https://doi.org/10.5194/nhess-2022-188
Preprint. Discussion started: 14 July 2022

Natural Hazards and Earth System Sciences
Discussions

A new skill score for ensemble flood maps: assessing spatial spread-skill with remote sensing observations.

Helen Hooker[1], Sarah L. Dance[1,2,3], David C. Mason[4], John Bevington[5], and Kay Shelton[5]

[1]Department of Meteorology, University of Reading, UK
[2]Department of Mathematics and Statistics, University of Reading, UK
[3]National Centre for Earth Observation (NCEO), Reading, UK
[4]Department of Geography and Environmental Science, University of Reading, UK
[5]Jeremy Benn Associates Limited (JBA Consulting), UK

Correspondence: Helen Hooker (h.hooker@pgr.reading.ac.uk)

Abstract. An ensemble of forecast flood inundation maps has the potential to represent the uncertainty in the flood forecast and provide a location specific, probabilistic, likelihood of flooding. This gives valuable information to flood forecasters, flood risk managers and insurers and will ultimately benefit people living in flood prone areas. Spatial verification of the ensemble flood map forecast against remotely observed flooding is important to understand both the skill of the ensemble forecast and the uncertainty represented in the variation or spread of the individual ensemble member flood maps. Previously, a scale-selective approach has been used to evaluate a convective precipitation ensemble forecast. This determines a skilful scale of ensemble performance. By extending this approach through a new application we evaluate the spatial predictability and the spatial spread-skill of an ensemble flood forecast across a domain of interest. The spatial spread-skill method computes an agreement scale at grid level between each unique pair of ensemble flood maps (ensemble spatial spread) and between each ensemble flood map with a SAR-derived flood map (ensemble spatial skill). By comparing these we can determine the spatial spread-skill performance. These methods are applied to an example flood event on the Brahmaputra River in the Assam region of India, August 2017. Both the spatial-skill and spread-skill relationship vary with location and can be related to physical characteristics of the flooding event. Routine validation and mapping of spatial predictability in an operational system would allow better quantification of model systematic biases and uncertainties. This would be particularly useful for ungauged catchments and would enable targeted model improvements to be made across different parts of the forecast chain.

1 Introduction

Forecast flood maps indicating the extent and depth of a predicted fluvial flood within an actionable lead time, are a vital tool for flood risk managers and emergency response teams prior to and during a flood event. Typically, forecast flood maps are presented as deterministic forecasts predicting precisely where flooding will occur. This can lead to incidents of false alarms or missed warnings and subsequent recriminations causing mistrust in the system (Cloke and Pappenberger, 2009; Savage et al., 2016). A timely prediction of exactly where and when flooding will occur is virtually impossible due to the chaotic nature of the atmosphere (Lorenz, 1969) which ultimately determines where heavy precipitation will fall. The ensemble forecasting

https://doi.org/10.5194/nhess-2022-188
Preprint. Discussion started: 14 July 2022


Natural Hazards and Earth System Sciences Discussions

approach aims to address the sensitive nature of the atmosphere to its initial conditions and through multiple model runs these uncertainties can be quantified (Leutbecher and Palmer, 2008). This results in a probabilistic weather forecast that indicates the predictability of the atmosphere at a given space and time. State-of-the-art operational ensemble flood forecasting systems link together a chain of forecast models to produce probabilistic streamflow and flood inundation forecasts at national and global scales (Cloke and Pappenberger, 2009; Emerton et al., 2016; Wu et al., 2020). Ensemble Numerical Weather Prediction (NWP) models provide meteorological inputs into land-surface, hydrological and hydraulic models, cascading the atmospheric uncertainty through to the flood forecast. Throughout this chain of models multiple sources of uncertainties exist (Beven, 2016; Matthews et al., 2022; Pappenberger et al., 2005; Zappa et al., 2011). Boelee et al. (2019) detail the uncertainties arising from meteorological inputs, measurements and observations, initial conditions, unresolved physics within the models and parameter estimates. A probabilistic flood inundation forecast should present a meaningful prediction of the likelihood of flooding, given the uncertainties represented in the system (Alfonso et al., 2016).

The accuracy of the location of flooding, predicted in advance, is defined as spatial predictability. The spatial predictability of ensemble forecasts of flood inundation could be verified by comparing with a remote observation of the flood from satellite or unmanned aerial vehicle (UAV) based sensors. Satellite-based optical and Synthetic Aperture Radar (SAR) sensors are well known for their flood detection capability (e.g. Horritt et al., 2001; Mason et al., 2012). SAR sensors are active, which enables them to scan the Earth through weather and clouds, and at night. The SAR backscatter intensity detected depends on the roughness of the surface with unobstructed flooded areas and other surface water bodies appearing relatively smooth and returning low backscatter values. Dasgupta et al. (2018a) detail some of the challenges along with approaches to solutions of flood detection using SAR. Examples of these challenges include; roughening of the water surface by heavy rain and strong wind, emergent or partially submerged vegetation and flood detection in urban areas. Accurate flood detection in urban areas particularly due to surface water flooding has become increasingly important and recent techniques have led to improved flood detection (Mason et al., 2018, 2021a, b). Optical instruments rely on solar energy and cannot penetrate cloud, making them less useful during a flooding situation. Recent studies have investigated the flood detection benefits from combining both optical and SAR imagery (Konapala et al., 2021; Tavus et al., 2020). Improvements in the spatial-temporal resolution of SAR images and their open source availability mean that they are an increasingly valuable tool for hydraulic and hydrodynamic model improvements through calibration, validation and data assimilation (e.g. García-Pintado et al., 2015; Hostache et al., 2018; Cooper et al., 2018, 2019; Di Mauro et al., 2021; Dasgupta et al., 2018b, 2021a, b). The Global Flood Monitoring (GFM) product (EU Science Hub, 2021; GFM, 2021; Hostache, R., 2021) of the Copernicus Emergency Management Service (CEMS) (Copernicus Programme, 2021) produces SAR-derived flood inundation maps for every Sentinel-1 image detecting flooding. Three flood detection algorithms provide uncertainty estimation and population affected estimates within 8 hours of the image acquisition. The European Space Agency (ESA) Copernicus Programme have recently included the ICEYE constellation of small satellites into the fleet of missions contributing to Europe's Copernicus environmental monitoring programme (ESA, 2021). ICEYE captures very high resolution (spot mode ground range resolution = 1 m) SAR images which brings the potential for increased

https://doi.org/10.5194/nhess-2022-188
Preprint. Discussion started: 14 July 2022


Natural Hazards and Earth System Sciences Discussions

accuracy of flood detection, particularly in urban areas.

To evaluate the accuracy of an ensemble forecast, a number of verification measures have been applied previously. Anderson et al. (2019) developed a joint verification framework for end-to-end assessment of the England and Wales Flood Forecasting Centre (FFC) ensemble flood forecasting system. Anderson et al. describe the verification metrics such as the continuous rank probability score (CRPS), rank histograms, Brier Skill Score (BSS) and the relative operative characteristics (ROC) diagrams that are commonly applied to assess the main ensemble attributes desirable in both precipitation and streamflow ensemble forecasts (e.g. Renner et al., 2009). These metrics refer to flooding events as part of a time series evaluated against a reference benchmark such as climatology to produce an average skill score. These differ from the ensemble spatial verification approaches applied at a specific time point considered here. The verification of ensemble forecasts usually involves comparing the RMSE of the ensemble mean against an observed quantity to assess the *skill* of the forecast with the ensemble standard deviation used as a measure of *spread*. A perfect ensemble should encompass forecast uncertainties such that the ensemble spread equals the RMSE. This *spread-skill* relationship was assessed by Buizza (1997) to investigate the predictability limits of the European Centre for Medium-Range Weather Forecasts (ECMWF) Ensemble Prediction System (EPS). This approach to ensemble verification is based on point values and makes the assumption that the ensemble mean is the forecast state with the highest probability and that the forecast distribution is Gaussian. Significant flooding events are, in their nature, a rare occurrence and in certain circumstances a few ensemble members can indicate a low probability of an extreme flood. Also, in particular atmospheric scenarios the ensemble forecast may result in a multi-modal forecast where two clusters of ensemble members are each equally likely. For example, both clusters may indicate flooding events but at different magnitudes. In both of these instances the individual ensemble member details are important and evaluation of the ensemble mean alone would not be meaningful: the ensemble mean field alone does not retain the spatial detail of the individual member forecasts.

The spatial predictability of the full ensemble must be evaluated against observations of flooding to determine the spatial spread-skill of the ensemble forecast. For a flood map ensemble to be considered spatially well-spread, the spread or variation between ensemble members should equal the spatial predictability, or skill of the ensemble members. Presently, to the best of our knowledge, quantitative evaluation methods assessing the spatial spread-skill of ensemble forecast flood maps do not exist. In numerical weather prediction previous work by Ben Bouallègue and Theis (2014) investigated the application of spatial techniques to ensemble precipitation forecasts using a neighbourhood, or fuzzy approach that allowed comparisons at larger scales than grid level. A location dependent approach to the spatial spread-skill evaluation of a convective precipitation ensemble forecast was developed by Dey et al. (2016). This method compares every ensemble member across a range of scales on a spatial field against an observation field to assess whether the ensemble forecast is spatially over-, under- or well-spread on average across a domain of interest (Chen et al., 2018). In a recent study, a scale-selective approach was developed and applied to evaluate a deterministic flood map forecast (Hooker et al., 2022a). This paper extends and applies this approach to assess the spatial predictability and the spatial spread-skill of an ensemble flood map forecast.

https://doi.org/10.5194/nhess-2022-188
Preprint. Discussion started: 14 July 2022


Natural Hazards and Earth System Sciences Discussions

In this paper we aim to address the following questions:

- How can we summarise the spatial predictability information in ensemble flood map forecasts?
- How can we evaluate the spatial spread-skill of an ensemble flood map forecast?
- How does the spatial spread-skill vary with location and how can this be presented?

In Section 2 we present a new approach to the evaluation of spatial predictability and the spatial spread-skill of an ensemble flood map forecast by comparing against a remotely observed flooding extent. We illustrate the features of the methods through en example case study of an extreme flooding event of the Brahmaputra river which impacted India and Bangladesh in August 2017; with focus on the Assam region of India. The flood event details are described in Section 3.1. The international ensemble version of the JBA Consulting Flood Foresight system provides forecast flood maps for the study and is described in Section 3.2. Observations of the flood are derived from satellite based Synthetic Aperture Radar (SAR) sensors and the method is explained in Section 3.3. The results including our new Spatial spread-skill (SSS) map are discussed in Section 4. Our results show that individual ensemble member spatial predictions of flooding are meaningful and that the full ensemble spatial detail should be evaluated. We conclude in Section 5 that the spatial spread-skill of the ensemble forecast varies with location across the domain and can be linked to physical characteristics of the flooding event.

