RESEARCH ARTICLE

Meteorological Applications RMetS

Flood hazard risk forecasting index (FHRFI) for urban areas: The Hurricane Harvey case study

Toni Jurlina[1,2] | Calum Baugh[1] | Florian Pappenberger[1] | Christel Prudhomme[1,3,4]

[1]Forecast Department, European Centre for Medium-Range Weather Forecasts (ECMWF), Reading, UK

[2]Croatian Meteorological and Hydrological Service, Zagreb, Croatia

[3]Centre for Ecology & Hydrology, Wallingford, UK

[4]Geography and Environment Department, Loughborough University, Loughborough, UK

Correspondence
Toni Jurlina, Forecast Department, European Centre for Medium-Range Weather Forecasts (ECMWF), Reading, UK.
Email: toni.jurlina@ecmwf.int

Abstract

Hurricane Harvey caused at least 70 confirmed deaths, with estimated losses in the Houston urban area of Texas reaching above US$150 billion, making it one of the costliest natural disasters ever in the United States. The study tests two types of forecast index to provide surface flooding (inundation) warning over the Houston area: a meteorological index based on a global numerical weather prediction (NWP) system, and a new combined meteorological and land surface index, the flood hazard risk forecasting index (FHRFI), where land surface is used to condition the meteorological forecast. Both indices use the total precipitation extreme forecast index (EFI) and shift of tails (SoT) products from the European Centre for Medium-Range Weather Forecasts (ECMWF) medium-range ensemble forecasting system (ENS). Forecasts at the medium range (3–14 days ahead) were assessed against 153 observed National Weather Service (NWS) urban flood reports over the Houston urban area between August 26 and 29, 2017. It is shown that the method provides skilful forecasts up to four days ahead using both approaches. Moreover, the FHRFI combined index has a hit ratio of up to 74% at 72 hr lead time, with a false-alarm ratio of only 45%. This amounts to a statistically significant 20% increase in performance compared with the meteorological indices. This first study demonstrates the importance of including land-surface information to improve the quality of the flood forecasts over meteorological indices only, and that skilful flood warning in urban areas can be obtained from the NWP using the FHRFI.

KEYWORDS
ensemble prediction, extreme weather, flash flood

1 | INTRODUCTION

1.1 | Flood and society damages

Flooding is a devastating natural hazard, with over 1 million deaths attributed to storms and floods between 1970 and 2012, and over US$400 billion in economic losses at the global scale (Golnaraghi *et al.*, 2014). In urban areas, the socioeconomic damage from flooding is greater than in non-urban areas (Hapuarachchi *et al.*, 2011). This is due to increased surface flooding (inundation) from changes in land surface (where runoff can increase by two- to six-fold; Ramachandra and Kumar, 2008), and increased population and infrastructure exposure. With the predicted rise of megacities around the world (Kraas *et al.*, 2013), surface flooding in the

urban environment is likely to continue to be a challenge facing emergency responders.

1.2 | Flood risk management at national and continental scales

The management of flood hazards and risks is important for public safety, but also to reduce extensive damage. For more effective management and strategic allocation of emergency responses during crisis, most flood early warning systems (EWS) operate at a national level (and require input data and local information) (Alfieri *et al.*, 2014) or a regional level (Hapuarachchi *et al.*, 2011). At that scale, operational EWS for surface water flooding rely on simplified forecast indices representative of extreme rainfall, for which there is no need for additional calibration parameters (Alfieri *et al.*, 2012). The extreme rainfall alert (Hurford *et al.*, 2012) (ERA) of the British Flood Forecasting Centre (FFC) and the Swiss warning system for point precipitation (Alfieri *et al.*, 2016) are two examples of such systems, both designed to give early indications on upcoming severe rainfall events potentially leading to surface water flooding, including in urban environments (Alfieri *et al.*, 2012). However, the large data requirement for continent-scale EWS limits their current application, with the European Flood Awareness System (EFAS) (Thielen *et al.*, 2009) and flash flood guidance (FFG) (Ntelekos *et al.*, 2006) being prominent examples for application in Europe and the world (Gourley *et al.*, 2012). In the case of the EFAS, delivered information is used by the Emergency Response Coordination Centre (ERCC) to compile reports on the flood situation and outlook and to co-ordinate the emergency response at the continental scale (Emerton *et al.*, 2016).

