A multi-system comparison of forecast flooding extent using a scale-selective approach

Hydrology Research

Hydrology Research Vol 54 No 10, 1115 doi: 10.2166/nh.2023.025

A multi-system comparison of forecast flooding extent using a scale-selective approach

Helen Hooker [a,*], Sarah L. Dance[a,b,c], David C. Mason[a], John Bevington[d] and Kay Shelton[d]

[a] Department of Meteorology, University of Reading, Reading, UK
[b] Department of Mathematics and Statistics, University of Reading, Reading, UK
[c] National Centre for Earth Observation (NCEO), Reading, UK
[d] Jeremy Benn Associates Limited (JBA Consulting), Skipton, UK
*Corresponding author. E-mail: h.hooker@pgr.reading.ac.uk

HH, 0000-0002-5135-3952

ABSTRACT

Fluvial flood forecasting systems increasingly couple river discharge to a flood map library or a real-time hydrodynamic model to provide forecast flood maps to humanitarian agencies. The forecast flood maps can be linked to potential impacts to inform forecast-based financing schemes. We investigate a new application of scale-selective verification by evaluating three flood forecasting systems. Two simulation library systems, Flood Foresight (30 m) and GloFAS Rapid Flood Mapping (1,000 m) and one hydrodynamically modelled system, the Bangladesh Flood Forecasting and Warning Centre (FFWC) Super Model (300 m), all made predictions of flooding extent at different spatial scales (grid lengths, in brackets) for the Jamuna River flood, Bangladesh, July 2020. The flood maps are validated against synthetic-aperture-radar-derived observations of flooding using a scale-selective approach that can compare directly across different spatial scales. At short forecast lead times, the Super Model outperforms the other systems. Near to the Bangladesh border, the trans-boundary benefits of the two global systems are evident. We find that scale-selective methods can quantify the skill of systems operating at different spatial scales so that the benefits and limitations can be evaluated. Multi-system comparison of flood maps is important for improving impact-based forecasts and ensuring funds and response activities are appropriately targeted.

Key words: flood forecasting systems, SAR, scale-selective verification

HIGHLIGHTS

- Through a new application of our scale-selective validation method we compare flood maps of different spatial scales (grid lengths) with SAR-derived observations.
- Three flood forecasting systems (two global ensemble and one local deterministic) operating in Bangladesh, July 2020 are evaluated.
- The results show the importance of accounting for spatial scale when interpreting skill scores in multi-system studies.

1. INTRODUCTION

Flood forecasting systems are increasingly used to improve preparedness ahead of a major flooding event (Stephens & Cloke 2014; Wu *et al.* 2020). One of the main action points from the recent Global Assessment Report (GAR2022) on Disaster Risk Reduction (DRR) is to 'design systems to factor in how human minds make decisions about risk' (UNDRR 2022). While flood forecasting systems have improved significantly and continue to improve both globally and locally, the reliance on government departments and disaster managers to make the right decisions when faced with a potential crisis can result in inappropriate actions and unpreparedness (e.g. Fekete & Sandholz 2021; Coughlan de Perez *et al.* 2022). The GAR2022 report shows that just 5.8% ($5.5 billion USD) of official development assistance contributes to disaster prevention and preparedness compared to 90.1% ($119.8 billion USD) for emergency response. Yet it has been demonstrated (for Europe) that financing for mitigation purposes such as flood forecasting systems can lead to overall cost savings (Pappenberger *et al.* 2015).

Forecast-based financing (FbF) schemes can form a major element of DRR strategies and aim to directly link forecasts of extreme events to humanitarian actions (Coughlan de Perez *et al.* 2015, 2016). FbF schemes work by quantifying risks in advance of crises or disasters, prepositioning funds, and agreeing in advance how funds will be released based on forecasts,

ahead of an event (OCHA 2020). Anticipation and risk financing allows humanitarians to be better prepared by making important decisions before disaster strikes. These proactive decisions, directly linked to action (rapid fund release) remove the potential for reactive, incorrect decisions in the midst of a disaster. The success of the FbF system largely depends on the threshold triggers set and on the performance of the flood forecasting system at mapping the flood hazard.

Advances in flood forecasting systems both at global and local levels link together meteorological and hydrological forecasts to hydrodynamic models, simulating flood-wave propagation (Emerton *et al.* 2016; Wu *et al.* 2020; Apel *et al.* 2022). The resulting flood maps when directly linked to impacts can be used to inform DRR schemes. Multiple trade-offs exist in the development of such systems that inherently depend on observation data availability and computing power. These determine whether the maps can be modelled in real-time or are pre-calculated and form part of a simulation library; the spatial scale (grid size) of the forecast flood maps and whether the maps are deterministic or probabilistic (Savage *et al.* 2016). A recent review of flood inundation prediction (Bates 2022) states that a key task to drive forward the development of better global hydraulic models will be more rigorous and comprehensive validation. Hoch & Trigg (2019) outline a Global Flood Model Validation Framework which includes a recommendation to routinely validate flood extent. Quantitative performance evaluation forms an important part of fitness-for-purpose assessment and continual system improvement. Currently, there is limited quantitative validation of operational flood forecasting systems producing flood maps and operating in the same area. A recent advancement in flood map validation (Hooker *et al.* 2022) means that quantitative comparisons can be made across flood maps at different spatial scales, which makes a multi-system evaluation possible.

The accuracy of ensemble forecasts of flood extent can be verified by comparing with observations of flooding from unmanned aerial vehicles or satellite-based sensors. Satellite-based synthetic-aperture radar (SAR) sensors are active, which means they can operate at night, through cloud and weather, and are well known for their flood detection capability (e.g. Horritt *et al.* 2001; Mason *et al.* 2012; Schumann *et al.* 2022). The SAR backscatter intensity depends on the smoothness of the surface, with unobstructed flooded areas returning low backscatter values. Recent techniques used to extract flood extent from SAR images have led to improved flood detection in urban areas (Mason *et al.* 2018, 2021a, 2021b). Since late 2021, SAR-derived flood maps are produced for every Sentinel-1 image detecting flooding around the world by the Global Flood Monitoring (GFM) service (EU Science Hub 2021; GFM 2021; Hostache 2021), part of the Copernicus Emergency Management Service (CEMS) (Copernicus Programme 2021). Within 8 h of the Sentinel-1 image acquisition, three flood detection algorithms are combined to give the flood class and uncertainty estimation per grid cell.

In this paper, through application of a new approach, we evaluate the inundation accuracy of three fluvial flood forecasting systems operating during severe flooding of the Jamuna River in Bangladesh, July 2020. The flood maps are compared against satellite SAR-derived flooding extent. The systems are: Flood Foresight, an FbF system run by JBA Consulting working in partnership with the Start Network (Revilla-Romero *et al.* 2017); the Global Flood Awareness System (GloFAS); Rapid Flood Mapping (RFM) service (Alfieri *et al.* 2013; GloFAS 2022a); and the Bangladesh Flood Forecasting and Warning Centre (FFWC) Super Model (BWDB 2020), each producing forecast flood maps at different spatial scales (30, 1,000, and 300 m, respectively). We investigate how a novel scale-selective spatial verification approach (Hooker *et al.* 2022) can be applied to multi-system studies where the forecast flood maps are presented at different spatial scales. Through applying this approach, we determine a skillful scale of each flood map that can be directly compared and discuss the benefits and limitations of each forecast system for flood mapping purposes that underpin the triggering of FbF schemes.

We describe the characteristics of the Jamuna River flood, July 2020 in Section 2. The three flood forecasting systems are described in Section 3 along with details of the SAR-derived observation of flooding. The scale-selective methods used to evaluate the forecast flood maps are outlined in Section 4. The performance of Flood Foresight with forecast lead time is presented in Section 5.1 and the multi-system flood map comparisons are presented in Section 5.2. In Section 5.3, we discuss the benefits and limitations of each system and conclude with recommendations in Section 6.

2. FLOOD EVENT ON THE JAMUNA RIVER, BANGLADESH JULY 2020

Bangladesh lies on the world's largest delta that drains the Tibetan plateau and the Himalayas to the north via the Brahmaputra and Ganges River systems. The river system capacity in Bangladesh is overwhelmed each monsoon season by the volume of water travelling through. Flooding is exacerbated in coastal regions where tidal surges from tropical cyclones impede the river drainage (Bernard *et al.* 2022). Due to its geographical location and the low lying, low slope nature of

the land, Bangladesh is susceptible and vulnerable to flooding year after year and faces a worsening situation as a result of climate change and sea level rise (Hossain *et al.* 2021).

This study focuses on the Brahmaputra River (locally named the Jamuna River), which is characterised by braided, meandering channels that migrate continually due to frequent silting and erosion, particularly during flood events. The total length of the Brahmaputra is 2,900 km with a catchment area of around 583,000 km^2. Several flashy tributaries such as the Teesta join the main channel in the north from steep catchments in the southern Himalayas. The main distributary of the Jamuna River is the Old Brahmaputra, described by the Bangladesh FFWC as a high flow spill river contributing largely to flooding, depending on the variations of siltation at the entry point (BWDB 2020).

Bangladesh is divided locally into administrative districts, locally named zilas. Four of these that align along the Jamuna River have been chosen for the comparison of flood mapping skill, these are Kurigram, Gaibandha, Jamalpur, and Sirajganj (Figure 1).