2 Ensemble flood map spatial predictability evaluation methods

Hooker et al. (2022a) described and applied a new scale-selective approach to evaluate the spatial skill of a deterministic flood map forecast in terms of a skilful scale at which the forecast captures the observed SAR-derived flood map. Here, we apply this verification measure to evaluate different aspects of an ensemble forecast. The Fraction Skill Score (FSS) method is outlined in Section 2.1. Agreement scale maps are defined for location specific comparisons between two fields in Section 2.2. These are used to assess the spatial relationship between each unique pair of ensemble member flood maps and between every ensemble member flood map and the observed SAR-derived flood map (Section 2.3). Visualisation methods of the spatial spread-skill relationship including our new *Spatial Spread-Skill (SSS)* map are presented in Section 2.4.

2.1 Fraction Skill Score

The FSS is a scale-selective verification measure that can determine the skilful scale of a modelled flood map, when compared against remotely sensed observations of a flood. We will call these the *model field array* and the *observed field array* respectively. For an ensemble forecast, the modelled flood map could be an individual ensemble member, or a summarised flood estimate derived from a combination of ensemble members (see Section 3.4).

Hooker et al. (2022a) describe the calculation of the FSS. "Both arrays are converted into binary fields using a threshold approach that is predetermined for the situation. For a flood map verification of spatial skill, the simplest example is applied

https://doi.org/10.5194/nhess-2022-188
Preprint. Discussion started: 14 July 2022


Natural Hazards and Earth System Sciences
Discussions

here and assigns each grid cell as unflooded (0) or flooded (1) for the entire domain. Alternative threshold approaches for flood depth maps could include applying thresholds to water depth percentiles.

Given a domain of interest, each of the grid cells are numbered according to their spatial coordinates (i, j), $i = 1 \dots N_x$ and $j = 1 \dots N_y$ where N_x is the number of columns in the domain and N_y is the number of rows. For each grid cell a square of length n forms an $n \times n$ neighbourhood surrounding the grid cell. The fraction of $1s$ in the square neighbourhood is calculated for each grid cell. This creates two fields of fractions over the domain for both the forecast M_{nij} and observed O_{nij} data. The fraction fields are compared against one another to calculate the mean squared error (MSE) for the neighbourhood

$$MSE_n = \frac{1}{N_x N_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} [O_{nij} - M_{nij}]^2. \tag{1}$$

Based on the fractions calculated for the model and observed fields a worst possible MSE is calculated

$$MSE_{n(ref)} = \frac{1}{N_x N_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} [O_{nij}^2 + M_{nij}^2]. \tag{2}$$

The FSS is given by

$$FSS_n = 1 - \frac{MSE_n}{MSE_{n(ref)}}. \tag{3}$$

In general, the FSS is calculated for each length of neighbourhood n. For a given neighbourhood size an FSS of 1 is said to have perfect skill and 0 means no skill. The FSS will increase as n increases up to an asymptote (see Fig. 3 from Roberts and Lean (2008)). If there is no model bias across the whole domain of interest (observed and forecast flooded areas are the same) then the asymptotic fraction skill score (AFSS) at $n = 2N - 1$, where N is the number of points along the longest side of the domain, will equal 1. Plotting FSS against spatial scale can indicate a range of scales where the model is deemed to be the most useful. This usefulness is a trade-off between being too smooth (larger n) or too fine, where the forecast skill is lost and the computation time lengthy. A target FSS score (FSS_T) is defined as

$$FSS_T \geq 0.5 + \frac{f_o}{2}, \tag{4}$$

where f_0 is the fraction of flood observed across the whole domain of interest and can be thought of as being equidistant between the skill of a random forecast and perfect skill. FSS_T will vary depending on the magnitude of the observed flood, relative to the domain area. This allows the comparison of the FSS_T scale across different domain sizes and floods of different magnitudes."

2.2 Location dependent agreement scales

The FSS (Section 2.1) gives a domain average measure of forecast performance and a minimum scale at which the forecast is deemed skilful. To enable the spatial spread-skill of the full ensemble to be evaluated, we first define a measure of similarity

https://doi.org/10.5194/nhess-2022-188
Preprint. Discussion started: 14 July 2022

in terms of location specific agreement scales. Here, two data fields are compared f_{1ij} and f_{2ij}, these could both be ensemble member flood maps or an ensemble member flood map and an observed flood map. These are non-thresholded fields and the agreement scale method can be applied to both binary flood extent maps as well as flood depth fields. Following Hooker et al. (2022a) "the aim is to find a minimum neighbourhood size (or scale) for every grid point such that there is an agreement between f_{1ij} and f_{2ij}. This is known as the agreement scale $S_{ij}^{A(f_1 f_2)}$. The relationship between the agreement scale and the neighbourhood size described previously in section 2.1 is given by $S_{ij}^{A(f_1 f_2)} = (n-1)/2$.

Firstly, all grid points are compared by calculating the relative MSE D_{ij}^{S} at the grid scale, $S = 0$ ($n = 1$),

$$D_{ij}^{S} = \frac{(f_{1ij}^{S} - f_{2ij}^{S})^2}{(f_{1ij}^{S})^2 + (f_{2ij}^{S})^2}. \tag{5}$$

If $f_{1ij} = 0$ and $f_{2ij} = 0$ (both dry) then D_{ij}^{S} = 0 (correct at grid level). Note that D_{ij}^{S} varies from zero to 1. The fields are considered to be in agreement at the scale being tested if:

$$D_{ij}^{S} \leq D_{crit,ij}^{S_{ij}} \quad \text{where} \quad D_{crit,ij}^{S} = \alpha + (1-\alpha)\frac{S}{S_{lim}} \tag{6}$$

and S_{lim} is a predetermined, fixed maximum scale. The parameter value α is chosen to indicate the acceptable bias at grid level such that $0 \leq \alpha \leq 1$. Here we set $\alpha = 0$ (no background bias). If $D_{ij}^{S} \geq D_{crit,ij}^{S}$ then the next neighbourhood size up is considered ($S = 1$, a 3 by 3 square). The process continues with increasingly larger neighbourhoods until the agreement scale

$$S_{ij}^{A(f_1 f_2)} \quad \text{or} \quad S_{lim} \quad \text{at} \quad D_{ij}^{S} \leq D_{crit,ij}^{S_{ij}} \tag{7}$$

is reached for every cell in the domain of interest. The agreement scale at each grid cell is then mapped onto the domain of interest."

A categorical scale map for an individual ensemble member or a combination of members such as the ensemble median can be created by combining the agreement scale map with a conventional contingency map (see Hooker et al. (2022a) for further details). Categorical scale maps may be used as a basis for comparison between ensemble members and observations, as we illustrate with our case study in Section 4.3.

2.3 Ensemble spatial spread-skill evaluation

We assume that each ensemble forecast flood map represents an equally likely future scenario and the evaluation of the full ensemble is needed to quantify the uncertainty and to evaluate the spatial spread-skill relationship. The ensemble flood map spatial characteristics vary with location and in order to preserve the location dependent information, we utilise a method developed to evaluate a convective ensemble precipitation forecast (Dey et al., 2016). Here, we outline the method and describe a new application to evaluate an ensemble forecast flood map (see Dey et al. (2016) for additional details of the method).

https://doi.org/10.5194/nhess-2022-188
Preprint. Discussion started: 14 July 2022


Natural Hazards and Earth System Sciences
Discussions

A neighbourhood approach (Section 2.2) is used to assess the spatial agreement scale $S_{ij}^{A(f_1 f_2)}$ or measure of similarity at each grid cell location (i, j) between each unique pair of ensemble flood maps. For an ensemble of N members, there are

$$N_p = \frac{N(N-1)}{2}, \tag{8}$$

unique pairs (1275 pairs for a 51 member ensemble). Between each flood map and the SAR-derived flood map there are 51 pairs. For an ensemble, the skillful scale can be renamed as a *believable scale*, which is the scale where ensemble members become sufficiently similar to observations such that they are a useful prediction. Every paired ensemble agreement scale field is averaged at each grid cell to produce a mean field, following from equation (7)

$$S_{ij}^{A(\overline{mm})} = \frac{1}{N_p} \sum_{f_1=1}^{N-1} \sum_{f_2=f_1+1}^{N} S_{ij}^{A(f_1 f_2)} \tag{9}$$

indicating the location specific believable scales of the forecast flood map ensemble. Maps of $S_{ij}^{A(\overline{mm})}$ summarise the spatial predictability and can be linked to physical processes. Each of the agreement scale fields between the ensemble members and the observations are also averaged at each grid cell to give

$$S_{ij}^{A(\overline{mo})} = \frac{1}{N} \sum_{f=1}^{N} S_{ij}^{A(f_0)}. \tag{10}$$

A measure of the spatial spread-skill of the ensemble can be found by comparing the average agreement scale between the ensemble members $S_{ij}^{A(\overline{mm})}$ representing the ensemble *spread* with the average agreement scale between the ensemble members and the observed flood field $S_{ij}^{A(\overline{mo})}$ representing the ensemble *skill*. Visualisation methods for evaluating the spatial spread-skill are presented in Section 2.4.

2.4 Spatial spread-skill visualisation methods

To evaluate the spatial spread-skill, $S_{ij}^{A(\overline{mm})}$ representing the ensemble *spread* must be compared in the same location as $S_{ij}^{A(\overline{mo})}$ representing the ensemble *skill*. A binned scatter plot compares groups of grid cells from two domains by averaging across a selected bin of cells at the same location on each domain. Dey et al. (2016) demonstrated for an idealised example that by plotting $S_{ij}^{A(\overline{mm})}$ against $S_{ij}^{A(\overline{mo})}$ as a binned scatter plot in order to preserve the spatial location of the comparison (Fig. 1), the ensemble can be classified as over-, under- or well-spread. The ensemble is deemed to be *well-spread* at a specific location in the domain of interest when the spread of the individual members represented at each grid cell by $S_{ij}^{A(\overline{mm})}$ equals the skill of the ensemble represented at each grid cell by $S_{ij}^{A(\overline{mo})}$, i.e. $S_{ij}^{A(\overline{mm})} - S_{ij}^{A(\overline{mo})} = 0$. The result would lie on a 1:1 line on the binned scatter plot. Where the spread between the ensemble members exceeds the skill of the ensemble forecast i.e. $S_{ij}^{A(\overline{mm})} > S_{ij}^{A(\overline{mo})}$ the ensemble is considered to be *over-spread* and the binned scatter plot will lie beneath the 1:1 line. The converse is true for an *under-spread* ensemble forecast where the agreement between members, the spread, is less than the agreement between the ensemble and the observations, the skill. Here, $S_{ij}^{A(\overline{mm})} < S_{ij}^{A(\overline{mo})}$ and the binned scatter plot would lie above the 1:1 line.

https://doi.org/10.5194/nhess-2022-188
Preprint. Discussion started: 14 July 2022

Natural Hazards and Earth System Sciences
Discussions

To summarise the spread-skill relationship for our example flood case we develop this visualisation further by plotting a hexagonal binned 2D histogram plot (hexbin, Section 4.3, Fig. 8). The domain is divided into a (pre-determined) number of hexagons. Hexagons reduce the sampling bias compared with a square grid due to the edge effects of the grid shape. This is related to the low perimeter-to-area ratio of the shape of the hexagon, a circle has the lowest ratio but cannot tessellate to form a continuous grid. Hexagons are closest to a circular shape and can tessellate to form an evenly spaced grid. The hexbin histogram plot colours represents the number of data points within each bin. This adds additional information to the standard binned scatter plot.