Currently, forecasting of surface water flooding, including urban flooding, is possible where numerical weather prediction (NWP) systems exist to drive flood-generation algorithms. Such systems can rely on, for example, limited area models (LAMs) or radar/nowcasting methods, and applications to flash floods exist in Europe (as part of the EFAS) (Thielen *et al.*, 2009; Raynaud *et al.*, 2015), in northern America (Gourley *et al.*, 2017), southern Africa (Georgakakos *et al.*, 2013), Australia and other regions (Hapuarachchi and Wang, 2008). Where such systems are absent, global NWP are possible alternatives owing to their ability to capture the synoptic signals that can result in localized extreme events at the medium range (3—14 days ahead), although they generally do not reproduce fine spatial scale processes, such as convection, also responsible for intense precipitation (Emerton *et al.*, 2016) and driving pluvial and urban flooding.

Surface water flooding arises from the occurrence of extreme rainfall rates (Doswell *et al.*, 1996), which are then conditioned by land-surface factors including impervious urban surfaces and river basin geometry (Penna *et al.*, 2013).

Identifying the occurrence of extreme rainfall rates from the NWP forecasts can be achieved using indices such as the extreme forecast index (EFI) (Lalaurette, 2003; Zsótér, 2006) and shift of tails (SoT) (Zsótér, 2006), both designed to identify in the forecasts abnormal situations compared with the expected modelled climatology. Applied to the European Centre for Medium-Range Weather Forecast (ECMWF) medium-range ensemble forecasting system (ENS), they were successful at identifying an extreme precipitation event in Greece up to three days in advance (Hewson and Tsonevsky, 2016).

1.3 | Land-surface influence on surface flooding

The NWPs are designed only to inform on the atmospheric forcing, but they do not account for the influence of the land surface as flood generation or natural attenuation mechanisms, known to condition when and where flood events do occur (Hapuarachchi and Wang, 2008). Because ignoring land surface properties might underestimate the risk of flooding, most flood warning systems transform the meteorological forecasts before issuing flood warnings. Physically based distributed hydrological models, known to be good tools for simulating hydrological extremes (Cole *et al.*, 2006), are often computationally demanding, require high-quality local information (e.g. Digital elevation model (DEM), land-use and soil characteristics maps) and local calibration (Hapuarachchi *et al.*, 2011). Several global flood and flash flood models with various levels of physics run operationally such as FLASH (http://flash.ou.edu), GloFAS, University of Maryland (flood.umd.edu; Wu *et al.*, 2014), University of Oklahoma (floods.global; Clark *et al.*, 2017), but run at a resolution that is often too coarse to capture urban flooding processes at the medium range. Alternatively, data-driven models (Thirumalaiah and Deo, 1998; Jain and Srinivasulu, 2004) require long-term data records not always available everywhere (Hapuarachchi *et al.*, 2011). Finally, simplified approaches (based on rainfall and soil moisture; Ntelekos *et al.*, 2006; Norbiato *et al.*, 2009; Javelle *et al.*, 2010) on runoff (Raynaud *et al.*, 2015), or process-based approaches (Panziera *et al.*, 2016; Antonetti *et al.*, 2019) have been shown to be as accurate as physically based models, particularly when transferred to ungauged river basins (Alfieri *et al.*, 2014). The present paper presents such a simplified methodology which accounts for meteorological and land-surface components together in a flood hazard risk forecast index (FHRFI), by combining information from the main flood-generating land-surface area (such as impervious surfaces, flood plains and wetlands footprints) within the meteorological hazard warning to create a spatial flood hazard index. It is applied to the surface flooding event over the

Houston metropolitan area following Hurricane Harvey during August 26–28, 2017, and benchmarked against a meteorological forecast index as a proof of concept to generate surface flood-risk warnings in a large conurbation, but the approach could be easily extended globally.