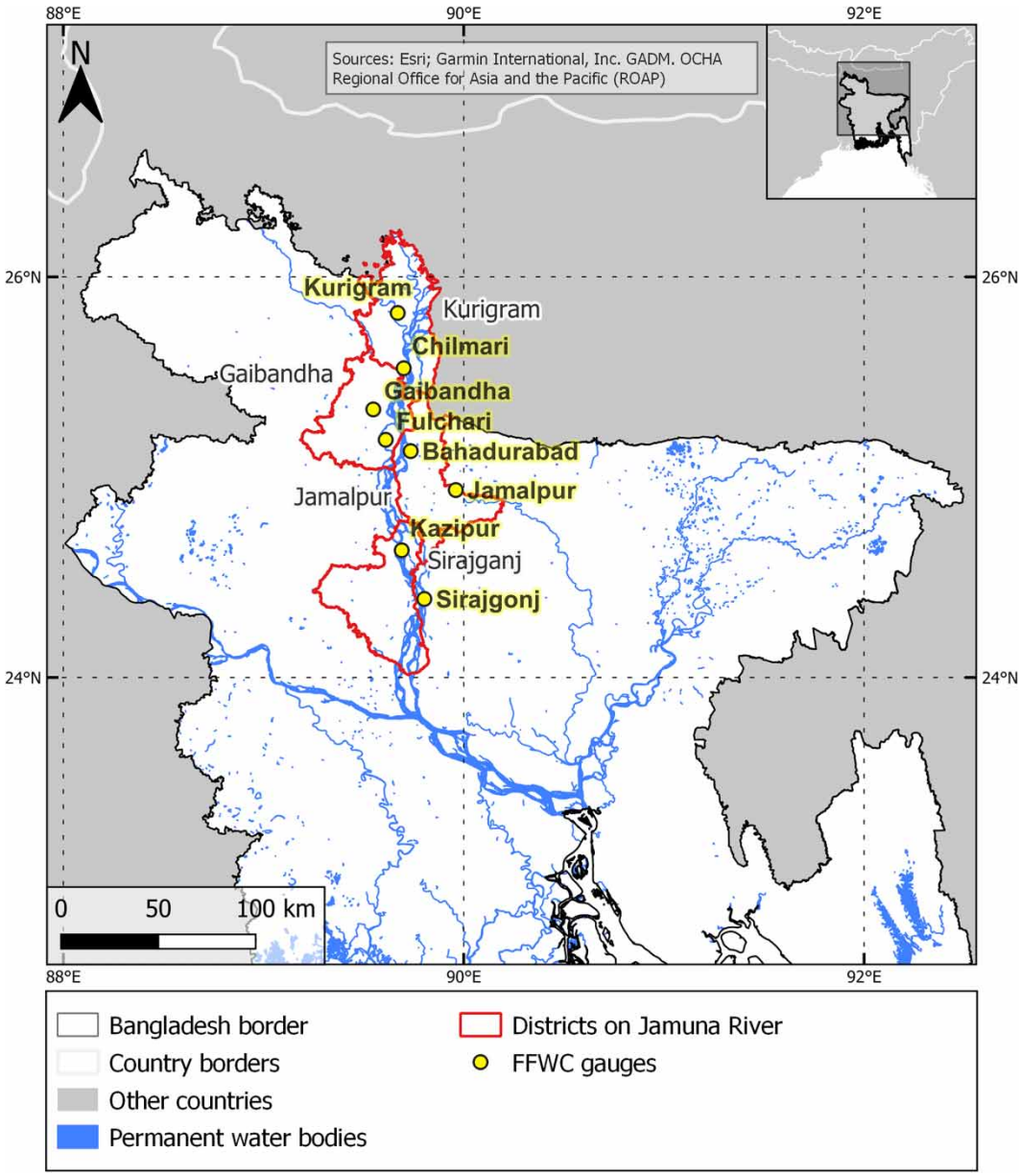

Figure 1 | Four districts (zilas) of interest in the Jamuna catchment in northern Bangladesh.

Bangladesh experienced an active monsoon season during the summer of 2020 which brought severe and prolonged flooding in multiple spells. An unusually wet May following cyclone Amphan meant that water levels were already raised ahead of the monsoon season (Hossain 2020). According to the Bangladesh Water Development Board (BWDB) the flooding had some remarkable characteristics. It began earlier than usual in late June and had a triple peak that had never been seen before. The flooding affected 40% of the country, inundating over 34,000 km^2. In 2020, this resulted in the second highest level of flooding since 1989 and the second longest flood duration since 1998. An estimated 5.5 million people were affected with 1 million houses waterlogged. Around 1.1 million people were displaced with almost 100,000 evacuated to over 1,500 shelters. Almost 1 million tube-wells and more than 100,000 latrines were damaged, 83,000 hectares of paddy fields were affected, and 257 people lost their lives.

The FFWC estimates that the Jamuna basin received 20% more rainfall in July than normal (BWDB 2020). Gaibandha recorded the highest 1-day maximum rainfall across the basin at 250 mm, with a 10-day consecutive maximum rainfall of 549.5 mm. The heavy monsoon rainfall in July caused two flood peaks in one month, the first peak around 15 July and the second around 25 July.

3. FLOOD FORECASTING SYSTEMS AND DATA

The focus of this multi-system comparison is to evaluate the performance of three systems at forecasting the flood inundation extent for the second flood peak on 25 July. In this section, we briefly outline these three flood forecasting systems and summarise their similarities and differences in simulating flood inundation extent (Sections 3.1, 3.2, and 3.3). Table 1 details the main system attributes. The flood map data used for comparison from each system are described in Section 3.4.

Table 1 | Flood forecasting system comparison

Attribute	Flood Foresight	GloFAS Rapid Flood Mapping	FFWC
System application	Forecast-based financing for humanitarian early action	Global, broad scale medium range flooding prediction for large river basins	National flood forecasting and warning
System type	Ensemble simulation library (globally scalable)	Global ensemble simulation library (upstream drainage area $>5{,}000$ km^2, river width $>$ 100 m)	Deterministic
Forecast type	Daily, 10-day lead time	Daily, maximum flood extent next 30 days	Daily, 5-day lead time
Meteorological model	ECMWF IFS	ECMWF IFS	BMD (WRF)
Hydrological model	LISFLOOD	LISFLOOD	MIKE II FF rainfall-runoff
Hydraulic model	RFLOW/JFLOW	CA2D	MIKE II GIS
Observed driving/input data	None	None	Rainfall and river water level
Grid length (m)	30	1,000	300
DSM/DEM	NEXTMAP World30 DSM	SRTM (adjusted)	Survey of Bangladesh (>10 years old)
DSM/DEM grid resolution (m)	30	90 (re-scaled to 1,000)	300
Modelled flood map return period thresholds (years)	20, 50, 100, 200, 500 and 1,500 plus 30 interpolated flood maps	10, 25, 50, 100, 250, 500 and 1,000	N/A
Flood map selection	$ens_{any}>$ 5year return period threshold	$ens_{mean}>$ 10-year return period threshold	N/A
Defences included?	No	No	Yes

Any ensemble member forecast discharge value is defined as ens_{any}. The mean forecast discharge value of all ensemble members is defined as ens_{mean}.

3.1. GloFAS RFM

GloFAS couples state-of-the-art numerical weather prediction (NWP) forecasts with a distributed hydrological model. With its continental scale set-up, it provides downstream countries with forecasts of upstream river conditions up to one month ahead as well as continental and global overviews for large river basins. As of version 3.1 (released in May 2021), this modelling chain is based on the full configuration of the LISFLOOD hydrological model, forced by an ensemble of meteorological inputs (GloFAS 2022a). The meteorological forecast data are provided to LISFLOOD by the ECMWF Integrated Forecasting System (IFS), the operational 51 ensemble member NWP system from ECMWF (Alfieri *et al.* 2013; GloFAS 2022a). The NWP data (precipitation, temperature, potential evapotranspiration, and evaporation rates for open water and bare soil surfaces) are taken as inputs into the hydrological model, LISFLOOD. LISFLOOD is a distributed hydrological rainfall-runoff model, simulating surface, groundwater and subsurface water flow and then routing the water to river channels and simulating the routing of the channel flow (LISFLOOD 2022). LISFLOOD includes consideration of snow melt, infiltration, vegetation interactions (interception, evapotranspiration, water uptake) and exchange of soil moisture between a 3-layer soil water balance sub-model. The runoff data produced are routed through a representation of the river network using a double kinematic wave approach. The river network used is taken from the HydroSHEDS dataset (Lehner 2014). GloFAS is calibrated using historical streamflow records from selected stations worldwide (Hirpa *et al.* 2018). For Bangladesh, four river gauges have been used to calibrate GloFAS as reported on the GloFAS web viewer (GloFAS 2022c). Two of these gauges are on river reaches that would impact the four Jamuna River districts: one at Bahadurabad on the main Jamuna River, and another upstream at Kaunia on the Teesta River, a tributary of the Jamuna River. The observation record at Kaunia is very short at around 7 years (1985–1992), while the record at Bahadurabad is over 35 years (1981–2015). Modified Kling-Gupta Efficiency (KGE) is calculated for each station; a performance measure that indicates how well the model reanalysis (at day 0) replicates the flows observed, greater than 0.8 is very good and less than 0.2 is very poor. Both stations have KGE values above 0.7, indicating relatively good hydrological model performance.

GloFAS RFM (GloFAS 2022b) displays a maximum flood extent over the next 30 days by matching the return periods from the GloFAS streamflow forecast to a catalogue of modelled inundation extents. Flood maps are triggered for basins greater than 5,000 km^2 and where the 10-year RP threshold is exceeded by the ensemble mean. The RP flood maps available are listed in Table 1. The flood maps were developed using the semi-inertial formulation of the CA2D hydraulic model which is a reduced complexity model based on the cellular automata approach and the diffusive wave equations, specifically designed to simulate flood inundation events involving wide areas (Dottori & Todini 2011). Dottori *et al.* (2016) describe the methods used to derive the flood maps at specified return periods on a global scale using a vegetation corrected version of the global DEM SRTM (Farr *et al.* 2007). The hydraulic modelling was performed at 1 km grid resolution.

3.2. Flood Foresight

The Start Network (Start Network 2022) is a charity and network of over 80 humanitarian agencies and aims to develop locally led early action by moving to a model of proactive funding to alleviate crises before they happen. JBA Consulting, in partnership with the Start Network, have developed a Disaster Risk Financing (DRF) system for the Jamuna River that links a fluvial probabilistic flood inundation forecasting system, Flood Foresight (Revilla-Romero *et al.* 2017), to populations impacted by flooding (Figure 2). The DRF system quantifies the flood risk to the population for the purposes of setting trigger threshold levels through a probabilistic global catastrophe risk model, FLY (Dunning 2019).

A domain of interest is divided into 'Impact Zones' (IZ) or sub-catchments using the HydroBASINS dataset (Lehner 2014). The Flood Foresight system links each IZ to GloFAS grid cells providing 51 ensemble member forecasts of river discharge (Section 3.1). Based on the forecast discharge for each IZ, a precomputed flood map is selected from a simulation library. The flood map library was hydrodynamically modelled using JFlow® (Bradbrook 2006) and RFlow using a detailed DSM at 30 m spatial scale at specified return period (RP) thresholds (detailed in Table 1). These were subsequently linearly interpolated at five intermediate intervals between each RP threshold and extrapolated between zero and the 20-year RP flood map (totalling 36 flood maps). The flood map selected is determined by the RP threshold exceeded within each IZ. An example forecast domain in Figure 3 shows neighbouring IZ trigger flood maps at different RP thresholds and the RP threshold is not exceeded in every IZ.