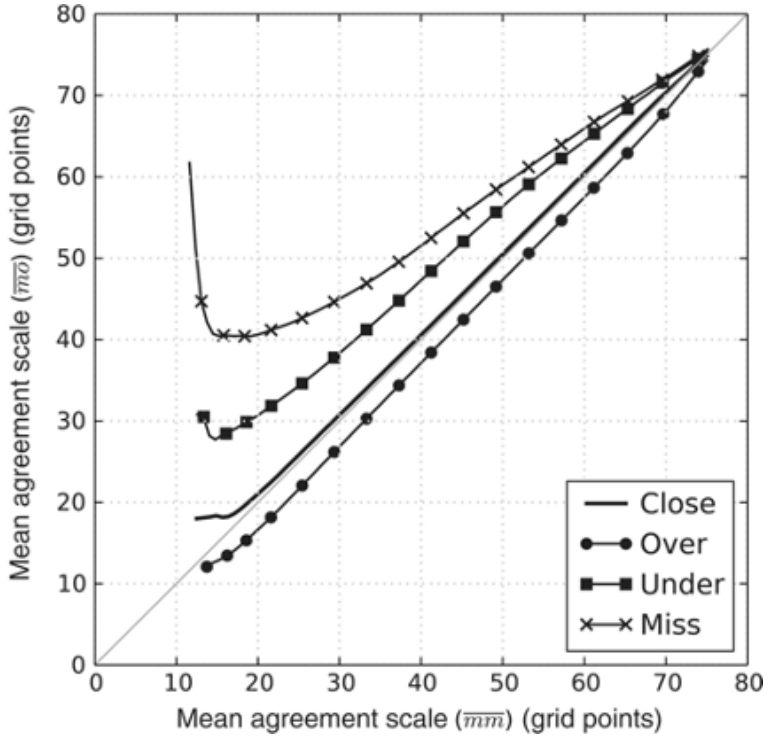

Figure 1. Figure reproduced with permission from Dey et al. (2016) showing results from an idealised experiment indicating the spatial spread-skill relationship between an ensemble forecast and the observation.

2.4.1 Spatial Spread-Skill (SSS) map

Whilst the hexbin plot is useful for gaining an understanding of the general spread-skill relationship of the ensemble flood map forecast, it does not tell us specifically where in the domain the ensemble spatial predictability is better or worse so that it can be linked to physical processes and to improve model performance. Our new *Spatial Spread-Skill* (SSS) map plots $S_{ij}^{A(\overline{mm})} - S_{ij}^{A(\overline{mo})}$ at every grid cell location so that the spread-skill is mapped across the domain and can be linked directly

https://doi.org/10.5194/nhess-2022-188
Preprint. Discussion started: 14 July 2022


Natural Hazards and Earth System Sciences Discussions

to different sub-catchments and surface features such as tributaries, embankments, bridges and importantly the underlying topography or DTM, which influence the derivation of the ensemble flood maps. Regions on the SSS map where the ensemble is over-spread are positive with negative areas indicating where the ensemble is under-spread, zero values show a well-spread ensemble. Note that this does not necessarily mean that the entire ensemble is in agreement with observations at grid level, but that the agreement scales between $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ are equal.

3 Ensemble forecasting flood event case study

In this section we describe an example flooding event used to demonstrate the application of our new spatial spread-skill evaluation approach. We evaluate a 1-day flood inundation 51 ensemble member forecast from the Flood Foresight system (Section 3.2) for the domain area against a satellite SAR-derived flood map (Section 3.3).

3.1 Brahmaputra flood, Assam India, August 2017

The origin of the Brahmaputra River lies in the Himalayan Kailas Range of southwestern Tibet, China. Draining an area of 543,000 km^2, the Brahmaputra flows for 2000 km across the Tibetan Plateau and a further 1000 km parallel to the Himalayan foothills through the Assam Valley, India before entering Bangladesh where the Brahmaputra joins the Ganges River (Palash et al., 2020). The Brahmaputra baseflow originates from the upstream glacial snow melt, however the streamflow rates are dominated by the summer monsoon precipitation. The basin receives up to 95% of its annual rainfall during the pre-monsoon and monsoon season, which usually runs from April to September and causes annual flooding of the Brahmaputra. The Assam region typically records on average 2300 mm of annual rainfall and up to 5000 mm in the Himalayan foothills (Dhar and Nandargi, 2000, 2003).

For this example case we focus on the third wave of flooding that occurred during the monsoon season in August 2017, peaking around the 12$^{\text{th}}$. Figure 2 shows the location of the Brahmaputra and of a chosen domain centred upon some of the worst flooding that occurred. This area includes a confluence zone where the Subansiri River meets the Brahmaputra. The monsoon flooding impacted an estimated 40 million people across India and Bangladesh. Locally in the Assam region, the flooding in August affected 3.3 million people and 3186 villages. Over 14,000 people were evacuated to one of 678 relief camps that were also needed to home the 183,584 people relocated. The local Assam State Disaster Management Authority (ASDMA, 2017) flood early warning system issued a low warning alert (disasters that can be managed at the district level) on the 10$^{\text{th}}$ August for the district.

In 2017, the southwest monsoon season rainfalls were predicted to be *normal*. However, the pre-monsoon season began early in the year with heavy thunderstorms affecting the region from March onwards. In the Assam region, June and July were 60% wetter than the previous three years and during August more locally intense rainfall was recorded compared with historical observations (Palash et al., 2020). In higher latitude areas, 30 km to the north of the domain at North Lakhimpur, 215.8 mm

https://doi.org/10.5194/nhess-2022-188
Preprint. Discussion started: 14 July 2022

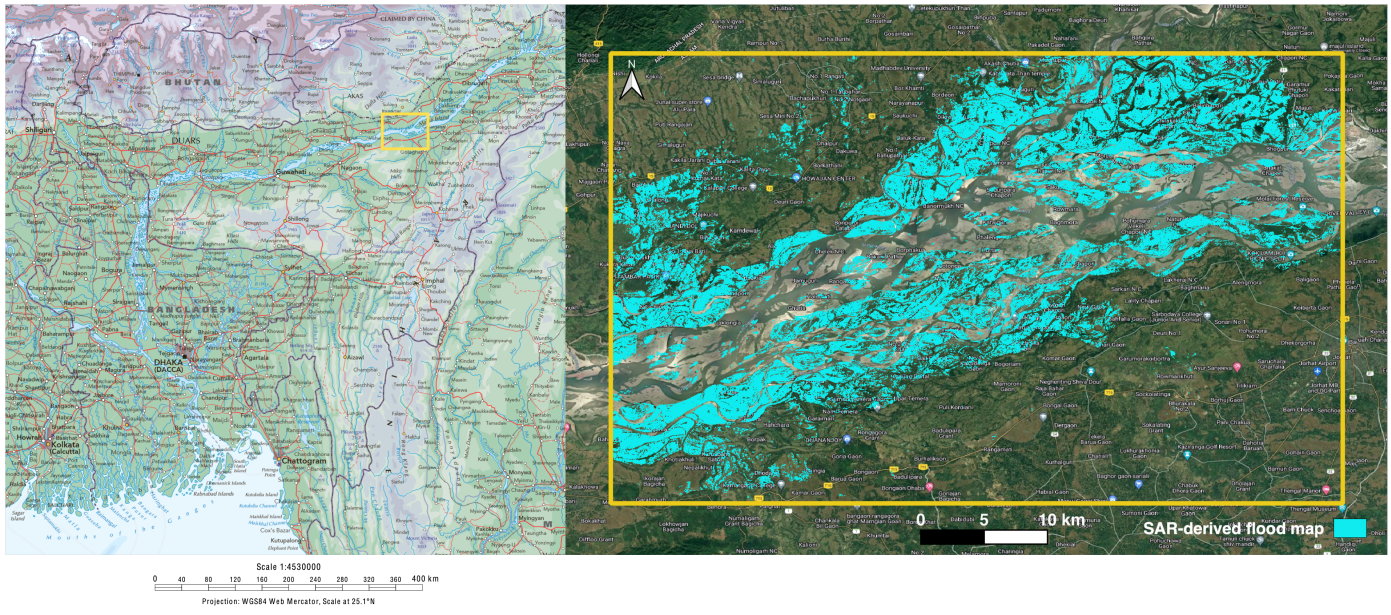

Figure 2. Left panel: domain location on the Brahmaputra River in the Assam region of India. Domain size is 57.5 km by 39.3 km. Right panel: Sentinel-1 SAR-derived flood map for the domain of interest. Base map from © Google Maps.

was recorded in the three days prior to the flood peak (Floodlist, 2017; Hossain et al., 2021). A flooding situation is declared in India when water levels (WL) rise above a pre-determined danger level (DL). If the WL increases beyond +1 m of the DL then a severe flood situation is triggered. The peak WL recorded at downstream Tezpur (DL 65.23 m) on August 14th was 66.12 m. There are reports of regional variations in maximum WL recorded with upland regions to the north of the Assam valley exceeding record levels. Flooding was exacerbated due to embankment breaches and deforestation has contributed to worsening flooding.

3.2 Ensemble flood forecasting system

The Flood Foresight system (Fig. 3), developed and operationally run by JBA Consulting, is a fluvial flood inundation mapping system that can be implemented at any river basin around the world. Flood Foresight utilises a simulation library approach to generate real-time and forecast flood inundation and water depth maps. The simulation library approach saves valuable computing time and allows the application of Flood Foresight in near continuous real-time at national and international scales. A pre-computed library of flood maps for a river basin or country are created using JFlow®, (Bradbrook, 2006) and RFlow. JFlow uses a raster-based approach with a detailed underlying digital terrain model (DTM) and a simplified form of the full 2D hydrodynamic shallow water flow equations. RFlow combines a 1D model based upon Normal Depth calculations, optimised for use on a Digital surface Model (DSM) with rapid 2D flood spreading and is calibrated against JFlow. These equations capture

https://doi.org/10.5194/nhess-2022-188
Preprint. Discussion started: 14 July 2022

Natural Hazards and Earth System Sciences
Discussions

the main controls of the flood routing for shallow, topographically driven flow. Six flood maps at 30 m resolution are created for 20, 50, 100, 200, 500 and 1500 year return period flood events (corresponding to annual exceedance probabilities (AEPs) of 5%, 2.5%, 1%, 0.5% and 0.2% and 0.07% respectively). These are interpolated to derive five intermediate maps between each adjacent pair of the JFlow maps, equally spaced in return period creating a total library of 36 flood maps. Flood Foresight takes inputs of rainfall from numerical weather prediction (NWP) models, river gauge data (both historical and real-time) and forecast streamflow and uses these to select the most appropriate flood map for the location and forecasts daily flood maps out to ten days.