2 | METHODS

2.1 | Flood-generating land-surface data

The method developed and tested here is simple, easy to implement and scalable to continental or global scales. It combines precipitation forecasts from the NWP to hydrological land-surface factors that condition flooding. The analysis was performed on a 1 km resolution; therefore, data coarser than this, including the NWP forecast, were converted to a 1 km grid using a nearest-neighbour approach, or using rasterization if the data set was in vector format (GDAL/OGR Contributors, 2019). Flood-generating land-surface data were derived from three sources: (1) a 100 year return period floodplain inundation extent provided by the European Union's Joint Research Centre (https://data.jrc.ec.europa.eu/dataset/da4d7f64-a5c3-403f-bd2b-11a97176031e) to highlight low lying areas near rivers; (2) the 500 m buffered (to match the 1 km resolution of the analysis) primary road and interstate networks from the Texas Department of Transport (https://tnris.org/data-catalog/entry/txdot-roadways/); and (3) the urban areas of Texas from the Census Bureau's geographical database (https://www.census.gov/geo/maps-data/data/cbf/cbf_ua.html) over the study area. Each 1 km gridded data set was converted to a Boolean data type of 0s and 1s. For the floodplain inundation depth grid, values of 1 were assigned to all grid cells where 100 year return period flood inundation depth was > 0.1 m. In the other two data sets, values of 1 were assigned wherever an urban area or road network was present. The three grids were combined by assigning a value of 1 in grid cells where at least one of the three input data sets had a value of 1, otherwise of 0. The domain of the case study is focused around the Houston metropolitan area and up to the Texas/Louisiana border (lower left corner (27°42´ N−97°48´ W), upper right corner (31°12´ N−92°6´ W). This extent was chosen to ensure coverage of the NWS storm reports as well as the entire Houston metropolitan area.

2.2 | Observation data

Flood reports for August 26–28, 2017, were retrieved from the National Weather Service (NWS) storm reports (http://www.spc.noaa.gov/climo/online/). Most reports originate from trained spotters, local law enforcement and emergency management officials within the warned areas. Floods are reported at specific point locations but often refer to flooding within a broader area such as an entire neighbourhood or section of road network (Calianno *et al*., 2013). Biases may also exist in the observations due to unreported flooding in evacuated areas which penalizes forecasts for producing false alerts, or missing observations resulting in the forecast being over-rewarded for producing a correct negative. To compensate for these, an area of influence of 20 km radius was buffered to each reporting point; this value was chosen to represent the typical scale of the street or neighbourhood level of the reports, as well as being similar to the spatial resolution of the ECMWF ensemble NWP used here (approximately 18 km). This results in a hit being attributable to a forecast within a 20 km radius of the reported point. By applying a spatial buffering, double-reported flood events were also avoided, typically resulting from the same flood event being reported by two different public agencies. The buffered observation polygons were then converted, by nearest-neighbour rasterization, onto the same 1 km forecast data grid, with the verification procedure performed on the same grid network.

2.3 | Forecast data

Forecasts from the medium-range ENS of the ECMWF were used. Forecasts for one and three day total precipitation for August 26–29, 2017, were accessed from the ECMWF meteorological archival and retrieval system (MARS) for the 0000 and 1200 UTC forecasts for lead times from 0 to 156 hr (13 forecasts of various lead times). The forecasts were available at each approximately 18 km longitude–latitude regular grid point covering the study area. The first time of forecast used in this study was at 0000 UTC on August 20, and the latest at 0000 UTC on August 26. The forecasts were expressed as two indices: the EFI and the SoT, described below.

Alternatively, surface runoff forecasts could be used from a fully coupled land-surface model which accounts for land-surface conditions (e.g. the tiled ECMWF ccheme for surface exchanges over land incorporating land surface hydrology [HTESSEL] from the ECMWF ensemble). However, their generally coarse spatial resolution (approximately 18 km for the HTESSEL) and simplified representation of urban landscapes are unlikely to resolve appropriately the small-scale land-surface and topographical conditioning that occurs within surface flooding in urban environments.

2.4 | Extreme forecast index (EFI)

The EFI (Lalaurette, 2003; Zsótér, 2006) compares the cumulative probability distributions (CDFs) of the ensemble

forecast and of the corresponding model climate to identify how extreme the forecast is. It is defined by:

$$\text{EFI} = \frac{2}{\pi} \int_0^1 \frac{p - F_f(p)}{\sqrt{p(1-p)}} dp, \qquad (1)$$

where $F_f(p)$ is the proportion of ensemble members that lie below the p-th quantile of the model climate. EFI ranges between −1 and 1; an EFI close to 1/–1 shows that the forecast ensemble is shifted to extreme conditions with a large part of the ensemble being close to the model climate extremes.

Although a high EFI indicates that an extreme event is more likely than usual, the values cannot directly be converted to probabilities. By construction, the EFI is limited to 1 when the entire forecast distribution is outside the model climate range, regardless of how extreme such a forecast is.