The simulation library approach enables rapid flood map selection so that the system can be run in near real-time. Where the forecast discharge exceeds a 5-year RP threshold the probabilistic flood maps are triggered and linked to populations impacted. IZ linked to a 2-year RP threshold (Figure 3) will not trigger a flood map due to low confidence in the RP threshold

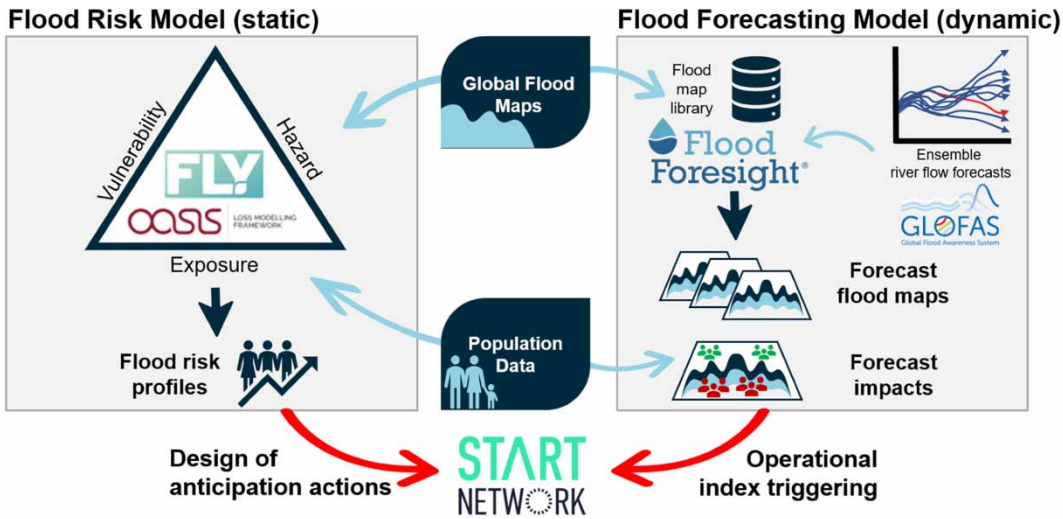

Figure 2 | Flood Foresight/Start Network ensemble flood inundation forecast and population impacts work flow.

levels and the uncertain flood map interpolation process at low discharge values. The system runs daily and produces 51 ensemble member flood extent and depth maps for forecast lead times up to 10 days ahead. Hooker *et al.* (2023a) recommended consideration of all ensemble members as indicators of potential flooding so we evaluate each grid cell where any ensemble member indicates flooding (ens_{any}).

3.3. Bangladesh FFWC

Historical catastrophic flooding in Bangladesh has led to well developed and forward-thinking flood forecasting and warning services. Under the Ministry of Water Resources, flood forecasting in Bangladesh is the responsibility of the BWDB following the BWDB Act-2000. The FFWC, established in 1972, is the lead organisation for flood forecasting and warning services. The FFWC acts as coordinators between other Bangladesh agencies and ministries involved in flood disaster management. During the event in July 2020, FFWC provided a 5-day deterministic and a 10-day probabilistic flood forecast. The FFWC have identified two main priority areas of improvement: to increase warning lead time and to make location specific flood forecasts (BWDB 2020). Operationally, FFWC uses real-time hydrological data from water level and rainfall stations at 3-h intervals. Rainfall estimates are based on the preceding 3 days of rainfall along with analysis derived from NWP forecasts from the Bangladesh Meteorology Department (BMD, NCAR (2022)). Forecast flood bulletins are prepared daily and disseminated through various modes to multiple recipients.

During the monsoon flood season, the FFWC generate a daily hydrodynamically modelled flood inundation map for the Jamuna, Ganges and Meghna river basins. The flood maps are generated using output files from MIKE 11 FF Rainfall-Runoff hydrological model and hydrodynamic modelling simulations using a customised MIKE 11 GIS model (Havnø *et al.* 1995; Gourbesville 1998). The Digital Elevation Model (DEM) used for the hydrodynamic modelling has a 300 m spatial resolution that was collected by the Survey of Bangladesh (SoB) more than 10 years previously.

3.4. Observation data

The data described here are used to evaluate the three flood forecasting systems. Observations of river water level from eight river gauges across the four districts were provided by the FFWC for validation purposes. The gauge locations are shown in Figure 1 in yellow, five are located on the main Jamuna River channel and three are located on tributaries/distributaries of the Jamuna River. Three satellite SAR images from Sentinel-1 (S1A) acquired on 25 July 2020 and three pre-flood images from the same track from 7 June were used to derive a remotely observed flood map. The HASARD flood mapping algorithm (Chini *et al.* 2017) hosted on WASDI (WASDI 2022) uses a statistical, hierarchical split-based approach to separate the two classes (flooded and unflooded) using the pre-flood and flood images. The HASARD mapping algorithm removes permanent water bodies, such as the river channel, reservoirs, and lakes. Flooded areas beneath vegetation, near to buildings and under bridges will not be detected using this method. To smooth the HASARD flood maps and allow a fairer comparison we apply a

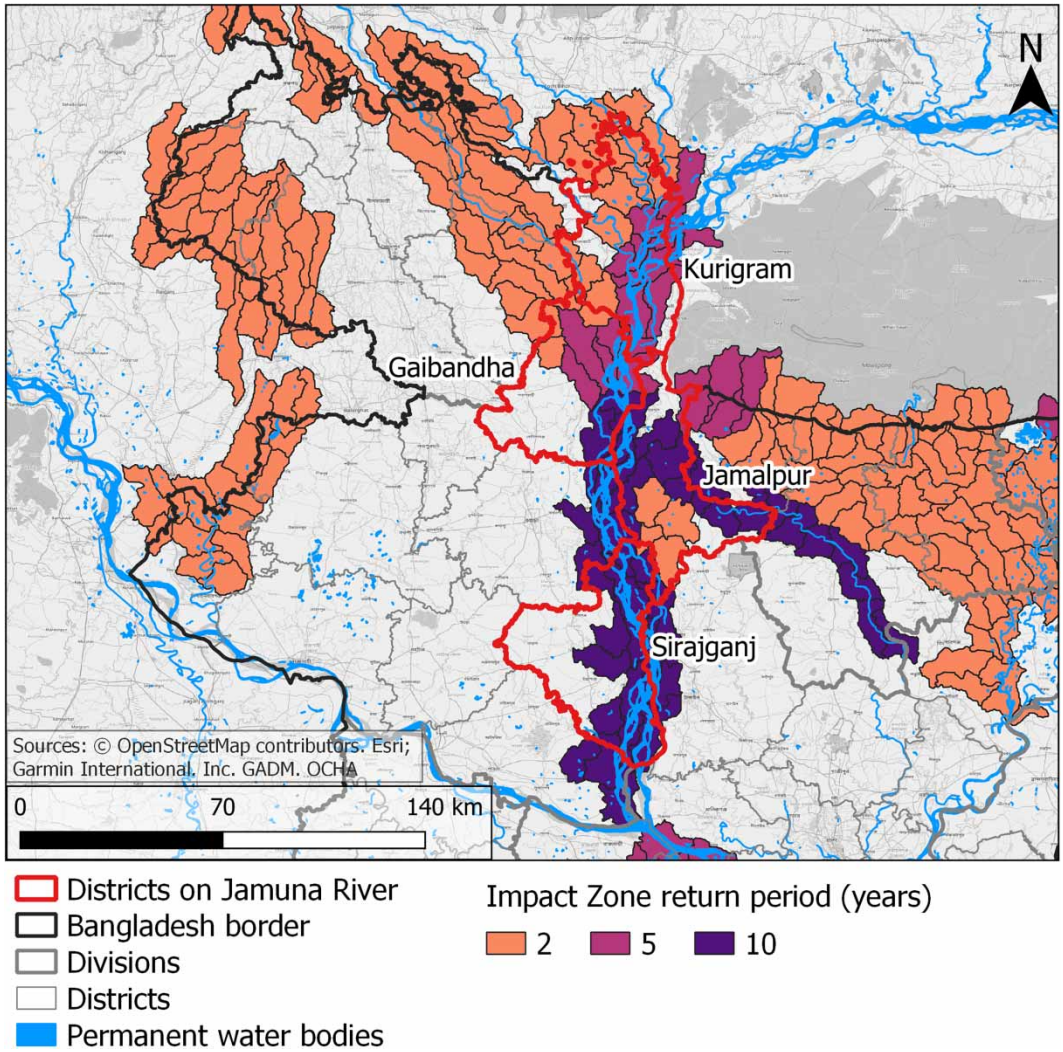

Figure 3 | Example Flood Foresight forecast domain divided into Impact Zones (shown in colour with black outline where the 2-year return period threshold is exceeded), each linked to a GloFAS grid cell. The Impact Zone colour shows the corresponding return period threshold exceeded, determined by the GloFAS forecast discharge.

morphological closing operation, without impacting the location of the flood extent, to flood fill buildings and vegetation. So that the flood prediction accuracy alone can be evaluated, the pre-flood occurrence of surface water using the JRC Global Surface Water database (Pekel *et al.* 2016) has been removed from the forecast inundation maps. The observed flood extent mosaic derived from the three SAR images at 20 m grid size was re-scaled to match the relevant forecast flood map grid lengths using majority (mode) aggregation.

4. SCALE-SELECTIVE EVALUATION METHODS

The flood forecasting systems detailed in Table 1 produce flood maps at 30, 300, and 1,000 m spatial scales (grid lengths). Validation of forecast flood maps against remotely observed flooding extent is typically carried out by labelling each grid cell using a contingency table with categories: correctly predicted flooded, under-prediction (miss), over-prediction (false alarm), and correctly predicted unflooded. After labelling, conventional binary performance measures such as Critical Success Index (CSI) and Pierce Skill Score (PSS) are calculated and give a domain average skill score (Stephens *et al.* 2014). Comparing the binary performance measures of these flood maps at different spatial scales would not be meaningful as the higher resolution maps would be overly penalised due to the double penalty impact (Roberts & Lean 2008; Hooker

et al. 2022). Several commonly applied binary performance measures will be included here for demonstration and comparison purposes only and the details of these can be found in Supplementary material, Appendix A, Table A1.

To tackle the issue of validating across differing spatial scales, we apply a scale-selective approach to flood map evaluation. The scale-selective evaluation approach includes calculation of the Fraction Skill Score (FSS, Roberts & Lean 2008) and location specific agreement scales (Dey *et al.* 2016), which are plotted on a Categorical Scale Map (CSM, Hooker *et al.* 2022, 2023a). A brief summary of the method is given here. For full methodology please see Roberts & Lean (2008); Dey *et al.* (2016); Hooker *et al.* (2022, 2023a). The FSS calculates the accuracy of a forecast flood map by comparing against an observed flood map across a range of neighbourhood lengths (n, see Figure 4). First, every grid cell is compared ($n = 1$). Then, the next largest neighbourhood size, $n = 3$, surrounding each grid cell is compared and the process continues to $n = 5$, $n = 7$ and so on. The fraction flooded (number of flooded grid cells in the neighbourhood divided by the total number of grid cells in the neighbourhood) in each of the forecast and observed flood maps are used to calculate the FSS at each n. The FSS calculation is based on the Brier Skill Score. The FSS for each n is derived by calculating the mean squared error (MSE) between the forecast and observed fractions and dividing this by a reference MSE. This value is subtracted from 1 to give the FSS score. Increasingly larger neighbourhoods are compared until a target FSS score has been reached and exceeded at which point the skillful scale has been met (e.g. see Figure 9). The target FSS score, given by $\mathrm{FSS_T} \geq 0.5 + f_o/2$ depends on the fraction of observed flooding in the domain of interest, f_o.