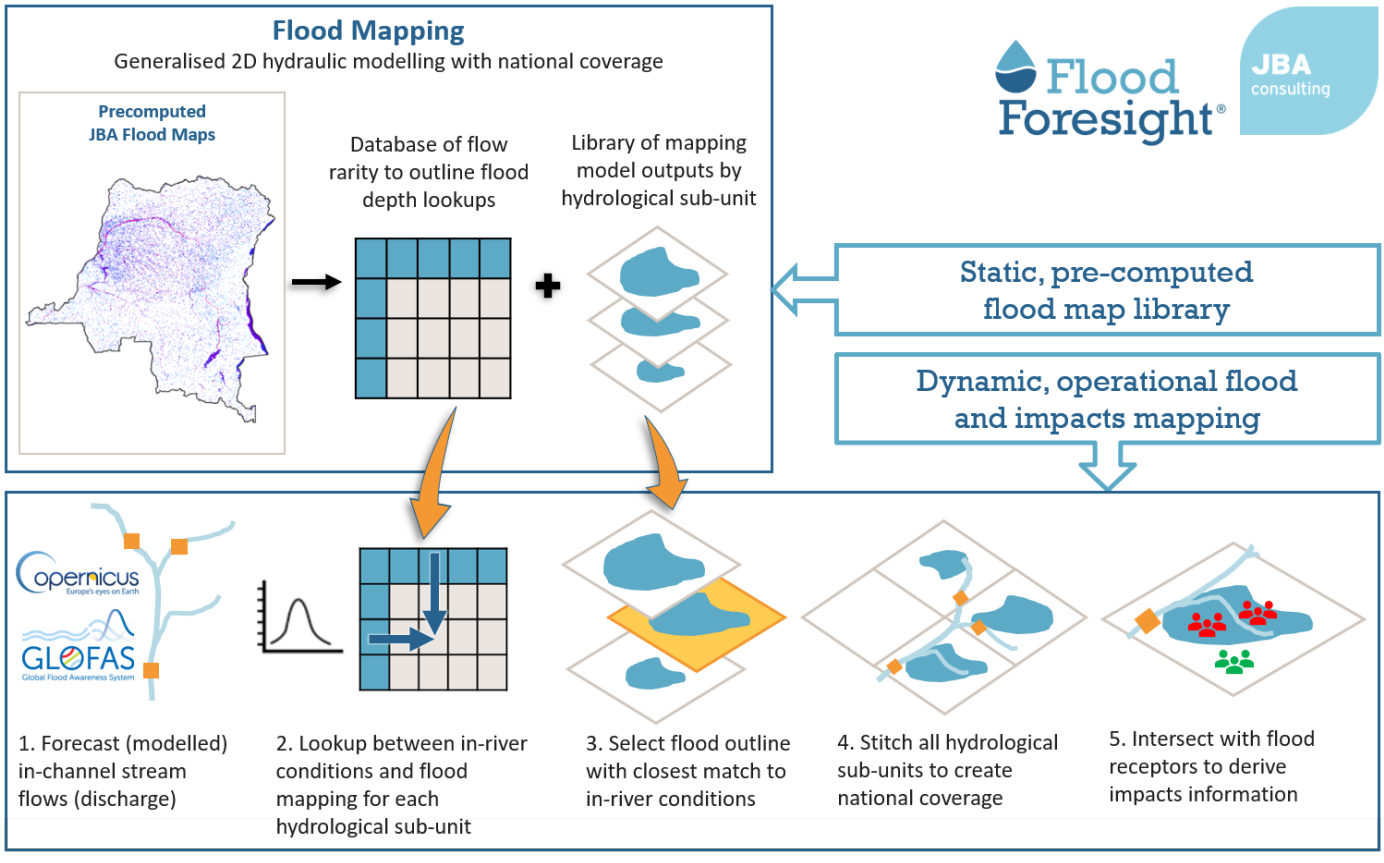

Figure 3. Flood Foresight ensemble forecast flood inundation and impact mapping work flow.

The global (non UK and Ireland) configuration of Flood Foresight uses ensemble streamflow forecast data from the Global Flood Awareness System (GloFAS) (Alfieri et al., 2013; GloFAS, 2021). GloFAS was jointly developed by the European Commission and the European Centre for Medium-Range Weather Forecasts (ECMWF) and is composed of an integrated hydro-meteorological forecasting chain that couples state-of-the-art weather forecasts with a land surface and hydrological model. With its continental scale set-up, GloFAS provides downstream countries with forecasts of upstream river conditions

https://doi.org/10.5194/nhess-2022-188
Preprint. Discussion started: 14 July 2022

up to one month ahead as well as continental and global overviews for large world river basins. Meteorological forecast data are provided by the ECMWF Ensemble (IFS) model, the operational ensemble weather forecasting product of the ECMWF. The meteorological forecast data provide inputs to the land surface module, HTESSEL (Hydrological Tiled ECMWF Scheme for Surface Exchange over Land). HTESSEL simulates the land surface response to the meteorological data, based on simulated interactions with soil conditions, idealised vegetation cover and land cover. From these simulations, HTESSEL outputs forecast global surface and sub-surface flows per grid cell. These simulated flows are then used by a simplified version of the hydrological model LISFLOOD, a one-dimensional (1D) routing model which simulates the movement of the surface and sub-surface flows. The runoff data produced is 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 and Grill, 2013). In summary, the meteorological 51 member ensemble input to the flood forecasting chain allows atmospheric evolution uncertainties to be represented within the ensemble streamflow forecast and the ensemble of inundation flood maps, thus creating a probabilistic flood map forecast, indicating the likelihood of flooding.

3.3 SAR-derived flood maps

A Sentinel-1 (S1A) image was acquired in interferometric wide swath mode (swath width 250 km) around the time of the flood peak at 11:48 on the 12th August 2017. A pre-flood image (February 2017) from the same satellite sensor and track was used to derive the flood map (Fig. 2). The ESA Grid Processing on Demand (GPOD) HASARD service (http://gpod.eo.esa.int/) has been utilised. The flood mapping algorithm (Chini et al., 2017) uses an automated, statistical, hierarchical split-based approach to distinguish between two classes (background and flood) using a pre-flood and flood image. Original SAR images (VV) are pre-proccesed, which involves: precise orbit correction, radiometric calibration, thermal noise removal, terrain correction, speckle reduction and re-projection to the WGS84 coordinate system. The HASARD mapping algorithm removes permanent water bodies, such as the unflooded river water, lakes and reservoirs. Flooded areas beneath vegetation, bridges and near to buildings will not be detected using this method. Flood Foresight forecast flood maps include the river channel and exclude surface features such as vegetation and buildings. To smooth the HASARD flood maps and allow a fairer comparison we apply a morphological closing operation (without impacting the location of the flood extent) to flood fill vegetation and buildings. The wide and braided Brahmaputra River in the Assam region covers a significant area of the selected domain. 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 Flood Foresight forecast inundation maps. The observed flood extent derived from satellite based SAR data at 20 m grid size is re-scaled to match the forecast flood map grid size (30 m) using spline interpolation.

3.4 Forecast data

The Flood Foresight forecast data for the Brahmaputra flood event contains a 51 member ensemble of flood maps indicating flooding extent, produced at a 1-day lead time. Vertically stacking each individual ensemble member flood map and adding vertically across every grid cell combines all ensemble members into a single flood map (all flooded grid cells are set to 1)

https://doi.org/10.5194/nhess-2022-188
Preprint. Discussion started: 14 July 2022


Natural Hazards and Earth System Sciences
Discussions
Open Access
EGU

showing where flooding is possible across all members (ens_{all}). A spatial median flood map is created (ens_{median}) where 26 members or more predict flooding at a particular grid cell location. Each of the ensemble member flood maps for the domain (Fig. 2) are plotted in Figure 4 along with ens_{all}, ens_{median} and the SAR-derived flood map.

https://doi.org/10.5194/nhess-2022-188
Preprint. Discussion started: 14 July 2022

Natural Hazards and Earth System Sciences
Discussions
Open Access
EGU

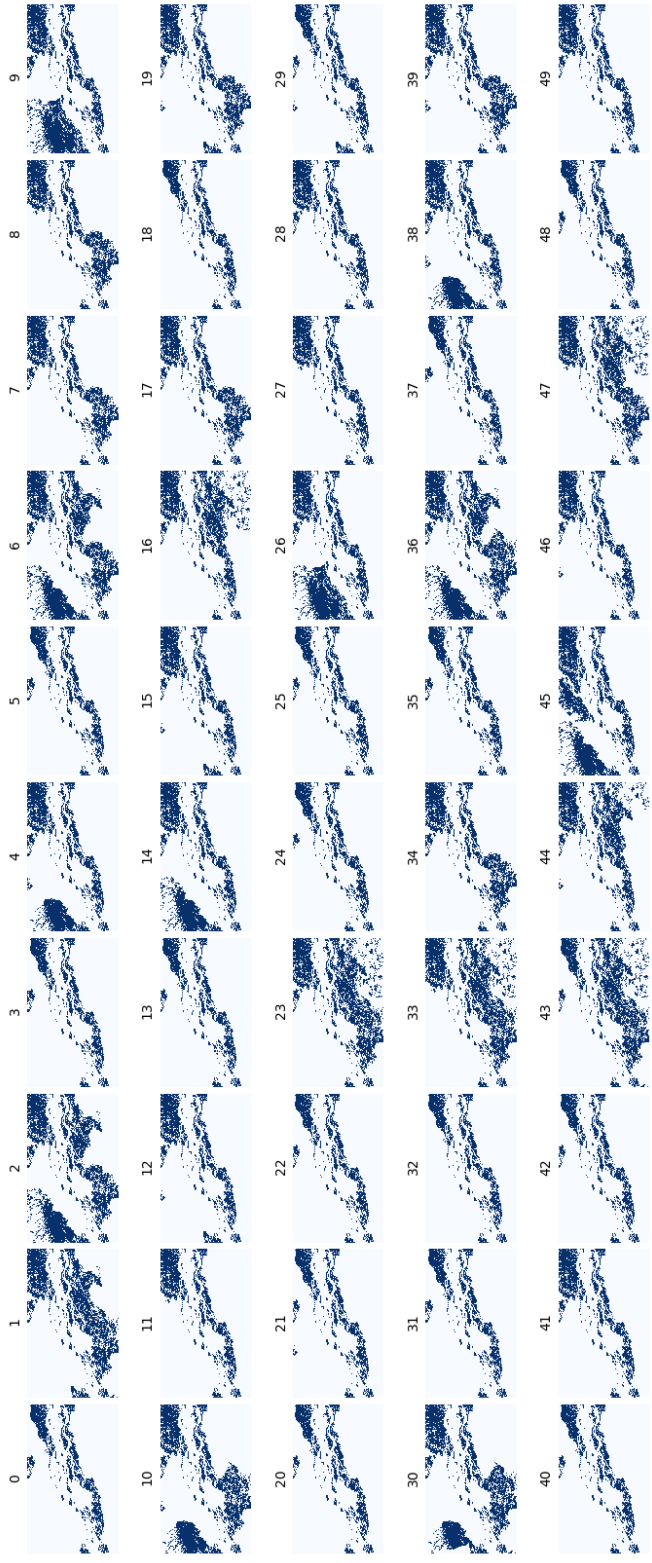

Figure 4. Brahmaputra River, Assam region, August 2017. 51 ensemble member forecast flood maps (labelled 0 to 50), ens_{median} and ens_{all} all at 1-day lead time and the Sentinel-1 SAR-derived flood map.

https://doi.org/10.5194/nhess-2022-188
Preprint. Discussion started: 14 July 2022

Natural Hazards and Earth System Sciences
Discussions

Figure 5 shows the amalgamated probabilistic ensemble forecast indicating the probability of flooding at each grid cell location. This was produced by vertically stacking each ensemble member flood map and adding vertically the number of flooded cells at each grid cell location across all ensemble members. The total is divided by 51 to calculate the probability. Dark blue colours near to the central river channel indicate agreement between all ensemble members and 100% forecast probability of flooding, lighter colours to the north of the river indicate a low probability of flooding.

