and sub-basin 3 (see Figure 4 (c)), respectively, Q was not sensitive to which DEM was used. The relative RMSEs (RRMSEs) for sub-basin 18 were 13.07% and 13.59%, and those for sub-basin 3 were 149.86% and 145.17% for SRTM and MERIT, respectively, confirming that the model exhibited similar performance with both DEMs. Figure 7(c) and Figure 7(d) show the time-varying simulated and observed WSE (ENVISAT) for sub-basins 18 and 3 (see Figure 4(c)) for some of the virtual stations. The RRMSE values for sub-basin 18 observations were 12.41% and 2.5%, and those for sub-basin 3 observations were 3.34% and 1.22% for SRTM and MERIT, respectively. These observations demonstrated significant improvement in the WSE simulations for MERIT compared to SRTM. This improvement was likely mainly due to the removal of tree biases (tree height) in the Amazon region, which decreases the WSE. Although SRTM performed better than MERIT (refer to Figure 7(e),(f)) in terms of predicting FA in sub-basin 18, with RRMSEs of 31.49% and 64.51% for SRTM and MERIT, respectively, the high peaks for sub-basin 3 were better represented by MERIT than by SRTM. RRMSE values of 31.21% and 28.92% for SRTM and MERIT, respectively, showed that the MERIT simulated FA slightly better than SRTM for sub-basin 3, suggesting that the accuracy assessment may depend on the choice of evaluation metrics.

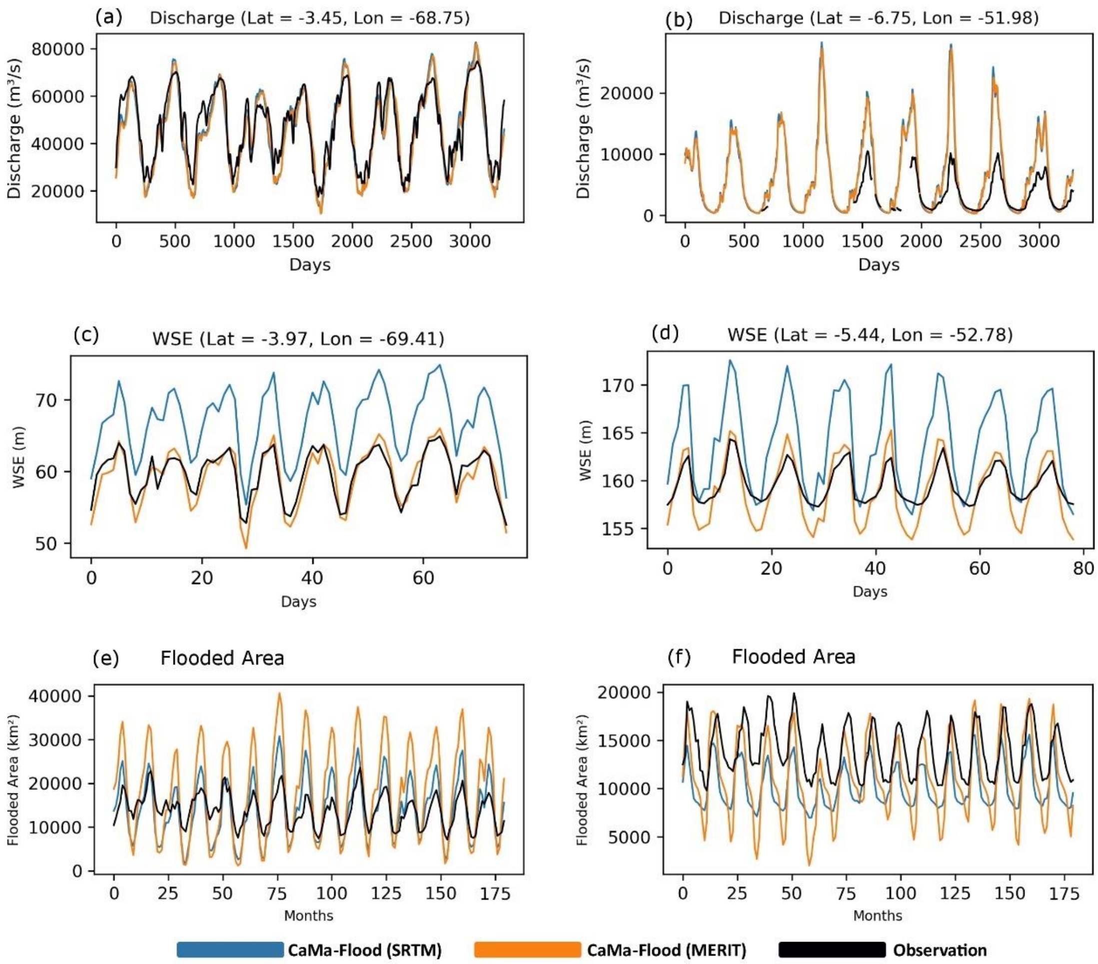


Figure 7: Time-series plots of simulated and observed Q for (a) sub-basin 18 (see Figure 4(c)) and (b) sub-basin 3 (see Figure 4(c)), simulated and observed WSE for (c) sub-basin 18 (see Figure 4(c)) and (d) sub-basin 3 (see Figure 4(c)), and simulated and observed FA for (e) sub-basin 18 (see Figure 4(c)) and (f) sub-basin 3 (see Figure 4(c)).