The FSS gives a domain averaged skill score. We also calculate a local agreement scale (S) for each grid cell that can be mapped onto a CSM. The relationship between S and n is given by $S = (n - 1)/2$. An acceptable level of background bias between the forecast and observed flood maps can be pre-set. This is used to determine an agreement criterion. Like the FSS method, the comparison begins at each grid cell, if the agreement criterion is met, the grid cell is labelled with an agreement scale $S = 0$. Where the criterion is not met, a larger neighbourhood size is compared (e.g. $n = 3$). The fraction flooded in

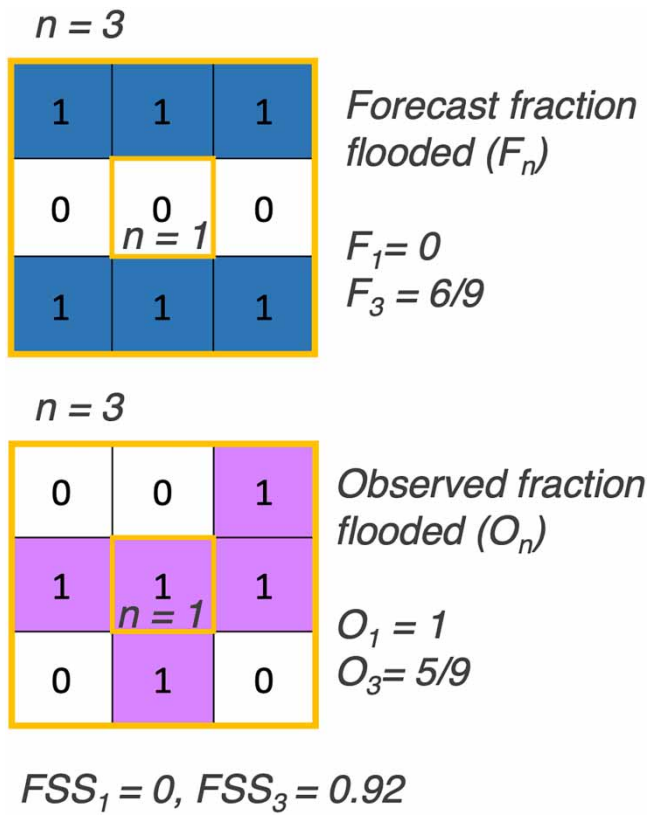

Figure 4 | An example FSS calculation applied to forecast flooding extent, 1 = flooded (in blue), 0 = unflooded (in white) compared to a remotely observed flood extent in pink. The FSS is calculated for two neighbourhood sizes, $n = 1$ (small gold box) and $n = 3$ (large gold box).

each of the forecast and observed flood maps are compared and if the criterion is met the agreement scale assigned would be $S = 1$. The process continues to larger neighbourhoods (e.g. $n = 5$, $S = 2$) until either the criterion is met or a predetermined limit is reached (S_{lim}, set to 9 for this application). The agreement scale at this limit would indicate a miss or false alarm for the grid cell. Combining a map of agreement scales with a conventional contingency map creates a CSM which shows the level of agreement (S) and whether the forecast is over- or under-predicting the flooding extent (e.g. see Figure 8). The CSM shows a location specific skill score that can be linked to different aspects of the forecast system such as IZ and their associated river discharge forecast or RP thresholds, river channel bathymetry, the DSM or flood defences.

5. RESULTS AND DISCUSSION

Forecast flood maps for the Jamuna River flooding, July 2020 from each of the three systems described in Sections 3.2 and 3.3 are compared against SAR-derived flood maps (described in Section 3.4) using scale-selective evaluation methods (as discussed in Section 4). First, in Section 5.1, the Flood Foresight system performance is evaluated against forecast lead time, where flood maps out to 10-day lead time are available. Performance evaluation of the other two systems with forecast lead time was not possible due to the availability of flood maps. In Section 5.2, we compare the forecast flood maps from each of the systems described in Section 3 and discuss their benefits and limitations in Section 5.3.

5.1. Flood Foresight in Jamuna River: a case study

The performance of the Flood Foresight system with forecast lead time is evaluated here. Flood Foresight predicts the flooding extent for second flood peak on 25 July at all lead times (out to 10 days). To evaluate the spatial accuracy within each district with forecast lead time, the absolute agreement scale score has been averaged across each district (solid lines, Figure 5).

An average agreement scale of zero would indicate the forecast and observed fields are in agreement at grid level, $S = 2$ means agreement is reached within a 5×5 neighbourhood. Across the 10-day forecast Flood Foresight performs best in Kurigram district in the north with consistently the worst performance in Sirajganj. Three of the districts (except Jamalpur) show a similar trend with forecast lead time with the best performance for all districts (smallest average agreement scale) occurring at a 10-day lead time (2.31) with a second peak of performance at a 4-day lead time (2.61). The performance worsens from a 4-day lead time to a 1-day lead time across all districts, with an average agreement scale at a 1-day lead time of 3.42. The unusual variation in skill with forecast lead time for Jamalpur prompted further investigation into the driving data.

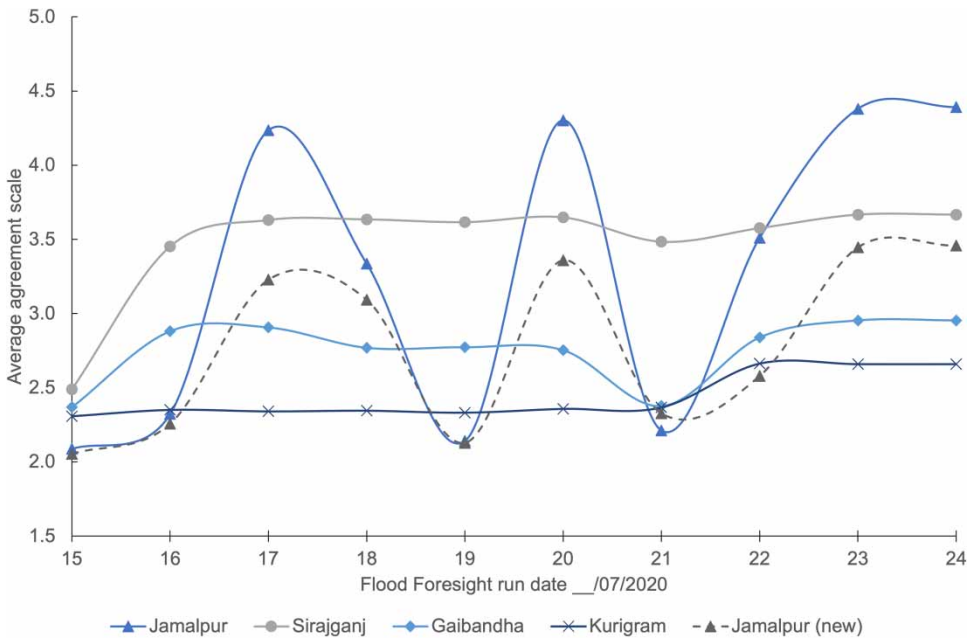

Figure 5 | Flood Foresight average agreement scale against forecast lead time for each district and updated Jamalpur following reassociation of IZ with GloFAS grid cells, forecast valid for 25 July.

Section 3.2 describes how the domain is divided into IZ with each of these linked to the driving data, GloFAS river discharge (Figure 3). The flood map selection within each IZ depends on the forecast discharge exceeding a particular RP threshold. A direct comparison of observed and modelled river conditions is not possible as observed data are river water levels, GloFAS provides forecast discharge, and the stage-discharge relationships are not available. The observed river water levels have been aligned using human judgment to the nearest GloFAS grid cell forecast discharge (1-day lead time control member) and compared for all river gauges. We aimed to match the trend in the two series while keeping the local station risk level close to the GloFAS 2 and 5-year RP threshold levels. Two of the eight river gauges are located in Jamalpur (Figure 1). Bahadurabad is located on the main Jamuna channel and Jamalpur is on a distributary of the Jamuna River crossing the Jamalpur district (Figure 6). Hydrographs for gauges located outside of Jamalpur are plotted in Supplementary material, Appendix A, Fig. A1. Bahadurabad (Figure 6(a)) is well calibrated in GloFAS (Section 3.1) and the overall forecast aligns well with the observed river water level; the second flood peak magnitude is slightly underestimated. The discharge forecast skill for Jamalpur station (Figure 6(b)) is very different to the observed river water level in terms of variation in water levels/discharge, timing and magnitude of flood peaks. A closer look at the river routing network in GloFAS shows that the Old Brahmaputra distributary in Jamalpur is not connected to the main river channel; this connection should occur at flood flows. To overcome the disconnection of the distributary in the river network, the IZ association to GloFAS grid cells in the Flood Foresight system design was manually updated. IZ aligning the Old Brahmaputra, crossing Jamalpur, were reassociated to GloFAS grid cells located upstream on the main Jamuna channel. The new forecast hydrograph for Jamalpur (Figure 6(c)) shows an improved forecast compared to observed river water levels. However, the first flood peak arrives quicker than observed and the second flood peak magnitude is underestimated.

Following the gauge reassociation update to the Flood Foresight system, updated CSMs were reanalysed and this led to an improvement in the overall skill for the Jamalpur district (Figure 5) but with variation in skill with forecast lead time. The variation is investigated by examining the CSMs for Jamalpur. Figure 7(a) maps the original CSM for Jamalpur for run date 20 July. CSMs show a location specific (at each grid cell) agreement scale between the forecast flood map and the observed SAR-derived flood map. There is a large area of under-prediction around Jamalpur caused by the disconnection of the Old Brahmaputra distributary (as discussed above).

The CSM change map (Figure 7(b)) is calculated by taking the difference between the absolute CSM values, |updated CSM| − |original CSM|. The CSM change map for Jamalpur shows where the reassociation has impacted the flood map. A negative agreement scale change indicates an increase in skill (purple, smaller grid size), whereas a positive agreement scale change indicates a decrease in skill (orange, larger grid size). The CSM change values can highlight areas where the agreement scale has increased/decreased but have not reached $S = 0$ (agreement at grid level). Areas surrounding the Old Brahmaputra (in purple) show most of the increased skill following the reassociation. However, there are regions of Jamalpur not impacted by the reassociation that remain under-predicted where smaller distributaries run, that would not be captured/calibrated in GloFAS due to their small size (Figure 7(b)). This results in some variation in skill with forecast lead time remaining after the reassociation (Figure 5). The visualisation of incremental improvements at specific locations will benefit future flood map development work. The CSM change map shows more sensitivity to changes in skill that would not be visible on a conventional contingency map.

5.2. Multi-system flood map comparison

GloFAS RFMs are displayed when the mean ensemble member exceeds the 10-year RP threshold within the next 30 days. By inspecting the reporting point hydrograph at Bahadurabad on each of the 10 days preceding the flood peak (25 July), we found that the 10-year RP threshold was exceeded just once by the ensemble mean on the 18 July (run date), which means only one forecast flood map was available from GloFAS RFM for evaluation. FFWC provided daily flood maps, based on most recent observations, valid for the same day. For the multi-system comparison, based on the availability of forecast flood maps, the following have been selected to compare in detail: Flood Foresight's flood map for the same run date as the available RFM forecast (18 July forecast date, valid date 25 July), the RFM (run date 18 July) and the FFWC flood map for 25 July (run date and forecast valid date) have each been compared to the SAR-derived flood map (25 July) and CSMs calculated (Figure 8). CSMs have also been calculated for all available forecast lead times for the Flood Foresight system (not shown).