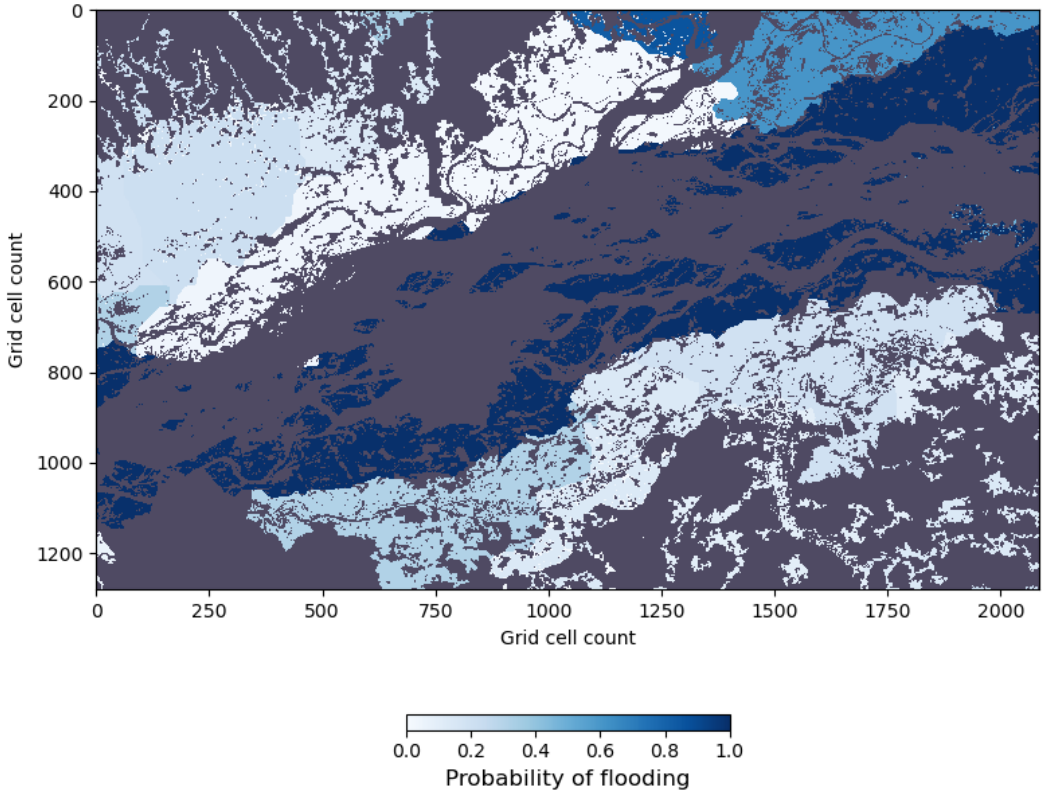

Figure 5. Colour shading from white (low) to dark blue (high) indicate the forecast probability of flooding based on a 1-day lead time, 51 ensemble member flood map forecast for the Brahmaputra River in the Assam region, August 2017. Dark grey areas are unflooded.

4 Results and discussion

To demonstrate an application of our new spatial scale approach to both ensemble forecast flood map evaluation of forecast skill and the spatial spread-skill relationship, we apply the methods outlined in Section 2 to the flooding case described in Section 3.1. First, in Section 4.1 we verify the full ensemble using a spatial scale approach to calculate a skilful scale of agreement

https://doi.org/10.5194/nhess-2022-188
Preprint. Discussion started: 14 July 2022

between each ensemble member and the SAR-derived flood map (Fig. 2) along with the combined ensemble (ens_{all}) and the ensemble spatial median (ens_{median}). We evaluate the location specific spatial skill of the ensemble by calculating categorical scale maps (Section 4.2) for ens_{all}, ens_{median} and a best and worst case ensemble members. In Section 4.3 we evaluate the spatial predictability of the full ensemble and show this on our new *Spatial Spread-Skill (SSS)* map, indicating regions where the ensemble is over-, under- or well-spread.

4.1 Ensemble spatial scale evaluation

Here we investigate how a scale-selective approach can be useful for extracting meaningful information from a flood map ensemble forecast where multiple forecast flood maps represent equally likely flooding scenarios (Fig. 4). A minimum skilful scale (where $FSS > FSS_T$) has been calculated for each individual member flood map, ens_{all} and ens_{median}. The results in Figure 6 show that individual ensemble member spatial skill varies considerably with FSS at grid level ranging from 0.35 to 0.59. One member ens_1, which would usually be disregarded as an outlier due to its low probability, outperformed all other members significantly with a skilful scale achieved at a neighbourhood size of $n = 3$. The combined ens_{all} showed more skill at grid level and smaller neighbourhood sizes compared with ens_{median}, both however exceeded FSS_T at $n = 41$, or 600 m. At neighbourhood sizes greater than $n = 41$, ens_{median} outperformed ens_{all}. An average agreement discrepancy distance of more than 600 m may appear large, however it is worth remembering the magnitude of this flood event in which the flood width exceeds tens of kilometers. There is a cluster of members showing similar skill to ens_{median} and ens_{all} and a second cluster, with more ensemble variation but indicating lower skill than the first cluster. These results show that all ensemble member flood maps, including outliers, should be considered individually as possible future flooding scenarios. Spatial variations across individual ensemble members indicate that it is not meaningful to consider only the ensemble median flood map to represent the information within the full ensemble.

4.2 Ensemble spatial predictability

The scale-selective skill scores calculated for different aspects of the ensemble forecast give a domain-averaged score and skilful scale. To understand location specific spatial predictability of the ensemble forecast, categorical scale maps are calculated and presented in Figure 7. These show how the agreement scale (Section 2.2) varies with location for (a) ens_{all}, (b) ens_{median}, (c) ens_1, the 'best' performing ensemble member and (d) ens_{21}, the 'worst' performing ensemble member. The ensemble summary map, ens_{all} (Fig. 7 (a)) captures most of the observed flooding (in grey) with small regions of under-prediction (red). However, as you might expect to see by including every potential flooding realisation there are significant regions of over-prediction (blue) or false alarm. The region of over-prediction to the south of the river is less evident in the ens_{median} categorical scale map (Fig. 7 (b)) which performs worse to the north by under-predicting flooding here. This flooding is captured well by ens_1 (Fig. 7 (c)) and in particular close to a confluence zone where the Subansiri River joins the Brahmaputra (grid cell location (1100, 250)). This ties in with the high rainfall totals accumulated just to the north of this

https://doi.org/10.5194/nhess-2022-188
Preprint. Discussion started: 14 July 2022

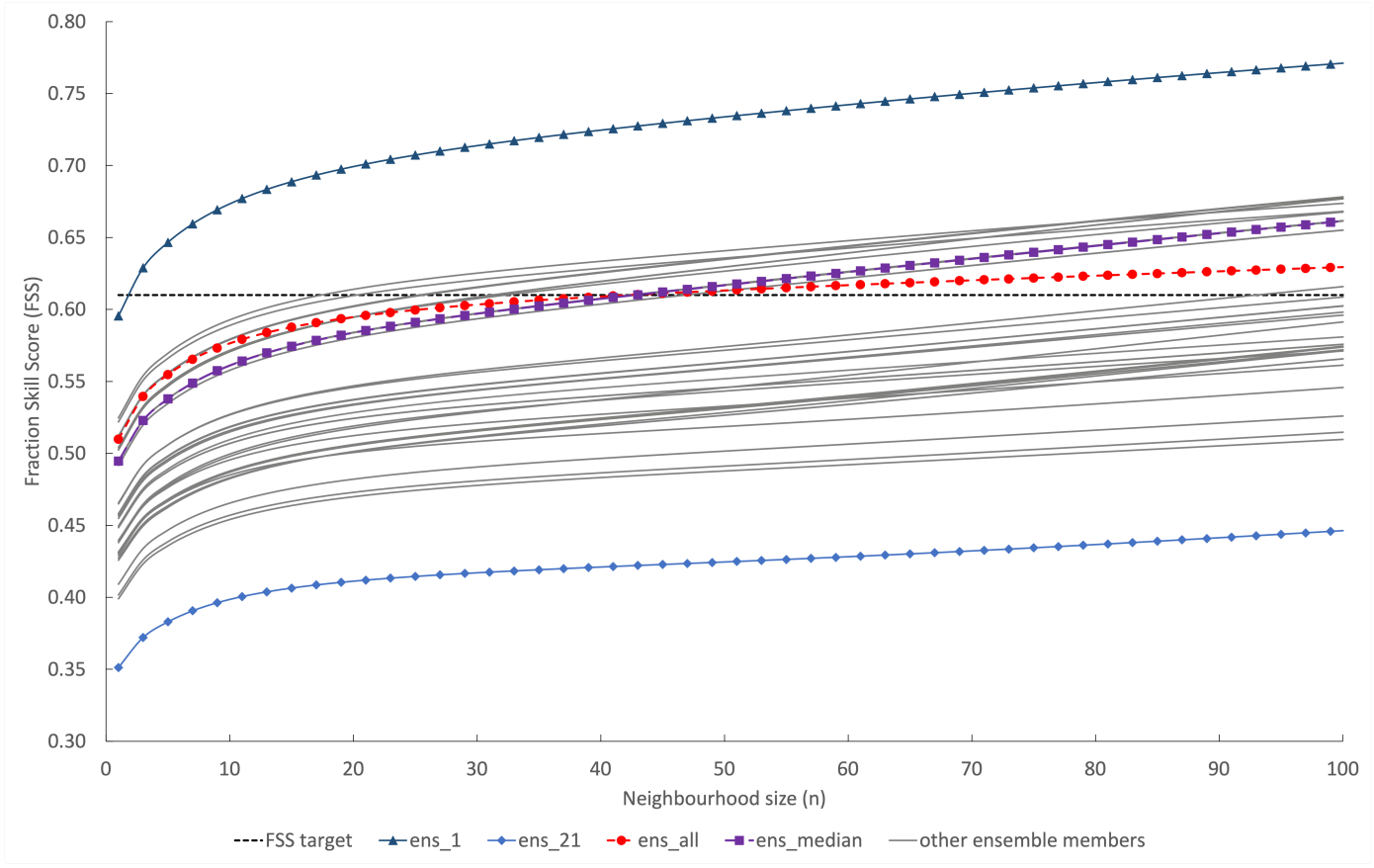

Figure 6. The spatial skill of each individual ensemble member forecast flood extent is evaluated along with the ens_{median} (a spatial median where 26 or more members predict flooding at a grid cell location) and ens_{all} (flooded grid cells from all ensemble members are combined). The FSS is calculated at increasing neighbourhood sizes to determine the scale at which the forecast becomes skilful at capturing the observed flood (FSS_T).