Figure 8 shows boxplots for the multivariable sub-basin skill score for all study parameters. The whiskers indicate the range of each sub-basin skill score. The mean of each boxplot as indicated by the black plus sign represents the OSK of CaMa-Flood for a particular DEM. The box bordered by red dashed lines contains boxplots of the maximum OSK, i.e., those for the best parameters based on the integrated multivariable evaluation of the CaMa-Flood river hydrodynamic model considering a particular DEM. The interquartile range for the best parameter set was small compared to most of the other parameters, suggesting that the best parameters performed almost equally well for most of the sub-basins.

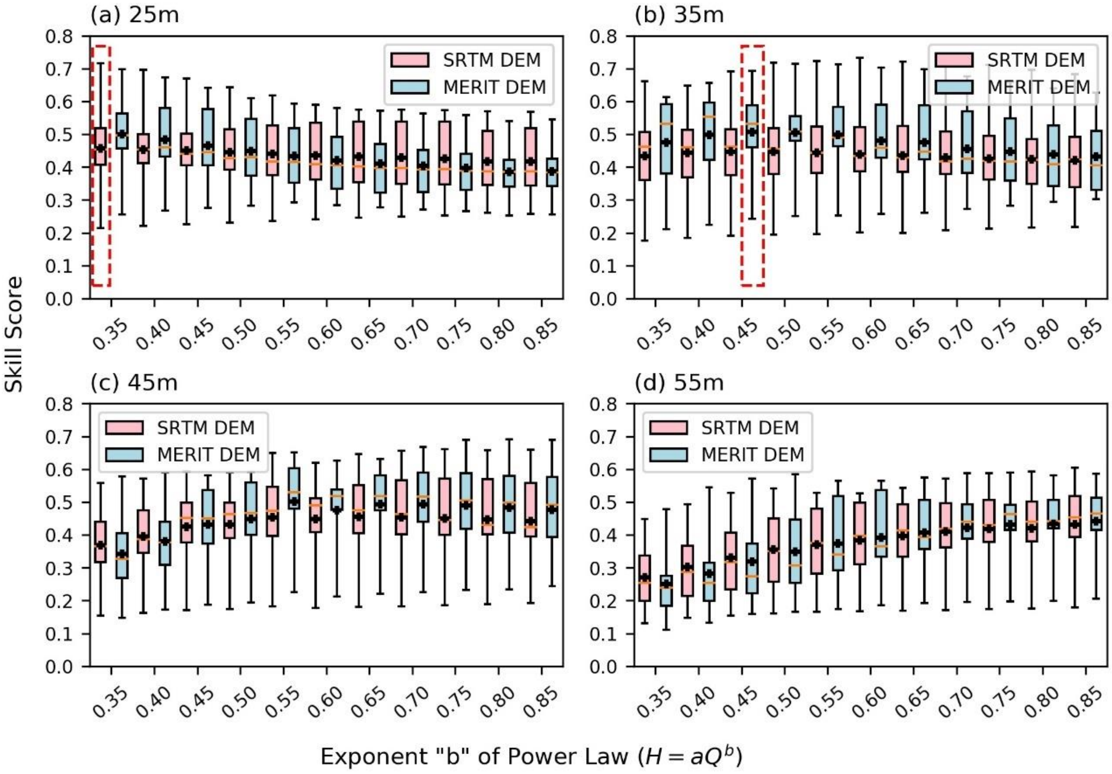
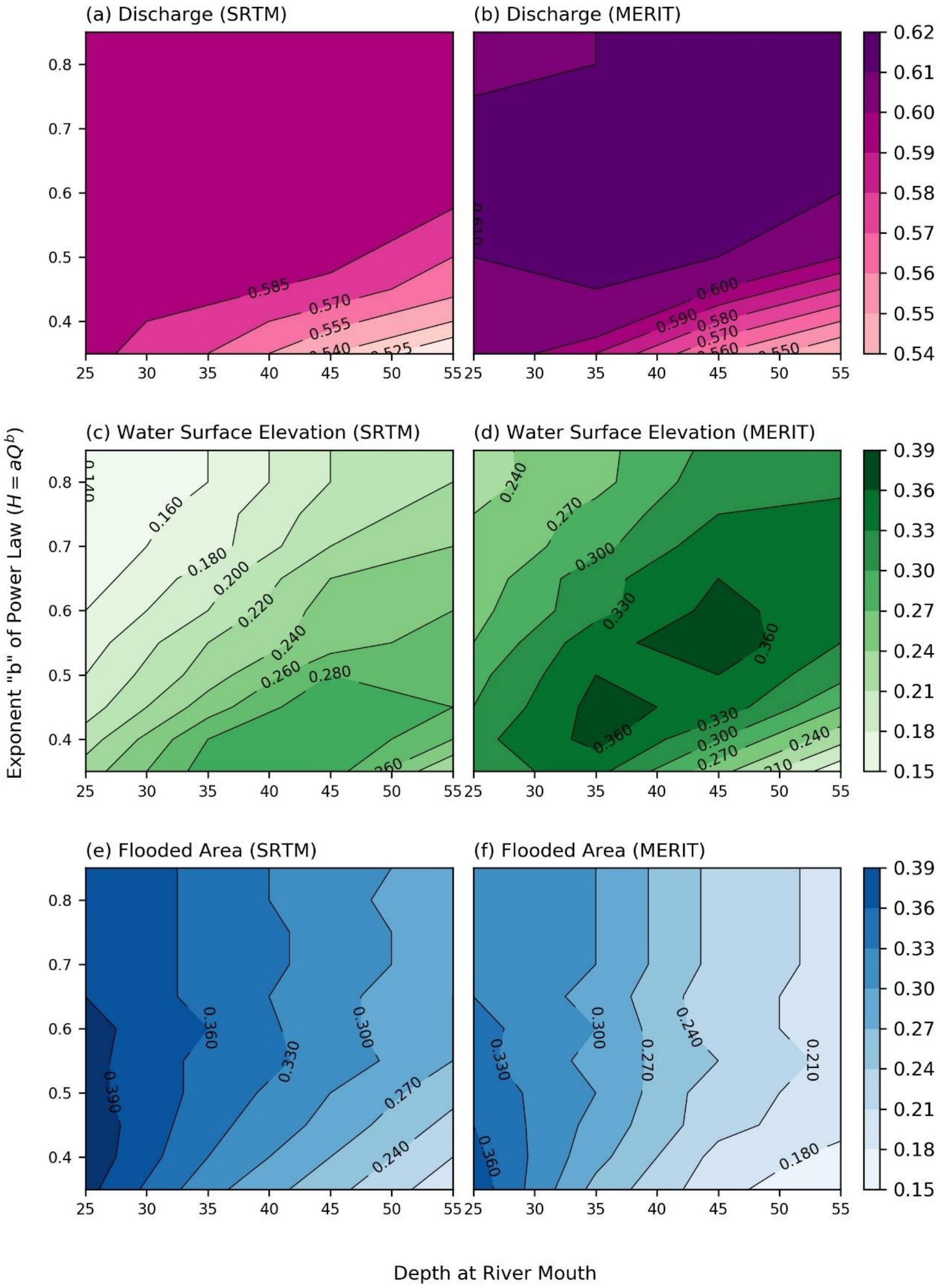