The large area of flood under-prediction (in red) to the northwest of Sirajganj on each CSM (Figure 8) can be linked to observed very heavy rainfall and possible surface water flooding (SWF) detected by the SAR data. The recorded rainfall

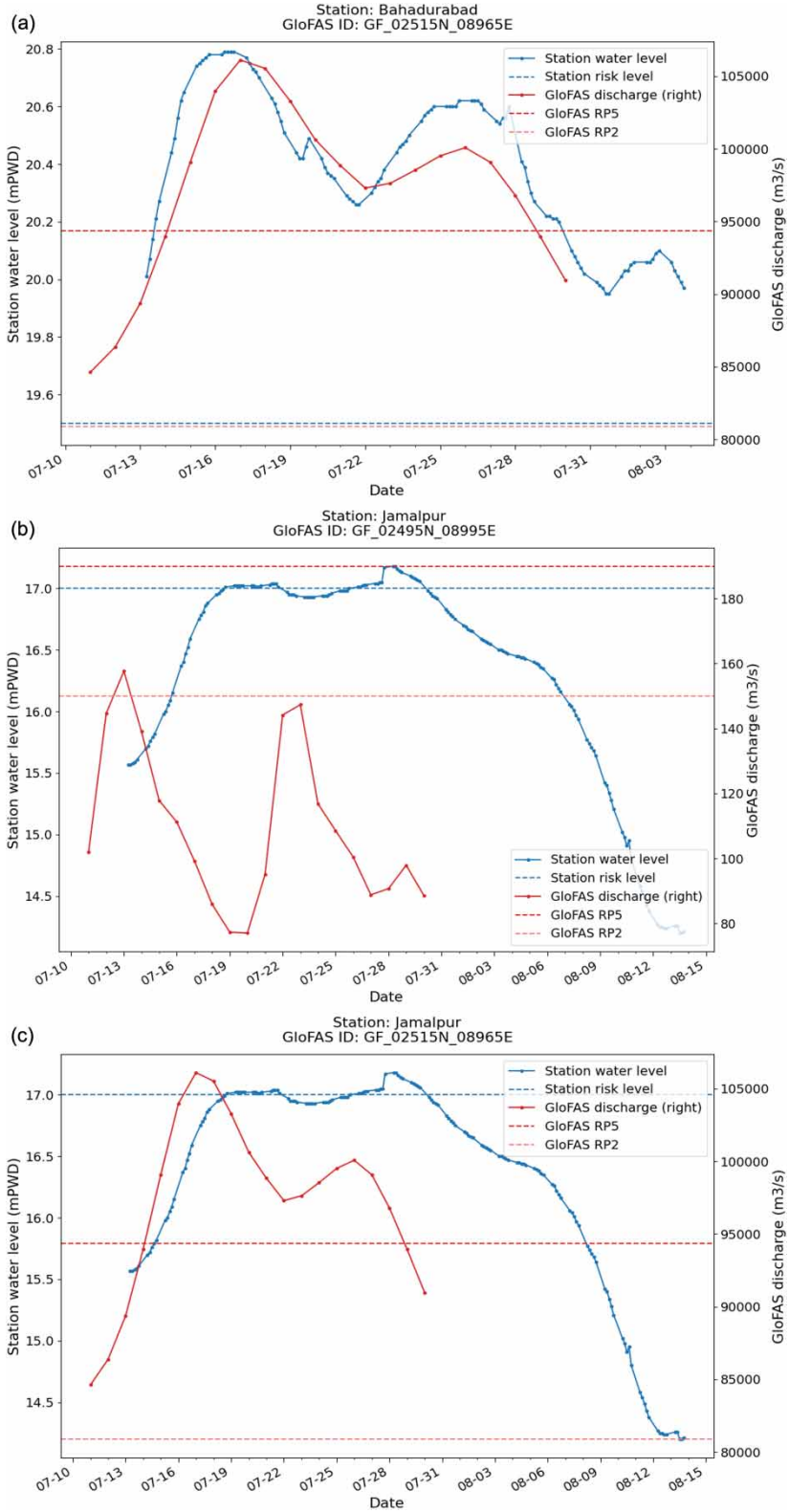

Figure 6 | (a) GloFAS forecast discharge (control member, 1-day lead time) compared to FFWC observed river water level for the main Jamuna channel at Bahadurabad and (b) the old Brahmaputra distributary in the Jamalpur district. (c) The old Brahmaputra distributary forecast discharge following reassociation of IZ with GloFAS grid cells. The GloFAS RP threshold levels are taken from the nearest GloFAS grid cell to the gauge station location. Station risk levels are provided by FFWC.

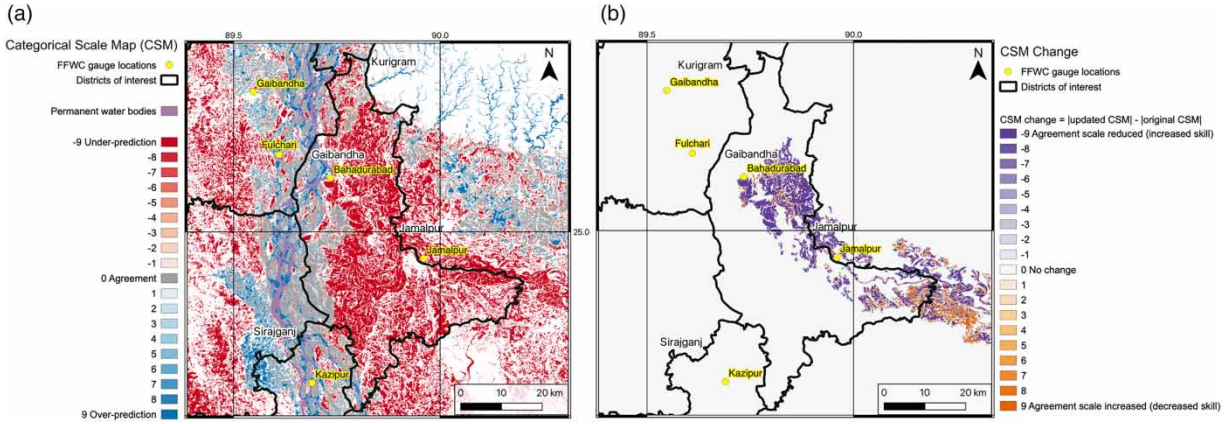

Figure 7 | (a) Original CSM for Jamalpur. (b) CSM change (updated CSM – original CSM) for Jamalpur following reassociation of IZ with GloFAS grid cells. Run date 20 July, forecast valid for 25 July for (a) and (b).

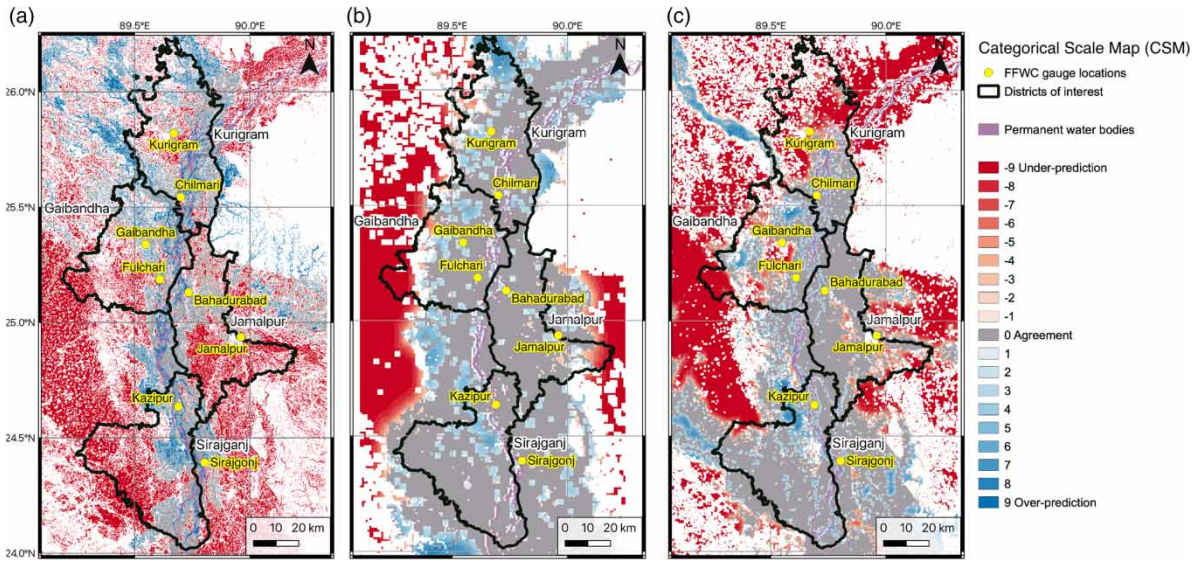

Figure 8 | CSM for flood inundation forecasts from three forecast systems for flood peak 25 July compared to SAR-derived flooding. (a) Flood Foresight run date 18 July, (b) GloFAS RFM run date 18 July, and (c) FFWC flood map run date 25 July.

described earlier (Section 2) is confirmed by rainfall derived from satellite data, (Climate Hazards Group InfraRed Precipitation with Station (CHIRPS) data (Funk *et al.* 2015)), which shows that the July rainfall anomaly in this area was 125–175 mm. The accumulated rainfall in GloFAS for the ensemble mean for the 20 days preceding 25 July amounts to a maximum of 300 mm, with up to 400 mm north of Bahadurabad. This is an under-prediction compared to observed rainfall (Gaibandha 1-day maximum rainfall 250 mm, 10-day consecutive maximum rainfall 549.5 mm), which explains the under-prediction seen in Figure 8(a) and 8(b). Note that each system forecasts fluvial flooding and does not account for SWF, which was likely given the extreme rainfall accumulations recorded. This is a limitation of both Flood Foresight and RFM systems and demonstrates the need for combining fluvial and SWF forecasting systems or combining the forecast flood maps with SAR data using data assimilation (e.g. García-Pintado *et al.* 2015; Cooper *et al.* 2019).

The CSM for Flood Foresight (Figure 8(a)) shows areas of over-prediction (in blue) next to the Jamuna River. It is likely that more of the river channel has been removed from the SAR image during the flood mapping process compared to the

12-month occurrence of water in the Global Surface Water database. Also, the Jamuna River channel migration will also contribute to errors in this area as the DSM was acquired in 2016. To the west of Sirajganj, the FFWC correctly maps flooding associated with the Atrai River, a tributary of the Jamuna River that is not forecast by the other two systems. Multiple tributaries flow across Gaibandha and Sirajganj that are not currently resolved by GloFAS. This also impacts the performance of Flood Foresight when flood maps were not triggered above the 5-year RP threshold.