region associated with localised very heavy rainfall. A region of under-prediction at grid cell location (750, 750) is missed by all members. A closer inspection of the DTM or surface features included/excluded in the hydraulic modelling, such as embankment heights, may indicate how this modelling could be improved. The 'worst' performing ensemble member ens_{21} (Fig. 7 (d)) accurately predicts flooding closer to the river channel, however under-prediction to the north along with over-prediction to the south show where the forecast was inaccurate. Categorical scale maps enable different ensemble flood map presentations to be evaluated so that the most useful presentation method can be determined for a particular flooding situation.

https://doi.org/10.5194/nhess-2022-188
Preprint. Discussion started: 14 July 2022

Natural Hazards and Earth System Sciences
Discussions

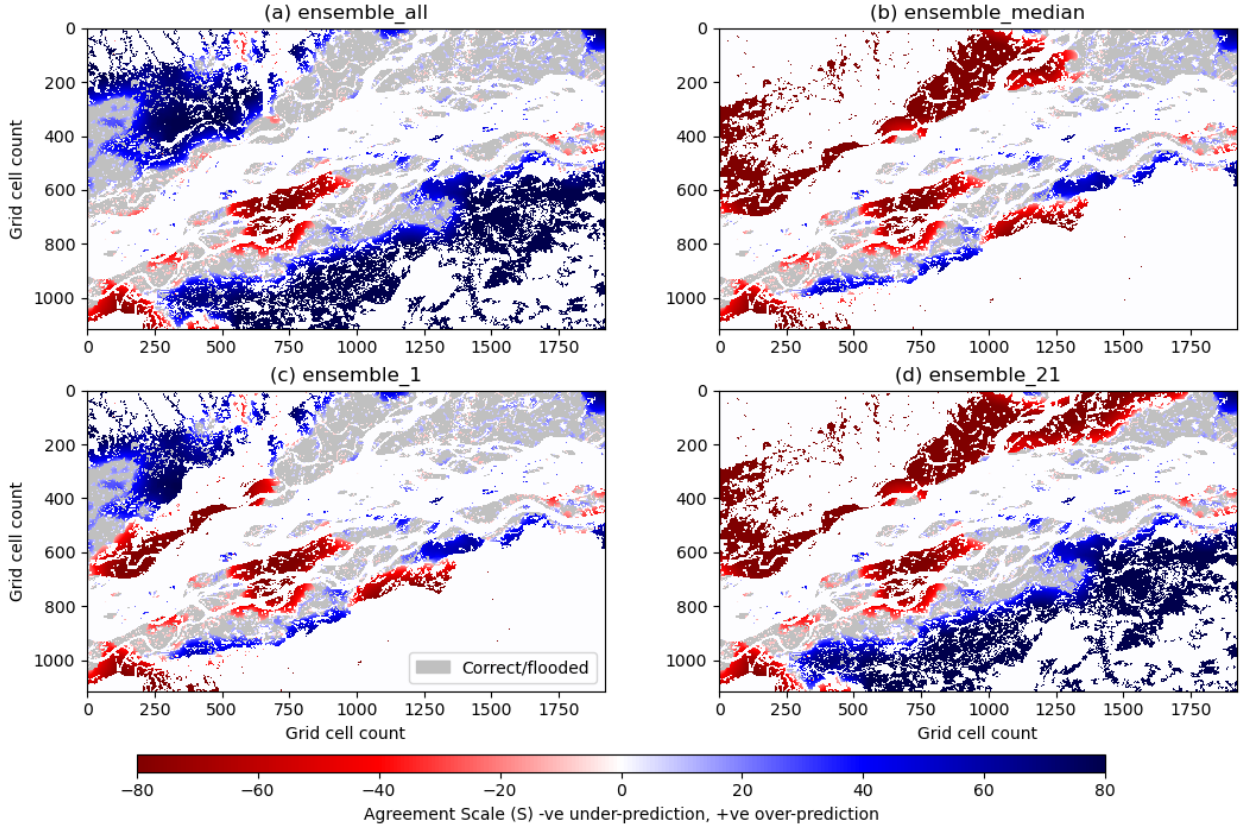

Figure 7. Categorical scale maps for (a) ens_{all} (flooded grid cells from all ensemble members are combined), (b) ens_{median} (a spatial median where 26 or more members predict flooding at a grid cell location), (c) individual ensemble member 1 and (d) individual ensemble member 21. Red areas indicate where the forecast is under-predicted and blue regions represent over-prediction. The colour shade gives the scale of agreement between the forecast and the observed flooding with lighter shading indicating a smaller agreement scale. Each grid cell is 30 m x 30 m.

4.3 Ensemble spatial spread-skill

To evaluate the location specific skill of the full ensemble, one option would be to calculate 51 categorical scale maps. This approach maintains the spatial detail held within each of the ensemble member flood maps, although does require multiple visual comparisons to be made by the flood forecaster or modeller, which takes time and effort. The categorical scale maps do not evaluate the ensemble spatial spread. To address this, we present a new summary ensemble Spatial Spread-Skill (SSS) map (Section 2.4) showing the spread-skill of the full ensemble forecast and keeping the location specific detail. All ensemble members are included in this analysis which evaluates both the spatial skill and the ensemble spatial spread of the forecast

https://doi.org/10.5194/nhess-2022-188
Preprint. Discussion started: 14 July 2022


Natural Hazards and Earth System Sciences
Discussions

against the remotely observed flooding extent.

Figure 8 shows how the average ensemble/ensemble agreement scale in (a) i.e. $S_{ij}^{A(\overline{mm})}$ calculated at each grid cell (representing ensemble *spread*) compares with the average ensemble/observed scale in (b) i.e. $S_{ij}^{A(\overline{mo})}$ (representing ensemble *skill*) along with the hexbin scatter plot in (c) which compares (a) and (b) to indicate the spatial spread-skill of the forecast. The hexagonal tessellation is used so that the distances along the diagonal are on the same scale as those along the abscissa and ordinate. Three numbered areas identify different ensemble spread-skill relationships. Area 1 shows that the agreement between ensemble members is close, but that they disagree with the observed flooding extent. This is displayed in orange shades as an under-spread or miss region in the SSS map, Figure 9. This is the region close to the confluence area described in Section 4.2. In area 2 on Figure 8, both (a) and (b) are in agreement at grid level, which indicates the ensemble is well-spread; these are shown in white on Figure 9. Away from the miss and well-spread regions in Figure 8, the overall impression is that the ensemble spread-skill lies below the 1:1 line and is over-spread, indicated by area 3. This corresponds to purple shading on the SSS map (Fig. 9). Overall Figure 8 tells us that the spread-skill relationship for this example case study is not uniform across the domain but is in fact location specific. The SSS map displays where the spatial spread-skill is over-, under-, or well-spread. From this we can infer how well the ensemble forecasting system encompasses the multiple sources of uncertainty and how meaningful the ensemble predictability of flood inundation actually is. This is an important evaluation tool for all catchments and becomes essential for validating flood forecasts in un-gauged or partially gauged rivers. A simulation library approach relies on the accuracy of the return period thresholds set, the (ensemble) forecast streamflow and the accuracy of the flood inundation map for a given streamflow. The approaches presented here enable these system attributes to be evaluated even where observed streamflow are limited or erroneous.

https://doi.org/10.5194/nhess-2022-188
Preprint. Discussion started: 14 July 2022