Figure 8: Sub-basin skill scores for the parameters considered by the SRTM and MERIT DEMs for depth at the river mouth of (a) 25 m, (b) 35 m, (c) 45 m, and (d) 55 m.

4 Discussion

To confirm the robustness of integrated multivariable evaluation, we performed a comparison against single-variable evaluation. Here, we evaluated a model of interest using average NNSE values for Q, WSE, and FA separately. Figure 9 shows the contour plot of OSKs, i.e., average basin NNSE values considering Q, WSE, and FA individually, for the whole Amazon Basin regardless of location and without considering the sub-basin approach. The best parameter set could not be determined using single-variable evaluation. The maximum performance of the model was seen for a wide range of parameters when considered separately; e.g., there were no peaks observed in the Q contour (Figure 9(a),(b)). It is challenging to find the best set of parameters by evaluating only one variable for the SRTM and MERIT DEMs.

Single-variable evaluation using Q, WSE, and FA for the CaMa-Flood river hydrodynamic model suggested that Q sensitivity is low and cannot be used for model calibration. Evaluation using WSE showed that the best parameter set obtained using only a single variable can lead to poor accuracy outcomes for the other variables, especially in the case of wrong/poor models (here, the SRTM DEM) (Figure 9(c)). Comparing the WSE contour plot (Figure 9(c),(d)) with the contour plots for other variables (Q and FA) (Figure 9(a),(b),(e),(f)), the best parameters (peaks in the contours) derived using WSE did not correspond to the best parameters (peaks in the contours) derived using the two other variables, and the model could perform well but for the wrong reasons. By contrast, the optimal value for FA could not be achieved, probably due to errors in the flood topography data, low spatial and temporal resolution of observation data, and low accuracy.



(a) Discharge (SRTM)
(b) Discharge (MERIT)
(c) Water Surface Elevation (SRTM)
(d) Water Surface Elevation (MERIT)
(e) Flooded Area (SRTM)
(f) Flooded Area (MERIT)
Exponent "b" of Power Law ($H = aQ^b$)
Depth at River Mouth

Figure 9: Basin skill score (average NNSE for the whole basin) contour plot considering (a) Q for the SRTM DEM, (b) Q for the MERIT DEM, (c) WSE for SRTM, (d) WSE for MERIT, (e) FA for SRTM, and (f) FA for MERIT.

The new multivariable evaluation technique is flexible, and the weights for each variable can be selected manually according to the user’s requirements. The method is flexible enough that the number of variables used for evaluation and averaging can be changed according to the user’s requirements. Figure 10 shows an example of metric calculation with different weighting of variables. The overall skill score was calculated by giving weights of 30%, 30%, and 40% to the Q, WSE, and FA observations, respectively. More weight was given to FA to allow optimization of the parameter location for better estimations of FA.

The maximum value of 0.49 was obtained with a = 0.349 and b = 0.35, corresponding to a depth at the river mouth of 25 m for the MERIT DEM, for the given range of parameters. Figure 11(a)–(c) show the sub-basin skill scores (average NNSE values) of the individual basins for each variable separately. These scores correspond to the parameters Q, WSE, and FA with assigned weights of 30%, 30%, and 40%, respectively. As shown in Figure 11(d) and 11(f), FA predictive performance was enhanced for most of the sub-basins with increases in the weight of the FA variable, whereas the change in Q predictive performance was not significant. The downstream sub-basin performance improved, whereas the FA predictive performance for a few upstream sub-basins appeared to worsen. This may be because a decrease in depth at the mouth further decreases the river depth upstream, which may not represent the true situation.

Figure 12 shows observation and simulated time-series plots of Q, WSE, and FA for the best set of parameters for the MERIT DEM with equal weights assigned to the three variables and with greater weighting of FA compared to the other two variables. The results showed significant improvement in FA predictive performance for sub-basin 3 with a change in RRMSE value from 27.61% to 17.33%. A significant decline was not observed for sub-basin 18, but the RRMSE value changed from 64.55% to 86.23%. The predictive performance of the other two variables (Q and WSE) did not change significantly for both of these sub-basins. The RRMSE value changed from 13.59% to 13.84% for sub-basin 18 Q observations and from 145.17% to 146.18% for sub-basin 3. The RRMSE changed from 2.5% to 3.15% for sub-basin 18 WSE observations and from 1.22% to 1.27% for sub-basin 3, when equal weights were assigned to the variables versus more weight assigned to FA. Our method can be used for the robust evaluation of river hydrodynamic models, and it is flexible enough to allow modification according to the user's requirements global optimization considering multiple variables.