Flood waters are (Figure 8(b)) spread further from the main river channel in the RFM, compared to the Flood Foresight flood map. This is due to the GloFAS configuration where clusters of cells are linked to the main channel reporting points. The RFM extent is also due to a smoother DEM created by re-scaling from 90 to 1,000 m. This will effectively remove flood barriers such as embankments and roads. An element of smoothing (albeit to a lesser extent compared to the RFM) would also occur in the Flood Foresight flood maps at 30 m grid length. The FFWC CSM shows good accuracy in Sirajganj and Jamalpur (Figure 8(c)). In Sirajganj, there is a region of over-prediction on both the Flood Foresight and the RFM CSM not present on the FFWC CSM. The FFWC model includes flood defences and water level observations from Kazipur, both could contribute to the better performance seen here. FFWC maps perform less well in Kurigram where the flooding extent is underestimated. Kurigram is next to the northern border of Bangladesh where upstream water level data are unavailable for hydrological model calibration and validation. The benefits of the trans-boundary systems of GloFAS and Flood Foresight are evident here.

To quantify the district performance of each system the FSS has been calculated for neighbourhood sizes up to $n = 19$ or larger (not plotted) where the FSS target has not been reached (Figure 9). The FSS gives a measure of spatial accuracy for each system flood map; however, the score is not directly comparable across different spatial scales since the scores are

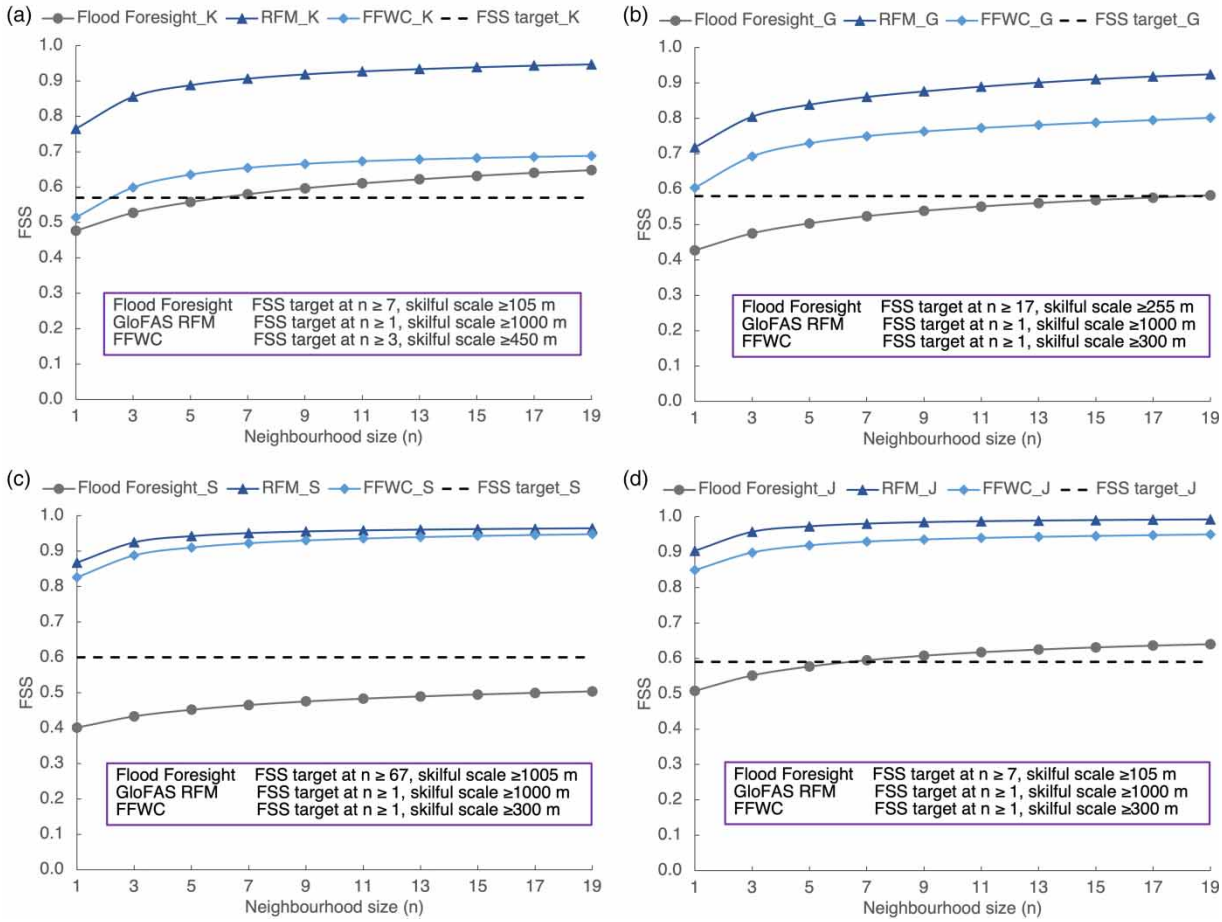

Figure 9 | Average FSS plotted against neighbourhood size (n) for each forecast system (run dates as described in Figure 8) in Kurigram (a), Gaibandha (b), Sirajganj (c). and Jamalpur (d) and the target skill score for each district.

calculated in terms of a neighbourhood size (Section 4). We can calculate a skillful scale, which is half of the neighbourhood size at which the FSS exceeds FSS_T. This accounts for the size of the grid cell and can be directly compared across each of the forecast systems. For example, in Figure 9(a) for Kurigram, at grid level ($n = 1$) the FFWC FSS (0.52) exceeds the Flood Foresight FSS (0.48). We saw on the CSM map (Figure 8) that the Flood Foresight map appeared to capture the observed flooding at Kurigram more accurately than the FFWC map. Despite this observation, the FFWC map has a higher CSI score (0.35, Supplementary material, Appendix A, Fig. A2 (c)) compared to the Flood Foresight CSI (0.31, Supplementary material, Appendix A, Fig. A1(a)) because the grid size is not accounted for by the CSI. However, neither of the two systems flood maps have reached the target FSS_T at grid level (Figure 9(a)). The FFWC map exceeds FSS_T at $n = 3$ and the Flood Foresight map exceeds FSS_T at $n = 7$. By accounting for the impact of the grid size, the skillful scale for Flood Foresight for Kurigram is 105 m ($(1/2)(7 \times 30)$) compared to 450 m ($(1/2)(3 \times 300)$) for the FFWC flood map indicating the Flood Foresight system is more accurate in Kurigram. In Gaibandha, the skillful scale is similar for Flood Foresight and the FFWC maps (255 m versus 300 m, Figure 9(b)). In Sirajganj, Flood Foresight requires a neighbourhood size similar to the grid size of the GloFAS RFM to exceed FSS_T at 1,005 m (Figure 9(c)). Following the reassociation of the IZ in Jamalpur, Flood Foresight has a skillful scale of 105 m here compared to a high FSS score for the FFWC model at 300 m grid level (Figure 9(d)). Across all districts, GloFAS RFM exceeds the FSS_T at grid level with the best performance in Jamalpur and the worst performance in Gaibandha which can be linked to the underestimation of flooding extent seen on the CSM (Figure 8(b)).

The scale-selective approach is also useful for comparing performance above the FSS_T line where two systems exceed FSS_T at grid level ($n = 1$). For example, in Jamalpur (Figure 9(d)) at $n = 1$ the RFM FSS is 0.90 compared to the FFWC FSS of 0.85. However, FFWC reaches the same score as the RFM (at $n = 1$, 0.90), at $n = 3$. In terms of spatial scale, the same FSS is reached by FFWC at $n = 3$ (450 m) as RFM at $n = 1$ (1,000 m), indicating that the FFWC flood map is more accurate than RFM in Jamalpur. The same result is true in Sirajganj (Figure 9(c)) where the FFWC FSS at $n = 3$ (0.89) exceeds the RFM FSS at $n = 1$ (0.87). In Gaibandha (Figure 9(b)) the FFWC FSS at $n = 5$ (0.73, 750 m) exceeds the RFM FSS at $n = 1$ (0.72). Overall, by accounting for grid length, the FFWC provides the most accurate forecast flood map (albeit at a 0-day lead time) in Jamalpur, Sirajganj and Gaibandha. RFM is most skillful in Kurigram with Flood Foresight outperforming the FFWC model here.

5.3. Discussion

GloFAS RFMs are designed to give an early indication, up to a month in advance that severe flooding is possible for large rivers across the globe. The RFMs are triggered when the ensemble mean discharge exceeds the 10-year RP threshold over the next 30 days. The RP threshold levels are calculated using ERA5 reanalysis data (Harrigan *et al.* 2020). GloFAS RFM is triggered for the Jamuna River only once at an 8-day lead time on 18 July (mapping maximum extent over the next 30 days) where the flood map shows a high level of skill across the districts of interest (Figure 8), comparable to the local FFWC flood maps, which use local observations. Unfortunately, the RFM skill reduces closer to the event with no flood maps triggered after 18 July. We also see this impacting the Flood Foresight skill, which reduces closer to the flood peak (Figure 5). The Flood Foresight system indicates flooding surrounding the Jamuna River at all forecast lead times. The flooding extent is generally under-predicted. This is partly due to the discharge forecast not exceeding the 5-year RP threshold, which also links to unresolved/uncalibrated smaller tributaries/distributaries in the GloFAS river network.

Boelee (2022) finds (for Africa, based on GloFAS ensemble-reforecast flows) that the forecast discharge exceedance above RP thresholds depends on both the forecast lead time and the number of ensemble members considered (the probability trigger set). Boelee found more flood occurrences were predicted at medium-range lead times, compared to short-range lead times. We infer that these results could also apply to Bangladesh and that they partly explain the RFMs best performance at 8-day lead time. The ERA5 reanalysis data is used to initialise the GloFAS forecast and determine the RP thresholds. Currently, the RP thresholds do not account for either forecast lead time or ensemble variability. The GloFAS hydrographs at Bahadurabad indicate a reanalysis discharge value of around the 2-year RP threshold for all lead times within 5 days of the flood peak. This is a significant underestimation compared to observed river levels (both in historical context and compared to FFWC danger levels and severe flood thresholds). High confidence is assigned to the initial conditions in the streamflow forecast and also in the short-term forecast before the ensembles show more variation at longer lead times. This leads to an overall under-estimation of the flood magnitude at shorter lead times. Zsoter *et al.* (2020) found that ensemble-reforecast-based thresholds would lead to an improved forecast at lead times beyond a few days as they can account for

variations in forecast skill with lead time and ensemble variability. Ensemble-reforecast-based thresholds could improve the flood map selection in both the RFM and Flood Foresight systems, which presently are the main limitation of both systems.

The RFM uses the ensemble mean or 50% of ensemble members must exceed the RP threshold to trigger a flood map. However, Boelee (2022) found that the percentage of flood events exceeding the threshold for any RP dropped to less than 50% as soon as the required ensemble size was increased to two ensemble members or more, for all the lead times. Extreme flood event prediction can lie in the ensemble member outliers (Hooker *et al.* 2023a). The Flood Foresight system uses information from all ensemble members for impact forecasts, which is a major benefit of this system. The automated probabilistic flood maps can be produced quickly in near real-time indicating a spread of possible conditions that could support the decision-making process. The RFM could see an improved forecast at more lead times if any ensemble member exceeding the RP threshold (rather than the mean) triggered the flood map selection. This would require further investigation in flood prone areas so that the number of ensemble members chosen can be optimised to avoid increasing false alarms. The RFM could also provide probabilistic information by combining flood maps from all ensemble members that exceed the RP threshold.