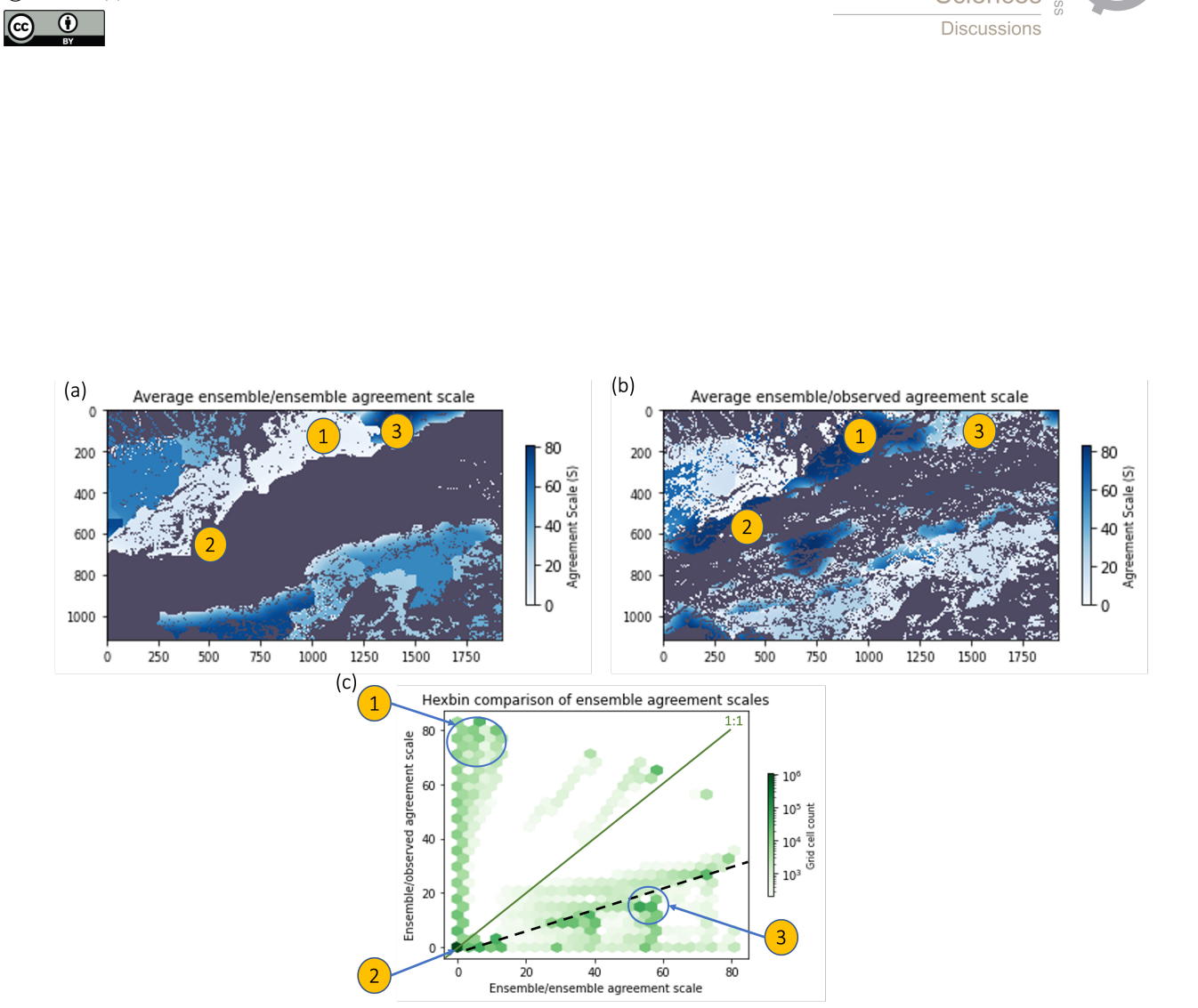

Figure 8. The average agreement scale map of each unique pair of forecast ensemble flood maps is plotted in (a) and between each ensemble member compared against the observed SAR-derived flood map in (b). A binned histogram scatter plot compares (a) and (b) to indicate the spatial spread-skill of the forecast ensemble. For a perfect ensemble forecast the average agreement scale between ensemble members should match the agreement scale between the ensemble forecast and observed flood map, i.e. they should align along the 1:1 line. The main trend from the whole domain is indicated with a dashed black line which shows that generally the forecast ensemble is over-spread (e.g. area 3), however there are variations in the spatial spread-skill across the domain with regions that are under-spread and misses (e.g. area 1) along with regions where all members show accurate spatial predictability of the observed flooded/unflooded grid cells (e.g. area 2) and the ensemble is considered well-spread.

https://doi.org/10.5194/nhess-2022-188
Preprint. Discussion started: 14 July 2022

Natural Hazards and Earth System Sciences
Discussions
Open Access
EGU

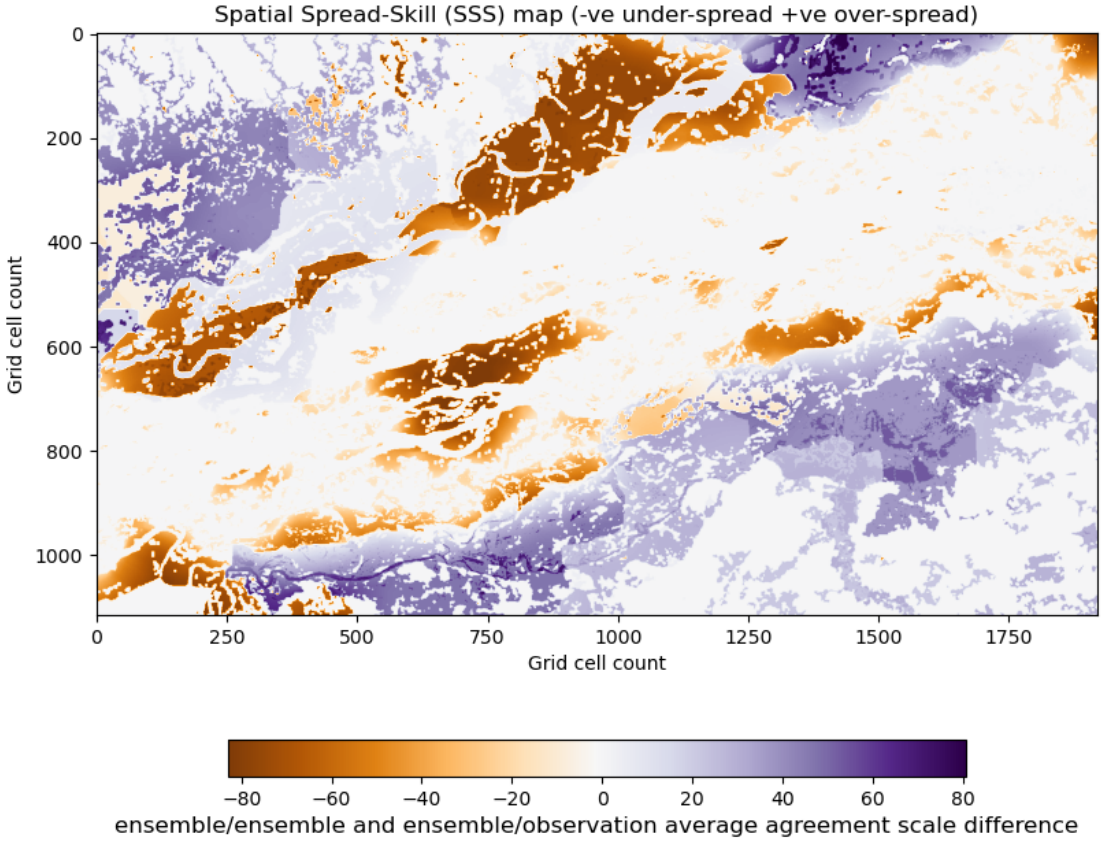

Figure 9. The Spatial Spread-Skill (SSS) map shows the difference between the ensemble/ensemble and the ensemble/observed average agreement scales at each grid cell. Negative orange areas indicate where the ensemble is under-spread and positive purple areas indicate where the ensemble is over-spread. White areas areas show that the the average agreement scales match and indicate good spatial spread-skill. Lighter shades show the average agreement scales are more closely matched with very dark shades indicating over- (dark purple) or under- (dark orange) prediction by the ensemble forecast.

https://doi.org/10.5194/nhess-2022-188
Preprint. Discussion started: 14 July 2022

5 Conclusions

The uncertainty information given by forecast ensemble flood maps is determined in large part by initial condition perturbations at the top of the hydro-meteorological forecast chain within the NWP system. Presently, there is limited understanding or evaluation of how these meteorological uncertainties link to mapped flooding predictability where multiple other sources of uncertainty exist. An evaluation of the spatial predictability and the spread-skill relationship of the ensemble flood map forecast will enable an improved understanding of the performance of the forecast system. Uncertainties in other parts of the forecast chain are not truly represented by the ensemble flood maps and evaluating their skill is important for understanding the likelihood of flooding that the ensemble flood maps capture. In this paper, we present a new scale-selective approach to assess the spatial predictability and spread-skill of an ensemble flood map forecast by comparing against a satellite SAR-derived observation of flooding extent. By calculating a skilful scale at each grid cell for every unique ensemble member pair we can determine the ensemble *spatial spread*, and between every ensemble member and the SAR-derived flood map we can determine the ensemble *spatial skill*. The difference between these skilful scales can be mapped onto our new Spatial Spread Skill (SSS) map which tells us for each specific location in the domain whether the ensemble is over-, under- or well-spread. The methods are applied to an example flooding event of the Brahmaputra in the Assam region of India in August 2017.

In operational practice there are multiple options of ensemble flood map presentation type for delivery to end-users and decision makers. Using a scale-selective approach we have evaluated the performance of individual ensemble members, a combined total ensemble and the spatial ensemble median compared to a SAR-derived observation of flooding extent. An important aspect of developing an inundation flood forecasting system is to determine the most useful way to present a spatial ensemble forecast. Other options could be to exclude ensemble member outliers, to spatially cluster similar ensemble members into groups of flooding extent or to present a most likely, best and worst case ensemble flood map. Whichever presentation method is chosen, this should be fully explored using the methods described here to evaluate the ensemble performance of historical flooding events. We found for this example flooding event that one ensemble member significantly outperformed the combined and median flood maps and that potentially in some cases this member would have been excluded as an outlier. This tells us that the ensemble spatial median could miss vital flooding information and that all members should be considered as potential future flooding scenarios.

Through mapping the spatial-spread skill relationship, which varies with location, links can be made between the spatial variations in spread-skill and the physical characteristics of the flooding event. An understanding of the spatial predictability is particularly important for un-gauged catchments where the calibration of both forecast streamflow and return period thresholds (used to select the simulation library flood map) is currently not possible. Ideally, in operational practice, these spatial verification approaches including the categorical scale and SSS maps could be calculated and stored routinely as flooding events coincide with SAR-derived or other remotely observed flood maps to build up a verification catalogue/database. This evaluation could be used to investigate the spatial spread-skill model performance under different scenarios such as forecast lead

https://doi.org/10.5194/nhess-2022-188
Preprint. Discussion started: 14 July 2022

time, month or season, or flood type. More locally, the impact of an improved DTM or the inclusion of a Digital Surface Model (DSM) or other surface features in the hydraulic model such as embankments could be considered. Over time, an evaluation database would improve our understanding of the spatial predictability of an ensemble flood map system and how well the uncertainties present are represented by the ensemble forecast.

Code and data availability. The functions used to evaluate the ensemble forecast flood maps using a scale-selective approach along with the SAR-derived flood maps are available on the following Zenodo page: https://doi.org/10.5281/zenodo.6603101 (Hooker et al., 2022b). The forecast flood maps from the JBA Flood Foresight system are commercial data used under license for this study.

Author contributions. JB and KS provided the forecast data. HH wrote the algorithms and ran the experiments, with input from SD, DM, JB and KS. HH prepared the manuscript with contributions from all the co-authors.

Competing interests. The authors declare that no competing interests are present.

Acknowledgements. This work was supported 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). SD and DM were funded in part by the UK EPSRC DARE project (EP/P002331/1). SD also received funding from NERC National Centre for Earth Observation.

https://doi.org/10.5194/nhess-2022-188
Preprint. Discussion started: 14 July 2022

References

Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., and Pappenberger, F.: 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, 2013.

Alfonso, L., Mukolwe, M. M., and Di Baldassarre, G.: Probabilistic Flood Maps to support decision-making: Mapping the Value of Information, Water Resources Research, 52, 1026–1043, https://doi.org/10.1002/2015WR017378, 2016.

Anderson, S. R., Csima, G., Moore, R. J., Mittermaier, M., and Cole, S. J.: Towards operational joint river flow and precipitation ensemble verification: considerations and strategies given limited ensemble records, Journal of Hydrology, 577, 123 966, https://doi.org/10.1016/j.jhydrol.2019.123966, 2019.

ASDMA: Assam State Disaster Management Authority Flood Alert, http://sdmassam.nic.in/download/alerts/10.08.2017.jpg, last access 10th November 2021, 2017.