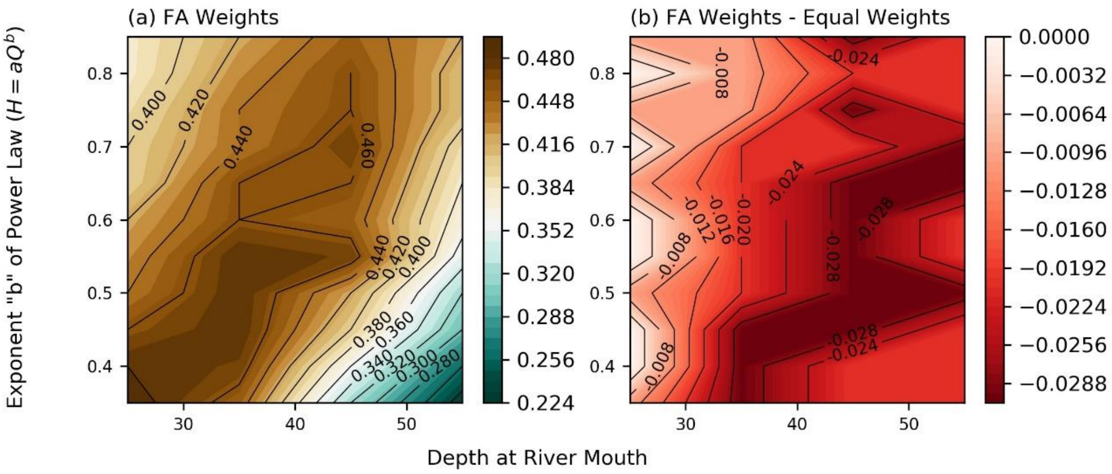


Figure 10: OSK contour plot for (a) greater weight assigned to FA (40%) and for (b) differences in score between more heavily weighted FA and equal weights across variables.

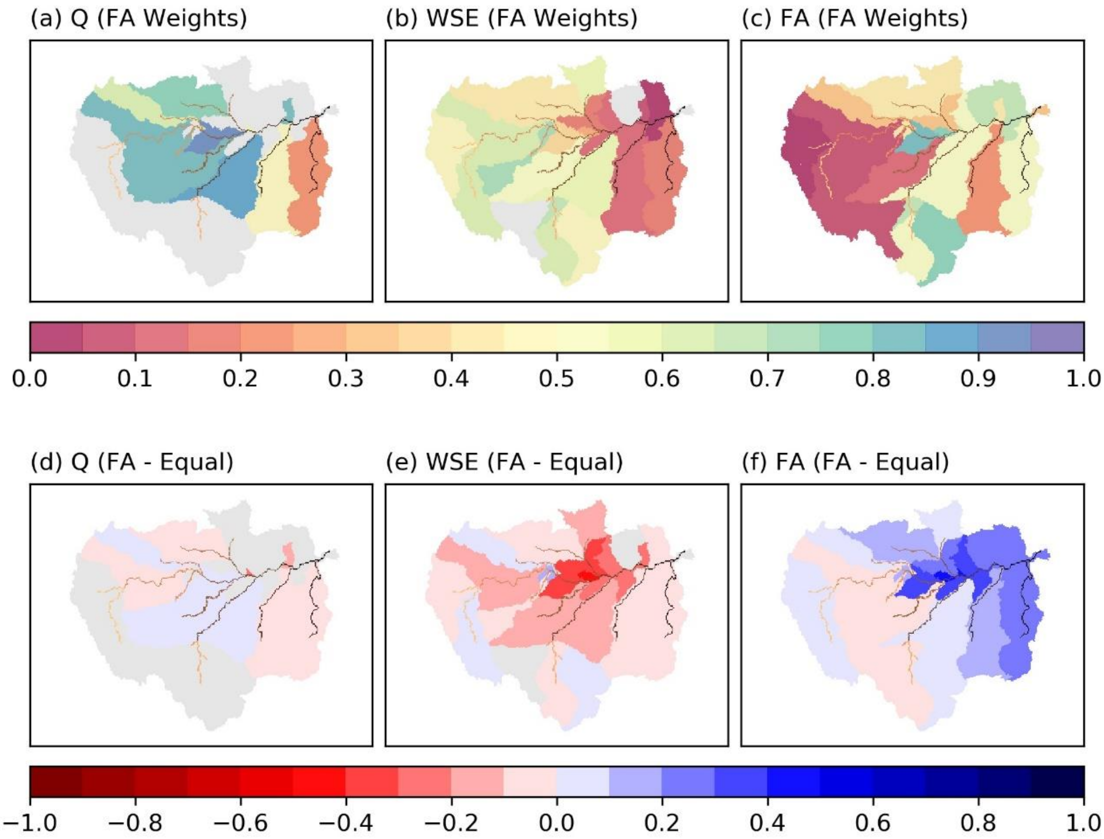


Figure 11: Sub-basin skill scores for (a) Q, (b) WSE, and (c) FA with the MERIT DEM, with weights of 30%, 30%, and 40% assigned to Q, WSE, and FA, respectively, and differences in sub-basin skill score between more heavily weighted FA and equal weighting of variables for (d) Q, (e) WSE, and (f) FA.