The FFWC model performs significantly better at shorter lead times compared to the other two systems, which is not surprising as locally observed river level and rainfall data are used as input data. The benefits of the FFWC model are that no RP thresholds need to be determined or exceeded to produce a flood map and there are no flood map interpolation uncertainties. The deterministic FFWC models forecast skill drops significantly with forecast lead time (BWDB 2020), which reduces the usefulness for flood mitigation and FbF purposes at longer lead times. The FFWC model performs less well compared to RFM and Flood Foresight near to the country's border where observations upstream are unavailable.

6. CONCLUSIONS

Forecast flood maps are increasingly sought to accurately link flooding hazard to populations impacted to inform FbF schemes. Humanitarian agencies in Bangladesh would like the impacts mapped in detail at Union level (4,571 Unions in Bangladesh) at long forecast lead times (out to 10 days) so that insurance funds can be locally targeted in good time. This creates a conflict between the detail or spatial scale (grid size) of the flood maps and the forecast skill of the flood forecasting system. Spatial validation of forecast flood maps from multiple systems has received little attention, partly due to the problem of comparing skill scores from maps at different spatial scales. Here, we applied a validation approach using scale-selective methods (Hooker *et al.* 2022) that determines a skillful scale, which can be directly compared across forecast systems.

We evaluated three flood forecasting systems: Flood Foresight (30 m), GloFAS RFM (1,000 m), and the FFWC Super Model (300 m) each predicting flooding extent at different spatial scales (shown in brackets) for the Jamuna River in July 2020. Each of the maps were compared against SAR-derived observations of flooding extent. Evaluating Flood Foresight skill at all lead times out to 10 days for four districts revealed issues with unresolved/uncalibrated tributaries/distributaries in GloFAS (used to input forecast discharge to Flood Foresight). Reconfiguring the Flood Foresight system led to an improved flood map forecast in one district (Jamalpur), but similar issues remained in other areas. This highlights one problem with trying to combine a gridded global hydrometeorological model with a detailed sub-catchment network and linking this to detailed flood maps. The flood mapping skill in the Jamuna basin is linked to the detail of the river network and whether flood maps are triggered, which depends on exceeding the RP threshold set. Where Flood Foresight maps were triggered, such as in Kurigram, the flood map accuracy outperformed the local FFWC model. This is due to its location next to the border of Bangladesh and the lack of upstream observations. In other areas the FFWC model captures more detail in the river network and shows less underprediction compared to the other systems. For FbF applications and humanitarian response in Bangladesh, a combination of Flood Foresight and the local FFWC model could produce flood inundation maps at a useful scale that can be linked to flooding impacts. Flood Foresight has the benefit of forecast skill at a longer lead times (up to 10 days) with probabilistic maps accounting for some of the forecast uncertainty. Flood Foresight could be supplemented or linked to driving data from the FFWC model at shorter lead times to incorporate local observations and a higher resolution river network. This would avoid regions of non-trigger where no flood map is selected from the Flood Foresight library due to the forecast discharge not exceeding the RP threshold. GloFAS RFM is designed as a deterministic 'heads-up' tool at a coarse resolution that would be difficult to link to impacts for FbF applications in its current configuration. All systems miss flooding across a wide area captured by the SAR data that is possibly due to SWF. These fluvial flood forecasting systems are not designed to map SWF and we recommend combining the forecast flood maps with SAR-derived flood maps through data assimilation so that SWF can be accounted for in post event impact calculations for FbF schemes. Alternatively, a combined fluvial/pluvial

flood forecasting system would be ideal; however, pluvial flood forecasting practice is currently less developed in part due to the difficulties in observing SWF and accurately predicting convective rainfall (Speight *et al.* 2021).

For future spatial validation of flood maps using SAR data, we recommend making use of the Copernicus GFM product (GFM 2021) which maps flooding detected from all Sentinel-1 images since October 2021. Importantly, an exclusion mask layer is available which can be used to exclude areas where SAR is unable to detect flooding such as in dense urban areas, under vegetation and near steep topography. The forecast flood maps would no longer be penalised for over-prediction in regions where the SAR cannot reliably detect flooding. Another opportunity for improvement lies with the flood map library. Both Flood Foresight and RFM maps are currently undefended. However, both would likely be improved if flood defence information from FFWC could be incorporated, allowing areas benefiting from those defences to be discounted when flood conditions are at return periods lower than the standard of protection of the defence. Alternatively, a new high resolution DTM including local defence features, ideally through acquiring LiDAR data (where locally obtained), would improve the local accuracy of the flood maps. The maps held within the simulation library could be hydrodynamically precomputed at lower discharge return periods, which would avoid inaccuracies caused by flood map interpolation beneath the lowest RP level. This would increase the confidence in these flood maps so that a lower RP threshold could be used to trigger FbF allowing more people access to insurance funds.

Fortunately, some of the issues discussed here such as the river network detail will be partly resolved by the major upgrade to GloFAS with version 4.0 due in 2023 (Grimaldi 2022). Significantly, the spatial resolution of GloFAS will increase four fold to approximately 5 km grid size. The river network will increase similarly, and more distributaries/tributaries will be included in Bangladesh. This upgrade along with the use of ensemble-reforecast-based RP thresholds should improve the flood map selection process used by both GloFAS RFM and Flood Foresight. Scale-selective validation methods will enable future system changes to be evaluated and compared meaningfully. Ideally, this would be automated and integrated into the flood forecasting system.

ACKNOWLEDGEMENTS

Many thanks to Sazzad Hossain and Tohidul Islam from the Flood Forecasting and Warning Centre in Bangladesh who provided model details, flood maps, and gauge data. Thanks to Calum Baugh from ECMWF who provided the GloFAS Rapid Flood Maps. Thanks to Ashraful Haque and Elizabeth Rees from the Start Network for their support in this project. This work was funded in part by the Natural Environment Research Council as part of a SCENARIO funded PhD project with a CASE award from the JBA Trust (NE/S007261/1). S.L.D. and D.C.M. were funded in part by the UK EPSRC DARE project (EP/P002331/1). S.L.D. also received funding from NERC National Centre for Earth Observation.

AUTHOR CONTRIBUTIONS

J.B. and K.S. provided the forecast data. H.H. wrote the algorithms and ran the experiments, with input from S.L.D., D.C.M., J.B., and K.S. H.H. prepared the manuscript with contributions from all the co-authors.

DATA AVAILABILITY STATEMENT

All relevant data are available from an online repository or repositories. Categorical Scale Maps along with the SAR-derived flood map are available on the following Zenodo page: {https://doi.org/10.5281/zenodo.7509980} \citep{Hooker2023b}.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J. & Pappenberger, F. 2013 GloFAS-global ensemble streamflow forecasting and flood early warning. *Hydrology and Earth System Sciences* **17**, 1161–1175. https://doi.org/10.5194/hess-17-1161-2013.

Apel, H., Vorogushyn, S. & Merz, B. 2022 Brief communication: impact forecasting could substantially improve the emergency management of deadly floods: case study July 2021 floods in Germany. *Natural Hazards and Earth System Sciences* **22**, 3005–3014. https://doi.org/10.5194/nhess-22-3005-2022.

Bates, P. D. 2022 Flood inundation prediction. *Annual Review of Fluid Mechanics* **54**, 287–315. https://doi.org/10.1146/annurev-fluid-030121-113138.

Bernard, A., Long, N., Becker, M., Khan, J. & Fanchette, S. 2022 Bangladesh's vulnerability to cyclonic coastal flooding. *Natural Hazards and Earth System Sciences* **22**, 729–751. https://doi.org/10.5194/nhess-22-729-2022.

Boelee, L. 2022 *Evaluation of Global Flood Forecasts in Ungauged Catchments*. Ph.D. thesis, The Open University, UK. Available from: http://oro.open.ac.uk/84856/.

Bradbrook, K. 2006 JFLOW: a multiscale two-dimensional dynamic flood model. *Water and Environment Journal* **20**, 79–86. https://doi.org/ https://doi.org/10.1111/j.1747-6593.2005.00011.x.

BWDB 2020 *Bangladesh Water Development Board Annual Flood Report 2020*. Available from: http://ffwc.gov.bd/, last access 26th October 2022.

Chini, M., Hostache, R., Giustarini, L. & Matgen, P. 2017 A hierarchical split-based approach for parametric thresholding of SAR images: flood inundation as a test case. *IEEE Transactions on Geoscience and Remote Sensing* **55**, 6975–6988. https://doi.org/10.1109/TGRS.2017.2737664.

Cooper, E. S., Dance, S. L., García-Pintado, J., Nichols, N. K. & Smith, P. J. 2019 Observation operators for assimilation of satellite observations in fluvial inundation forecasting. *Hydrology and Earth System Sciences* **23**, 2541–2559. https://doi.org/10.5194/hess-23-2541-2019.

Copernicus Programme. 2021 *Copernicus Emergency Management Service*. Available from: https://emergency.copernicus.eu/ last access 14th September 2021.

Coughlan de Perez, E., Van Den Hurk, B., Van Aalst, M. K., Jongman, B., Klose, T. & Suarez, P. 2015 Forecast-based financing: an approach for catalyzing humanitarian action based on extreme weather and climate forecasts. *Natural Hazards and Earth System Sciences* **15**, 895–904. https://doi.org/10.5194/nhess-15-895-2015.

Coughlan de Perez, E., Van Den Hurk, B., Van Aalst, M. K., Amuron, I., Bamanya, D., Hauser, T., Jongma, B., Lopez, A., Mason, S., De Suarez, J. M., Pappenberger, F., Rueth, A., Stephens, E., Suarez, P., Wagemaker, J. & Zsoter, E. 2016 Action-based flood forecasting for triggering humanitarian action. *Hydrology and Earth System Sciences* **20**, 3549–3560. https://doi.org/10.5194/hess-20-3549-2016.

Coughlan de Perez, E., Berse, K. B., Depante, L. A. C., Easton-Calabria, E., Evidente, E. P. R., Ezike, T., Heinrich, D., Jack, C., Lagmay, A. M. F. A., Lendelvo, S., Marunye, J., Maxwell, D. G., Murshed, S. B., Orach, C. G., Pinto, M., Poole, L. B., Rathod, K., Shampa, C. & Sant, C. V. 2022 Learning from the past in moving to the future: invest in communication and response to weather early warnings to reduce death and damage. *Climate Risk Management* **38** (100), 461. https://doi.org/10.1016/j.crm.2022.100461.

Dey, S. R., Roberts, N. M., Plant, R. S. & Migliorini, S. 2016 A new method for the characterization and verification of local spatial predictability for convective-scale ensembles. *Quarterly Journal of the Royal Meteorological Society*. https://doi.org/10.1002/qj.2792.

Dottori, F. & Todini, E. 2011 Developments of a flood inundation model based on the cellular automata approach: testing different methods to improve model performance. *Physics and Chemistry of the Earth* **36**, 266–280. https://doi.org/10.1016/j.pce.2011.02.004.

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.

Dunning, P. 2019 FLY. Available from: https://www.jbarisk.com/news-blogs/fly-technology-revolutionising-the-world-of-catastrophe-modelling/, last access 20th January 2023.