Ben Bouallègue, Z. and Theis, S. E.: Spatial techniques applied to precipitation ensemble forecasts: From verification results to probabilistic products, Meteorological Applications, 21, 922–929, https://doi.org/10.1002/met.1435, 2014.

Beven, K.: Facets of uncertainty: Epistemic uncertainty, non-stationarity, likelihood, hypothesis testing, and communication, Hydrological Sciences Journal, 61, 1652–1665, https://doi.org/10.1080/02626667.2015.1031761, 2016.

Boelee, L., Lumbroso, D. M., Samuels, P. G., and Cloke, H. L.: Estimation of uncertainty in flood forecasts—A comparison of methods, Journal of Flood Risk Management, https://doi.org/10.1111/jfr3.12516, 2019.

Bradbrook, K.: JFLOW: A multiscale two-dimensional dynamic flood model, Water and Environment Journal, https://doi.org/10.1111/j.1747-6593.2005.00011.x, 2006.

Buizza, R.: Potential forecast skill of ensemble prediction and spread and skill distributions of the ECMWF ensemble prediction system, Monthly Weather Review, 125, 99–119, https://doi.org/10.1175/1520-0493(1997)125<0099:PFSOEP>2.0.CO;2, 1997.

Chen, X., Yuan, H., and Xue, M.: Spatial spread-skill relationship in terms of agreement scales for precipitation forecasts in a convection-allowing ensemble, Quarterly Journal of the Royal Meteorological Society, 144, 85–98, https://doi.org/10.1002/qj.3186, 2018.

Chini, M., Hostache, R., Giustarini, L., and Matgen, P.: 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, 2017.

Cloke, H. L. and Pappenberger, F.: Ensemble flood forecasting: A review, Journal of Hydrology, 375, 613–626, https://doi.org/10.1016/j.jhydrol.2009.06.005, 2009.

Cooper, E. S., Dance, S. L., Garcia-Pintado, J., Nichols, N. K., and Smith, P. J.: Observation impact, domain length and parameter estimation in data assimilation for flood forecasting, Environmental Modelling Software, 104, 199–214, https://doi.org/10.1016/J.ENVSOFT.2018.03.013, 2018.

Cooper, E. S., Dance, S. L., García-Pintado, J., Nichols, N. K., and Smith, P. J.: 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, 2019.

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

Dasgupta, A., Grimaldi, S., Ramsankaran, R., Pauwels, V. R. N., Walker, J. P., Chini, M., Hostache, R., and Matgen, P.: Flood Mapping Using Synthetic Aperture Radar Sensors From Local to Global Scales, pp. 55–77, https://doi.org/10.1002/9781119217886.ch4, 2018a.

https://doi.org/10.5194/nhess-2022-188
Preprint. Discussion started: 14 July 2022

Dasgupta, A., Grimaldi, S., Ramsankaran, R. A., Pauwels, V. R., and Walker, J. P.: Towards operational SAR-based flood mapping using neuro-fuzzy texture-based approaches, Remote Sensing of Environment, 215, 313–329, https://doi.org/10.1016/j.rse.2018.06.019, 2018b.

Dasgupta, A., Hostache, R., Ramasankaran, R., Schumann, G. J., Grimaldi, S., Pauwels, V. R. N., and Walker, J. P.: On the impacts of observation location, timing and frequency on flood extent assimilation performance, Water Resources Research, https://doi.org/10.1029/2020wr028238, 2021a.

Dasgupta, A., Hostache, R., Ramsankaran, R. A., Schumann, G. J., Grimaldi, S., Pauwels, V. R., and Walker, J. P.: A Mutual Information-Based Likelihood Function for Particle Filter Flood Extent Assimilation, Water Resources Research, 57, 1–28, https://doi.org/10.1029/2020WR027859, 2021b.

Dey, S. R., Roberts, N. M., Plant, R. S., and Migliorini, S.: 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, 2016.

Dhar, O. N. and Nandargi, S.: A study of floods in the Brahmaputra basin in India, International Journal of Climatology, 20, 771–781, https://doi.org/10.1002/1097-0088(20000615)20:7<771::AID-JOC518>3.0.CO;2-Z, 2000.

Dhar, O. N. and Nandargi, S.: Hydrometeorological Aspects of Floods in India, pp. 1–33, Springer Netherlands, Dordrecht, https://doi.org/10.1007/978-94-017-0137-2_1, 2003.

Di Mauro, C., Hostache, R., Matgen, P., Pelich, R., Chini, M., Van Leeuwen, P. J., Nichols, N. K., and Blöschl, G.: Assimilation of probabilistic flood maps from SAR data into a coupled hydrologic-hydraulic forecasting model: A proof of concept, Hydrology and Earth System Sciences, 25, 4081–4097, https://doi.org/10.5194/hess-25-4081-2021, 2021.

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., and Cloke, H. L.: Continental and global scale flood forecasting systems, Wiley Interdisciplinary Reviews: Water, 3, 391–418, https://doi.org/10.1002/wat2.1137, 2016.

ESA: ICEYE commercial satellites join the EU Copernicus programme, https://www.esa.int/Applications/Observing_the_Earth/Copernicus/ICEYE_commercial_satellites_join_the_EU_Copernicus_programme, last access 28th October 2021, 2021.

EU Science Hub: The Joint Research Centre launches a revolutionary tool for monitoring ongoing floods worldwide as part of the Copernicus Emergency Management Service, 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, 2021.

Floodlist: India – Third Wave of Flooding Hits Assam, 2 Million Affected, http://floodlist.com/asia/india-assam-floods-august-2017, last access 10th November 2021, 2017.

García-Pintado, J., Mason, D. C., Dance, S. L., Cloke, H. L., Neal, J. C., Freer, J., and Bates, P. D.: 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, 2015.

GFM: GloFAS global flood monitoring (GFM), https://www.globalfloods.eu/technical-information/glofas-gfm/, last access 28th October 2021, 2021.

GloFAS: GloFAS Methods, https://www.globalfloods.eu/general-information/glofas-methods/, last access 15th November 2021, 2021.

Hooker, H., Dance, S. L., Mason, D. C., Bevington, J., and Shelton, K.: Spatial scale evaluation of forecast flood inundation maps, https://doi.org/10.31223/X5DG9C, 2022a.

Hooker, H., Dance, S. L., Mason, D. C., Bevington, J., and Shelton, K.: Ensemble flood map spatial verification [Data set], https://doi.org/10.5281/zenodo.6603101, 2022b.

https://doi.org/10.5194/nhess-2022-188
Preprint. Discussion started: 14 July 2022

Horritt, M. S., Mason, D. C., and Luckman, A. J.: 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, 2001.

Hossain, S., Cloke, H. L., Ficchì, A., Turner, A. G., and Stephens, E. M.: 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, 2021.

Hostache, R., Chini, M., Giustarini, L., Neal, J., Kavetski, D., Wood, M., Corato, G., Pelich, R. M., and Matgen, P.: Near-Real-Time Assimilation of SAR-Derived Flood Maps for Improving Flood Forecasts, Water Resources Research, 54, 5516–5535, https://doi.org/10.1029/2017WR022205, 2018.

Hostache, R.: A first evaluation of the future CEMS systematic global flood monitoring product, https://events.ecmwf.int/event/222/contributions/2274/attachments/1280/2347/Hydrological-WS-Hostache.pdf, last access 4th August 2021, 2021.

Konapala, G., Kumar, S. V., and Khalique Ahmad, S.: Exploring Sentinel-1 and Sentinel-2 diversity for flood inundation mapping using deep learning, ISPRS Journal of Photogrammetry and Remote Sensing, 180, 163–173, https://doi.org/10.1016/J.ISPRSJPRS.2021.08.016, 2021.

Lehner, B. and Grill, G.: Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems, Hydrological Processes, 27, 2171–2186, https://doi.org/https://doi.org/10.1002/hyp.9740, 2013.

Leutbecher, M. and Palmer, T. N.: Ensemble forecasting, Journal of Computational Physics, 227, 3515–3539, https://doi.org/10.1016/J.JCP.2007.02.014, 2008.

Lorenz, E. N.: The predictability of a flow which possesses many scales of motion, Tellus, 21, 289–307, https://doi.org/10.3402/tellusa.v21i3.10086, 1969.

Mason, D. C., Schumann, G. J., Neal, J. C., Garcia-Pintado, J., and Bates, P. D.: 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, 2012.

Mason, D. C., Dance, S. L., Vetra-Carvalho, S., and Cloke, H. L.: 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, 2018.

Mason, D. C., Dance, S. L., and Cloke, H. L.: 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, 2021a.

Mason, D. C., Bevington, J., Dance, S. L., Revilla-Romero, B., Smith, R., Vetra-Carvalho, S., and Cloke, H. L.: 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, 2021b.

Matthews, G., Barnard, C., Cloke, H., Dance, S. L., Jurlina, T., Mazzetti, C., and Prudhomme, C.: Evaluating the impact of post-processing medium-range ensemble streamflow forecasts from the European Flood Awareness System, Hydrology and Earth System Sciences, 26, 2939–2968, https://doi.org/10.5194/hess-26-2939-2022, 2022.

Palash, W., Akanda, A. S., and Islam, S.: The record 2017 flood in South Asia: State of prediction and performance of a data-driven requisitely simple forecast model, Journal of Hydrology, 589, https://doi.org/10.1016/j.jhydrol.2020.125190, 2020.

Pappenberger, F., Beven, K. J., Hunter, N. M., Bates, P. D., Gouweleeuw, B. T., Thielen, J., and de Roo, A. P.: Cascading model uncertainty from medium range weather forecasts (10 days) through a rainfall-runoff model to flood inundation predictions within the European Flood Forecasting System (EFFS), Hydrology and Earth System Sciences, 9, 381–393, https://doi.org/10.5194/hess-9-381-2005, 2005.

https://doi.org/10.5194/nhess-2022-188
Preprint. Discussion started: 14 July 2022


Natural Hazards and Earth System Sciences
Discussions

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

Renner, M., Werner, M. G., Rademacher, S., and Sprokkereef, E.: Verification of ensemble flow forecasts for the River Rhine, Journal of Hydrology, 376, 463–475, https://doi.org/10.1016/j.jhydrol.2009.07.059, 2009.

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

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

Tavus, B., Kocaman, S., Nefeslioglu, H. A., and Gokceoglu, C.: A fusion approach for flood mapping using sentinel-1 and sentinel-2 datasets, International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives, 43, 641–648, https://doi.org/10.5194/isprs-archives-XLIII-B3-2020-641-2020, 2020.

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

Zappa, M., Jaun, S., Germann, U., Walser, A., and Fundel, F.: Superposition of three sources of uncertainties in operational flood forecasting chains, Atmospheric Research, 100, 246–262, https://doi.org/10.1016/J.ATMOSRES.2010.12.005, 2011.