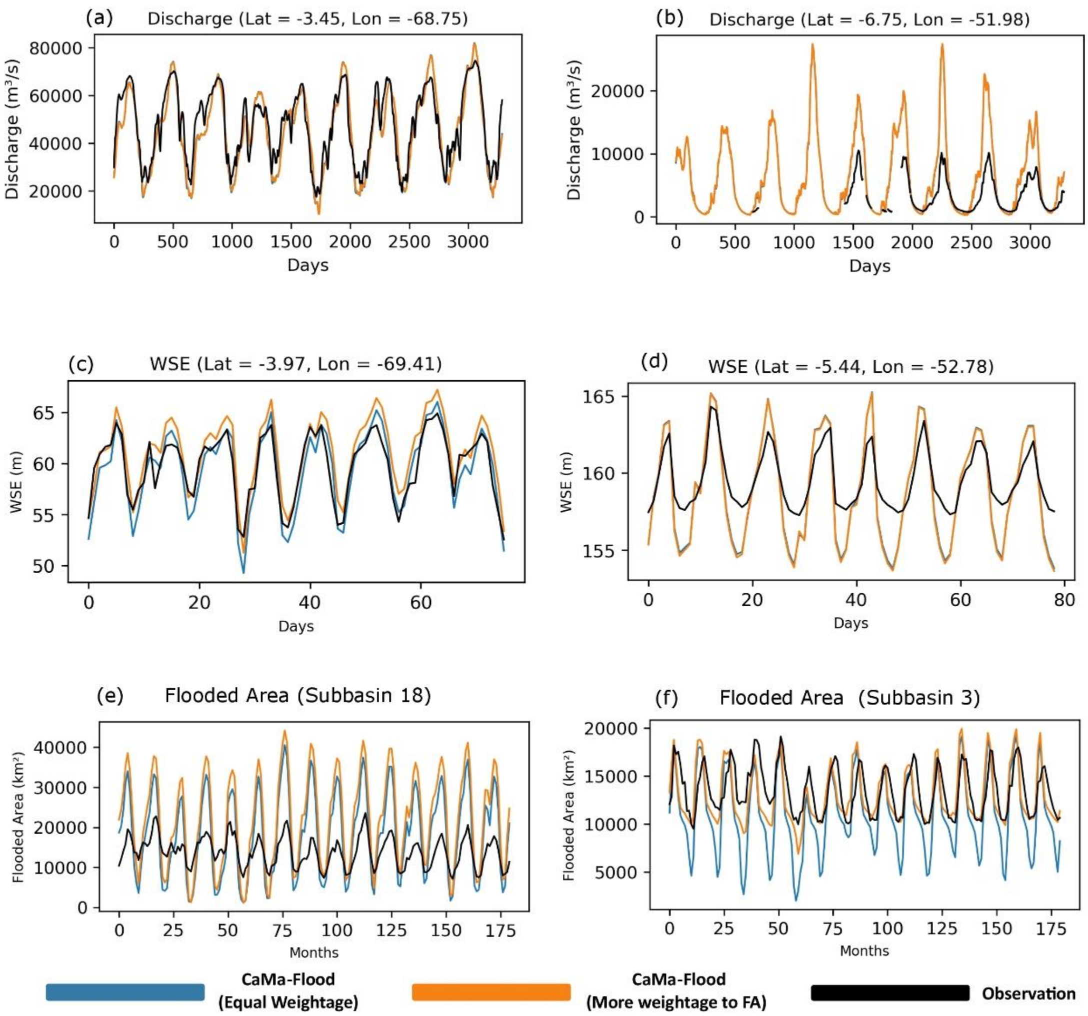


Figure 12: Time-series plots of simulated and observed (a) Q, (c) WSE, and (e) FA for sub-basin 18 (Figure 4(c)) and simulated and observed (b) Q, (d) WSE, and (f) FA for sub-basin 3 (Figure 4(c)) with the MERIT DEM using the best parameter set with equal weighting of variables.

Although we used the NNSE as the target metric, the method can be used with other metrics, e.g., flood extent can be evaluated using performance indices (Aronica et al., 2002; Bates & De Roo, 2000) that measure the agreement between predicted and observed flood extent, especially with static data (Hess et al., 2003). The same method can also be applied using other metrics, such as the Kling-Gupta Efficiency (KGE) (Gupta et al., 2009) and NSE (Nash & Sutcliffe, 1970), among others. The current method can be modified such that different specific weights are assigned to the observations, e.g., assigning greater weight to main river channel observations or downstream sub-basin observations.

5 Summary and Limitations

We developed an integrated multivariable evaluation technique for the robust assessment of river hydrodynamic models and confirmed that the MERIT DEM performed better than the SRTM DEM. The developed multivariable evaluation method uses the NNSE as the evaluation metric for simplicity of calculation. A sub-basin skill score approach for calculating an integrated metric across multiple variables was adopted to ease the difficulty associated with the spatial distribution of observations. The results obtained showed that the model can be evaluated across multiple variables at the same time with the proposed technique. With this method, the prediction of multiple variables is shown to improve with a better model, and combining multiple variables in the evaluation produces a more balanced estimation of performance.

The technique was applied to compare the performance of CaMa-Flood with two different DEMs, the SRTM and MERIT DEMs. As shown in Figure 5 and 6, the maximum value of the integrated metric for CaMa-Flood (OSK) with the MERIT DEM was 0.51, compared to the value of 0.46 with the SRTM DEM with equal weighting of each variable. The sub-basin skill scores for each variable (Figure 6(a)–(f)) separately showed significant improvement in WSE prediction as the DEMs (topography) directly affect the accuracy of WSE simulation. The results shown in Figure 6(g) and (h) indicate that with the MERIT DEM, the predictive performance for upstream sub-basins improved considerably. This improvement implied that the upstream sub-basins were affected to a greater extent due to tree biases in the SRTM DEM, whereas there was significant bias correction in the MERIT DEM. Single-variable evaluation (see Section 4) confirmed that robust evaluation of a hydrodynamic model is challenging to achieve using a single variable at a time. Figure 9(a), 9(b), 9(e), and 9(f) show that Q and FA are not suitable for use in calibration as the sensitivities are low and the same skill score values were obtained for a wide range of parameters, whereas the peaks obtained for WSE (Figure 9(c),(d)) did not correspond to the peaks of the other two variables, and the best parameter may be obtained in an incorrect manner.