Emerton, R. E., Stephens, E. M., Pappenberger, F., Pagano, T. C., Weerts, A. H., Wood, A. W., Salamon, P., Brown, J. D., Hjerdt, N., Donnelly, C., Baugh, C. A. & Cloke, H. L. 2016 Continental and global scale flood forecasting systems. *Wiley Interdisciplinary Reviews: Water* **3**, 391–418. https://doi.org/10.1002/wat2.1137.

EU Science Hub. 2021 *The Joint Research Centre Launches A Revolutionary Tool for Monitoring Ongoing Floods Worldwide as Part of the Copernicus Emergency Management Service*. Available from: https://ec.europa.eu/jrc/en/news/jrc-launches-revolutionary-tool-for-monitoring-floods-worldwide-part-copernicus-emergency-management-service, last access 28th October 2021.

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D. & Alsdorf, D. 2007 The shuttle radar topography mission. *Reviews of Geophysics* **45**. https://doi.org/https://doi.org/10.1029/2005RG000183.

Fekete, A. & Sandholz, S. 2021 Here comes the flood, but not failure? lessons to learn after the heavy rain and pluvial floods in Germany 2021. *Water (Switzerland)* **13**, 1–20. https://doi.org/10.3390/w13213016.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A. & Michaelsen, J. 2015 The climate hazards infrared precipitation with stations – A new environmental record for monitoring extremes. *Scientific Data* **2**, 1–21. https://doi.org/10.1038/sdata.2015.66.

García-Pintado, J., Mason, D. C., Dance, S. L., Cloke, H. L., Neal, J. C., Freer, J. & Bates, P. D. 2015 Satellite-supported flood forecasting in river networks: a real case study. *Journal of Hydrology* **523**, 706–724. https://doi.org/10.1016/J.JHYDROL.2015.01.084.

GFM 2021 *GloFAS Global Flood Monitoring (GFM)*. Available from: https://www.globalfloods.eu/technical-information/glofas-gfm/, last access 25th November 2022.

GloFAS 2022a *GloFAS Methods*. Available from: https://www.globalfloods.eu/general-information/glofas-methods/, last access 4th November 2022.

GloFAS 2022b *GloFAS Rapid Flood Mapping*. Available from: https://confluence.ecmwf.int/display/CEMS/GloFAS+Rapid+Flood+Mapping+and+Rapid+Impact+Assessment, last access 4th November 2022.

GloFAS 2022c *GloFAS Map Viewer*. Available from: https://www.globalfloods.eu/, last access 8th December 2022.

Gourbesville, P. 1998 MIKE 11 GIS: interest of GIS technology for conception of flood protection systems. In: *Proceedings of the 3rd Hydroinformatics Conference – Hydroinformatics'98*, Copenhagen, Denmark.

Grimaldi, S. 2022 *GloFAS v4.0 Hydrological Reanalysis*. Available from: https://www.globalfloods.eu/news/113-glofas-v34-release-and-glofas-v40-hydrological-reanalysis-dataset-announcement/, last access 25th November 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**, 2043–2060. https://doi.org/10.5194/essd-12-2043-2020.

Havnø, K., Madsen, M. N. & Dørge, J., 1995 MIKE 11 – a generalized river modelling package. In: *Computer Models of Watershed Hydrology* (Singh, V. P., ed.). Water Resources Publications, Colorado, USA, pp. 733–782.

Hirpa, F. A., Salamon, P., Beck, H. E., Lorini, V., Alfieri, L., Zsoter, E. & Dadson, S. J. 2018 Calibration of the global flood awareness system (GloFAS) using daily streamflow data. *Journal of Hydrology* **566**, 595–606. https://doi.org/10.1016/j.jhydrol.2018.09.052.

Hoch, J. M. & Trigg, M. A. 2019 Advancing global flood hazard simulations by improving comparability, benchmarking, and integration of global flood models. *Environmental Research Letters* **14**. https://doi.org/10.1088/1748-9326/aaf3d3.

Hooker, H., Dance, S. L., Mason, D. C., Bevington, J. & Shelton, K. 2022 Spatial scale evaluation of forecast flood inundation maps. *Journal of Hydrology* 128170. https://doi.org/10.1016/j.jhydrol.2022.128170.

Hooker, H., Dance, S. L., Mason, D. C., Bevington, J. & Shelton, K. 2023a Assessing the spatial spread-skill of ensemble flood maps with remote-sensing observations. *Natural Hazards and Earth System Sciences* **23**, 2769–2785. https://doi.org/10.5194/nhess-23-2769-2023.

Hooker, H., Dance, S. L., Mason, D. C., Bevington, J. & Shelton, K. 2023b A multi-system comparison of forecast flooding extent using a scale-selective approach. https://doi.org/10.5281/zenodo.7509980.

Horritt, M. S., Mason, D. C. & Luckman, A. J. 2001 Flood boundary delineation from synthetic aperture radar imagery using a statistical active contour model. *International Journal of Remote Sensing* **22**, 2489–2507. https://doi.org/10.1080/01431160116902.

Hossain, S. 2020 *GloFAS Case Study: Bangladesh 2020*. Available from: https://www.globalfloods.eu/get-involved/case-study-bangladesh-2020/, last access 18th October 2022.

Hossain, S., Cloke, H. L., Ficchì, A., Turner, A. G. & Stephens, E. M. 2021 Hydrometeorological drivers of flood characteristics in the Brahmaputra river basin in Bangladesh. *Hydrology and Earth System Sciences Discussions* **2021**, 1–28. https://doi.org/10.5194/hess-2021-97.

Hostache, R. 2021 *A First Evaluation of the Future CEMS Systematic Global Flood Monitoring Product*. Available from: https://events.ecmwf.int/event/222/contributions/2274/attachments/1280/2347/Hydrological-WS-Hostache.pdf, last access 4th August 2021.

Lehner, B. 2014 *HydroBASINS Global Watershed Boundaries and sub-Basin Delineations Derived From HydroSHEDS Data at 15 Second Resolution*. Available from: https://www.hydrosheds.org/images/inpages/HydroBASINS_TechDoc_v1c.pdf, last access 5th November 2022.

LISFLOOD. 2022 *LISFLOOD*. Available from: https://ec-jrc.github.io/lisflood-model/1_1_introduction_LISFLOOD/, last access 4th November 2022.

Mason, D. C., Schumann, G. J., Neal, J. C., Garcia-Pintado, J. & Bates, P. D. 2012 Automatic near real-time selection of flood water levels from high resolution Synthetic Aperture Radar images for assimilation into hydraulic models: a case study. *Remote Sensing of Environment*. https://doi.org/10.1016/j.rse.2012.06.017.

Mason, D. C., Dance, S. L., Vetra-Carvalho, S. & Cloke, H. L. 2018 Robust algorithm for detecting floodwater in urban areas using synthetic aperture radar images. *Journal of Applied Remote Sensing* **12**, 1. https://doi.org/10.1117/1.jrs.12.045011.

Mason, D. C., Dance, S. L. & Cloke, H. L. 2021a Floodwater detection in urban areas using Sentinel-1 and WorldDEM data. *Journal of Applied Remote Sensing* **15**, 1–22. https://doi.org/10.1117/1.jrs.15.032003.

Mason, D. C., Bevington, J., Dance, S. L., Revilla-Romero, B., Smith, R., Vetra-Carvalho, S. & Cloke, H. L. 2021b Improving urban flood mapping by merging synthetic aperture radar-derived flood footprints with flood hazard maps. *Water (Switzerland)* **13**. https://doi.org/10.3390/w13111577.

NCAR 2022 *Weather Research & Forecasting Model*. Available from: https://www.mmm.ucar.edu/models/wrf, last access 14th November 2022.

OCHA 2020 *Bangladesh Monsoon Flooding 2020: Anticipatory Action Pilot*. Available from: https://www.unocha.org/our-work/humanitarian-financing/anticipatory-action/summary-bangladesh-pilot, last access 18th October 2022.

Pappenberger, F., Cloke, H. L., Parker, D. J., Wetterhall, F., Richardson, D. S. & Thielen, J. 2015 The monetary benefit of early flood warnings in Europe. *Environmental Science and Policy* **51**, 278–291. https://doi.org/10.1016/j.envsci.2015.04.016.

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

Revilla-Romero, B., Shelton, K., Wood, E., Berry, R., Bevington, J., Hankin, B., Lewis, G., Gubbin, A., Griffiths, S., Barnard, P., Pinnell, M. & Huyck, C. 2017 Flood Foresight: A near-real time flood monitoring and forecasting tool for rapid and predictive flood impact assessment. In: *EGU General Assembly Conference Abstracts, EGU General Assembly Conference Abstracts*, p. 1230.

Roberts, N. M. & Lean, H. W. 2008 Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Monthly Weather Review*. https://doi.org/10.1175/2007MWR2123.1.

Savage, J. T. S., Bates, P., Freer, J., Neal, J. & Aronica, G. 2016 When does spatial resolution become spurious in probabilistic flood inundation predictions? *Hydrological Processes* **30**, 2014–2032. https://doi.org/10.1002/hyp.10749.

Schumann, G., Giustarini, L., Tarpanelli, A. & Jarihani, B. 2022 Flood modeling and prediction using earth observation data. *Surveys in Geophysics*. https://doi.org/10.1007/s10712-022-09751-y.

Speight, L. J., Cranston, M. D., White, C. J. & Kelly, L. 2021 Operational and emerging capabilities for surface water flood forecasting. *Wiley Interdisciplinary Reviews: Water* **8**, 1–24. https://doi.org/10.1002/wat2.1517.

Start Network. 2022 *Anticipation and Risk Financing*. Available from: https://startnetwork.org/anticipation-and-risk-financing, last access 25th October 2022.

Stephens, E. & Cloke, H. 2014 Improving flood forecasts for better flood preparedness in the UK (and beyond). *Geographical Journal* **180**, 310–316. https://doi.org/10.1111/geoj.12103.

Stephens, E., Schumann, G. & Bates, P. 2014 Problems with binary pattern measures for flood model evaluation. *Hydrological Processes*. https://doi.org/10.1002/hyp.9979.

UNDRR 2022 *Global Assessment Report on Disaster Risk Reduction*. Available from: https://www.undrr.org/gar2022-our-world-risk#container-downloads. last access 25th October 2022.

WASDI 2022 WASDI. Available from: https://www.wasdi.net/#!/marketplace, last access 4th November 2022.

Wu, W., Emerton, R., Duan, Q., Wood, A. W., Wetterhall, F. & Robertson, D. E. 2020 Ensemble flood forecasting: current status and future opportunities. *WIRES Water* **7**, 1–32. https://doi.org/10.1002/wat2.1432.

Zsoter, E., Prudhomme, C., Stephens, E., Pappenberger, F. & Cloke, H. 2020 Using ensemble reforecasts to generate flood thresholds for improved global flood forecasting. *Journal of Flood Risk Management* **13**, 1–14. https://doi.org/10.1111/jfr3.12658.

First received 26 January 2023; accepted in revised form 13 September 2023. Available online 26 September 2023