This study had some limitations. Estimating the channel depth using the power law equation is not realistic, as each tributary or river segment can have local characteristics that cannot be explained by the power law equation. The river model error cannot be explained only by the channel depth parameter. There are many possible error sources, such as errors in input runoff and flood plain topography, limitations due to simplification of physics, and representation of the river as a rectangular channel. Although we did not account for the sizes of the sub-basins and corresponding river sizes when weights were assigned to the observations during the evaluation, the evaluation method can be easily altered to incorporate these variations. The developed method can be used for more robust and accurate evaluation after incorporating new schemes to overcome model limitations in future.

The developed integrated multivariable evaluation technique can overcome limitations such as lower discharge sensitivity for calibration, by incorporating other variables into the evaluation. This method can reduce errors by considering the required data for evaluation and is flexible enough to be adapted according to the user's requirements. The method can modified or use a combination of other metrics, such as the NSE, KGE, and flood performance indices. An evaluation can be performed even if observations for some variables are missing for a few sub-basins by eliminating these variables from the evaluation for those particular sub-basins. Future studies will define sub-basin-scale parameters and evaluate models at the basin scale by including other metrics in evaluation process.

Aknowledgement

This research was supported by Japan Society for Promotion of Science (JSPS) Grant-in-Aid for Scientific Research (JSPS KAKENHI Grant Numbers 16H06291, and 20H02251). The authors would like to express gratitude to Dirk Eilander for archiving the E2O runoff dataset, and Dr. Xudong Zhou for processing the GIEMS flooded area data.

Data Availability

All data in this study are publicly available and were accessed at the links given in the text (HYBAM , http://www.ore-hybam.org; ENVISAT, http://hydroweb.theia-land.fr/;), from the literature (GIEMS, Papa et al., 2010). The Global Hydrodynamic Model CaMa-Flood is available from (Yamazaki et al., 2021).

References

Alsdorf, D., Han, S. C., Bates, P., & Melack, J. (2010). Seasonal water storage on the Amazon floodplain measured from satellites. *Remote Sensing of Environment*, *114*(11), 2448–2456. https://doi.org/10.1016/j.rse.2010.05.020

Archer, L., Neal, J. C., Bates, P. D., & House, J. I. (2018). Comparing TanDEM-X Data With Frequently Used DEMs for Flood Inundation Modeling. *Water Resources Research*, *54*(12), 10,205-10,222. https://doi.org/10.1029/2018WR023688

Aronica, G., Bates, P. D., & Horritt, M. S. (2002). Assessing the uncertainty in distributed model predictions using observed binary pattern information within GLUE. *Hydrological Processes*, *16*(10), 2001–2016. https://doi.org/10.1002/hyp.398

Balsamo, G., Viterbo, P., Beijaars, A., van den Hurk, B., Hirschi, M., Betts, A. K., & Scipal, K. (2009). A revised hydrology for the ECMWF model: Verification from field site to terrestrial water storage and impact in the integrated forecast system. *Journal of Hydrometeorology*, *10*(3), 623–643. https://doi.org/10.1175/2008JHM1068.1

Bates, P. D., & De Roo, A. P. J. (2000). A simple raster-based model for flood inundation simulation. *Journal of Hydrology*, *236*(1–2), 54–77. https://doi.org/10.1016/S0022-1694(00)00278-X

Bates, P. D., Horritt, M., & Hervouet, J. M. (1998). Investigating two-dimensional, finite element predictions of floodplain inundation using fractal generated topography. *Hydrological Processes*, *12*(8), 1257–1277. https://doi.org/10.1002/(SICI)1099-1085(19980630)12:8<1257::AID-HYP672>3.0.CO;2-P

Bates, Paul D., Horritt, M. S., Hunter, N. M., Mason, D., & Cobby, C. (2005). *Numerical Modelling of Floodplain Flow*. *Computational Fluid Dynamics: Applications in Environmental Hydraulics* (Vol. 4). https://doi.org/10.1002/0470015195.ch11

Bates, Paul 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

Baugh, C. A., Bates, P. D., Schumann, G., & Trigg, M. A. (2013). SRTM vegetation removal and hydrodynamic modeling accuracy. *Water Resources Research*, *49*(9), 5276–5289. https://doi.org/10.1002/wrcr.20412

Beven, K. (2006). A manifesto for the equifinality thesis. *Journal of Hydrology*, *320*(1–2), 18–36. https://doi.org/10.1016/j.jhydrol.2005.07.007

Brêda, J. P. L. F., Paiva, R. C. D., Bravo, J. M., Passaia, O. A., & Moreira, D. M. (2019). Assimilation of Satellite Altimetry Data for Effective River Bathymetry. *Water Resources Research*, *55*(9), 7441–7463. https://doi.org/10.1029/2018WR024010

Callow, J. N., Van Niel, K. P., & Boggs, G. S. (2007). How does modifying a DEM to reflect known hydrology affect subsequent terrain analysis? *Journal of Hydrology*, *332*(1–2), 30–39. https://doi.org/10.1016/j.jhydrol.2006.06.020

Chen, J. L., Wilson, C. R., & Tapley, B. D. (2010). The 2009 exceptional Amazon flood and interannual terrestrial water storage change observed by GRACE. *Water Resources Research*, *46*(12), 1–10. https://doi.org/10.1029/2010WR009383

Chow, V. T. (1959). Open-channel hydraulics. *McGraw-Hill Civil Engineering Series*.

Dutra, E., Balsamo, G., & Calvet, J.-C. (2017). Report on the improved Water Resources Reanalysis, (June), 94. https://doi.org/10.13140/RG.2.2.14523.67369

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., et al. (2007). The Shuttle Radar Topography Mission. *Reviews of Geophysics*, *45*(2), RG2004. https://doi.org/10.1029/2005RG000183

García-Díez, M., Fernández, J., & Vautard, R. (2015). An RCM multi-physics ensemble over Europe: multi-variable evaluation to avoid error compensation. *Climate Dynamics*, *45*(11–12), 3141–3156. https://doi.org/10.1007/s00382-015-2529-x

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, *377*(1–2), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Hawker, L., Rougier, J., Neal, J., Bates, P., Archer, L., & Yamazaki, D. (2018). Implications of Simulating Global Digital Elevation Models for Flood Inundation Studies. *Water Resources Research*, *54*(10), 7910–7928. https://doi.org/10.1029/2018WR023279

Hawker, L., Bates, P., Neal, J., & Rougier, J. (2018). Perspectives on Digital Elevation Model (DEM) Simulation for Flood Modeling in the Absence of a High-Accuracy Open Access Global DEM. *Frontiers in Earth Science*, *6*(December), 1–9. https://doi.org/10.3389/feart.2018.00233

Hawker, L., Neal, J., & Bates, P. (2019). Accuracy assessment of the TanDEM-X 90 Digital Elevation Model for selected floodplain sites. *Remote Sensing of Environment*, *232*(November 2018). https://doi.org/10.1016/j.rse.2019.111319

Hess, L. L., Melack, J. M., Novo, E. M. L. M., Barbosa, C. C. F., & Gastil, M. (2003). Dual-season mapping of wetland inundation and vegetation for the central Amazon basin. *Remote Sensing of Environment*, *87*(4), 404–428. https://doi.org/10.1016/j.rse.2003.04.001

Jarihani, A. A., Callow, J. N., McVicar, T. R., Van Niel, T. G., & Larsen, J. R. (2015). Satellite-derived Digital Elevation Model (DEM) selection, preparation and correction for hydrodynamic modelling in large, low-gradient and data-sparse catchments. *Journal of Hydrology*, *524*, 489–506. https://doi.org/10.1016/j.jhydrol.2015.02.049

Jung, H. C., & Jasinski, M. F. (2015). Sensitivity of a floodplain hydrodynamic model to satellite-based DEM scale and accuracy: Case study-the Atchafalaya Basin. *Remote Sensing*, *7*(6), 7938–7958. https://doi.org/10.3390/rs70607938

Kirchner, J. W. (2006). Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. *Water Resources Research*, *42*(3), 1–5. https://doi.org/10.1029/2005WR004362

Lehner, B., Verdin, K., & Jarvis, A. (2008). New global hydrography derived from spaceborne elevation data. *Eos*, *89*(10), 93–94. https://doi.org/10.1029/2008EO100001

Liu, K., Song, C., Ke, L., Jiang, L., Pan, Y., & Ma, R. (2019). Global open-access DEM performances in Earth's most rugged region High Mountain Asia: A multi-level assessment. *Geomorphology*, *338*, 16–26. https://doi.org/10.1016/j.geomorph.2019.04.012

Meade, R. H., Rayol, J. M., Da Conceicão, S. C., & Natividade, J. R. G. (1991). Backwater effects in the Amazon River basin of Brazil. *Environmental Geology and Water Sciences*, *18*(2), 105–114. https://doi.org/10.1007/BF01704664

Meyer Oliveira, A., Fleischmann, A. S., & Paiva, R. C. D. (2021). On the contribution of remote sensing-based calibration to model hydrological and hydraulic processes in tropical regions. *Journal of Hydrology*, 126184. https://doi.org/10.1016/j.jhydrol.2021.126184

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, *10*(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Nossent, J., & Bauwens, W. (2012). Application of a normalized Nash-Sutcliffe efficiency to improve the accuracy of the Sobol' sensitivity analysis of a hydrological model. *Eguga*, *14*(2011), 237.

Paiva, R. C. D., Collischonn, W., & Tucci, C. E. M. (2011). Large scale hydrologic and hydrodynamic modeling using limited data and a GIS based approach. *Journal of Hydrology*, *406*(3–4), 170–181. https://doi.org/10.1016/j.jhydrol.2011.06.007

Paiva, R. C. D., Collischonn, W., & Buarque, D. C. (2013). Validation of a full hydrodynamic model for large-scale hydrologic modelling in the Amazon. *Hydrological Processes*, *27*(3), 333–346. https://doi.org/10.1002/hyp.8425

Papa, F., Prigent, C., Aires, F., Jimenez, C., Rossow, W. B., & Matthews, E. (2010). Interannual variability of surface water extent at the global scale, 1993-2004. *Journal of Geophysical Research Atmospheres*, *115*(12), 1–17. https://doi.org/10.1029/2009JD012674

Patro, S., Chatterjee, C., Singh, R., & Raghuwanshi, N. S. (2009). Hydrodynamic modelling of a large flood-prone river system in India with limited data. *Hydrological Processes*, *23*(19), 2774–2791. https://doi.org/10.1002/hyp.7375

Richey, J. E., Mertes, L. A. K., Dunne, T., Victoria, R. L., Forsberg, B. R., Tancredi, A. C. N. S., & Oliveira, E. (1989). Sources and routing of the Amazon River Flood Wave. *Global Biogeochemical Cycles*, *3*(3), 191–204. https://doi.org/10.1029/GB003i003p00191

Rizzoli, P., Martone, M., Gonzalez, C., Wecklich, C., Borla Tridon, D., Bräutigam, B., et al. (2017). Generation and performance assessment of the global TanDEM-X digital elevation model. *ISPRS Journal of Photogrammetry and Remote Sensing*, *132*, 119–139. https://doi.org/10.1016/j.isprsjprs.2017.08.008

Sampson, C. C., Smith, A. M., Bates, P. D., Neal, J. C., & Trigg, M. A. (2016). Perspectives on Open Access High Resolution Digital Elevation Models to Produce Global Flood Hazard Layers. *Frontiers in Earth Science*, *3*. https://doi.org/10.3389/feart.2015.00085

Sanders, B. F. (2007). Evaluation of on-line DEMs for flood inundation modeling. *Advances in Water Resources*, *30*(8), 1831–1843. https://doi.org/10.1016/j.advwatres.2007.02.005

Santos da Silva, J., Calmant, S., Seyler, F., Rotunno Filho, O. C., Cochonneau, G., & Mansur, W. J. (2010). Water levels in the Amazon basin derived from the ERS 2 and ENVISAT radar altimetry missions. *Remote Sensing of Environment*, *114*(10), 2160–2181. https://doi.org/10.1016/j.rse.2010.04.020

Siqueira, V. A., Paiva, R. C. D., Fleischmann, A. S., Fan, F. M., Ruhoff, A. L., Pontes, P. R. M., et al. (2018). Toward continental hydrologic–hydrodynamic modeling in South America. *Hydrology and Earth System Sciences Discussions*, 1–50. https://doi.org/10.5194/hess-2018-225

Stisen, S., Koch, J., Sonnenborg, T. O., Refsgaard, J. C., Bircher, S., Ringgaard, R., & Jensen, K. H. (2018). Moving beyond run-off calibration—Multivariable optimization of a surface–subsurface–atmosphere model. *Hydrological Processes*, *32*(17), 2654–2668. https://doi.org/10.1002/hyp.13177

Tadono, T., Takaku, J., Tsutsui, K., Oda, F., Nagai, H., & Build, T. R. (2015). STATUS OF " ALOS WORLD 3D (AW3D)" GLOBAL DSM GENERATION Japan Aerospace Exploration Agency Remote Sensing Technology Center of Japan NTT DATA Corporation. *IEEE International Geoscience and Remote Sensing Symposium (IGARSS) 2015*, 3822–3825.

Weedon, G. P., Balsamo, G., Bellouin, N., Gomes, S., Best, M. J., & Viterbo, P. (2014). Data methodology applied to ERA-Interim reanalysis data. *Water Resources Research, 50*, 7505–7514. https://doi.org/10.1002/2014WR015638.Received

Yamazaki, D., Oki, T., & Kanae, S. (2009). Deriving a global river network map and its sub-grid topographic characteristics from a fine-resolution flow direction map. *Hydrology and Earth System Sciences, 13*(11), 2241–2251. https://doi.org/10.5194/hess-13-2241-2009

Yamazaki, Dai;, Revel, M., Zhou, X., & Nitta, T. (2021). global-hydrodynamics/CaMa-Flood_v4: CaMa-Flood (Version v4.00). Zenodo. https://doi.org/http://doi.org/10.5281/zenodo.4609655

Yamazaki, Dai, Kanae, S., Kim, H., & Oki, T. (2011). A physically based description of floodplain inundation dynamics in a global river routing model. *Water Resources Research*, *47*(4). https://doi.org/10.1029/2010WR009726

Yamazaki, Dai, Lee, H., Alsdorf, D. E., Dutra, E., Kim, H., Kanae, S., & Oki, T. (2012). Analysis of the water level dynamics simulated by a global river model: A case study in the Amazon River. *Water Resources Research*, *48*(9), 1–15. https://doi.org/10.1029/2012WR011869

Yamazaki, Dai, De Almeida, G. A. M., & Bates, P. D. (2013). Improving computational efficiency in global river models by implementing the local inertial flow equation and a vector-based river network map. *Water Resources Research, 49*(11), 7221–7235. https://doi.org/10.1002/wrcr.20552

Yamazaki, Dai, O'Loughlin, F., Trigg, M. A., Miller, Z. F., Pavelsky, T. M., & Bates, P. D. (2014). Development of the Global Width Database for Large Rivers. *Water Resources Research, 50*(4), 3467–3480. https://doi.org/10.1002/2013WR014664

Yamazaki, Dai, Ikeshima, D., Tawatari, R., Yamaguchi, T., O'Loughlin, F., Neal, J. C., et al. (2017). A high-accuracy map of global terrain elevations. *Geophysical Research Letters*, *44*(11), 5844–5853. https://doi.org/10.1002/2017GL072874

Yamazaki, Dai, 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

Zhou, X., Ma, W., Echizenya, W., & Yamazaki, D. (2020). Uncertainty in flood frequency analysis of hydrodynamic model simulations. *Natural Hazards and Earth System Sciences*, (August), 1–31. https://doi.org/10.5194/nhess-2020-272

The English in this document has been checked by at least two professional editors, both native speakers of English. For a certificate, please see:

http://www.textcheck.com/certificate/N6oGzy

The above statement is here to inform reviewers—who may not be native speakers of English—that the English in this document has been professionally checked.