2.5 | Shift of tails (SoT)

The SoT (Zsótér, 2006) complements the EFI by comparing specifically the tails of the ensemble forecast and model climate distributions. It measures the probability distance in the upper (SoT_+) and lower (SoT_-) tails of the forecast distribution as expressed by:

$$\text{SOT}_+(p) = -\frac{Q_f(p) - Q_c(0.99)}{Q_c(p) - Q_c(0.99)}$$

$$\text{SOT}_-(p) = -\frac{Q_f(p) - Q_c(0.01)}{Q_c(p) - Q_c(0.01)} \qquad (2)$$

where $Q_c(0.01)$ and $Q_c(0.99)$ are reference minimums and maximums of the model climatology; and $Q_c(p)$ and $Q_f(p)$ are the p-th quantile of the model climate and the forecast. Positive SoT_+ (SoT_-) index values indicate that the p-th quantile of the forecast distribution is more extreme than the reference maximum (minimum) of the model climate, hence the whole tail (from the p-th quantile) is also more extreme than the reference value. The SoT is, therefore, designed to highlight situations that might be unlikely but potentially very extreme. Here, the 99th percentile was used to highlight the locations with the greatest extremity, but future work could investigate the sensitivity of the verification scores to different SoT percentiles.

2.6 | Verification methods

The calculation of the FHRFI is made by overlapping the EFI (SoT) one/three day total precipitation forecasts with a
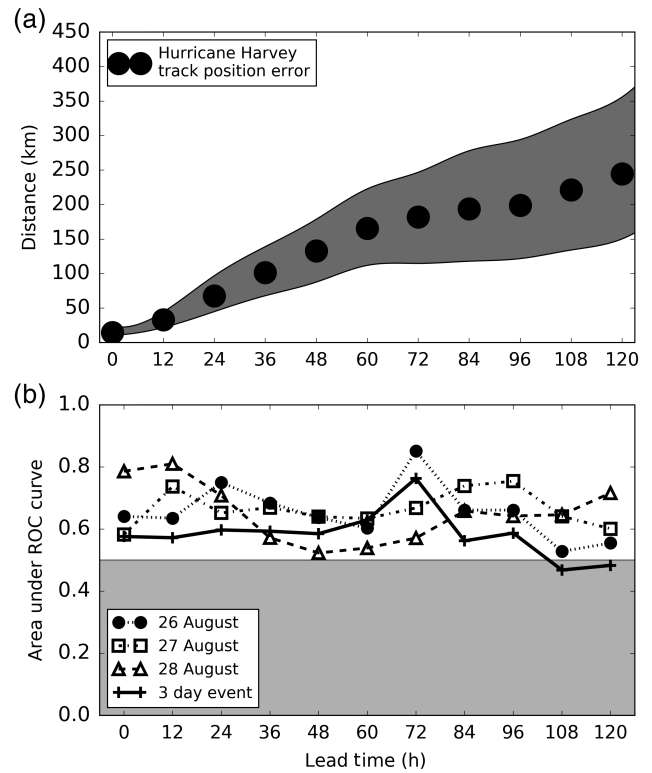

FIGURE 1 Meteorological forcing and skill over forecast lead time: (a) tropical cyclone track forecast errors for Hurricane Harvey for a three day time window, August 26–29, 2017 (seven time stamps combined, every 12 hr, from 0000 August 26 to 0000 August 29); and (b) area under the relative operative characteristics (AROC) curves for one day extreme forecast index (EFI) total precipitation calculated for August 26–28 (dashed lines) and three day EFI total precipitation (solid line) calculated for August 26–29, 2017. Lead times are given at the start of the forecast

local flood-generating land-surface map, resulting in the FHRFI$_\text{EFI}$ (FHRFI$_\text{SOT}$) indices. Verification scores were calculated for all forecast indices (EFI and SoT, meteorological forecasts only; and FHRFI$_\text{EFI}$ and FHRFI$_\text{SOT}$, which integrates land-surface information, calculated over one and three day accumulated precipitation) and lead times. The skill was evaluated from indices derived from a contingency table of forecasted and observed events.

- *Hit area* is defined as the intersection between the forecast area, the area over which EFI (SoT) forecast is greater than or equal to a given threshold EFI$_\text{T}$ (SoT$_\text{T}$), and the buffered observation area.
- *Miss area* is defined as the sum of all buffer observation areas not intersecting the forecasted area. If the 20 km buffer zone of an observed report intersects the area of the forecast, then the residual buffer zone, that is, the remainder of the buffered observation area which does not intersect the forecast area, is excluded from miss area
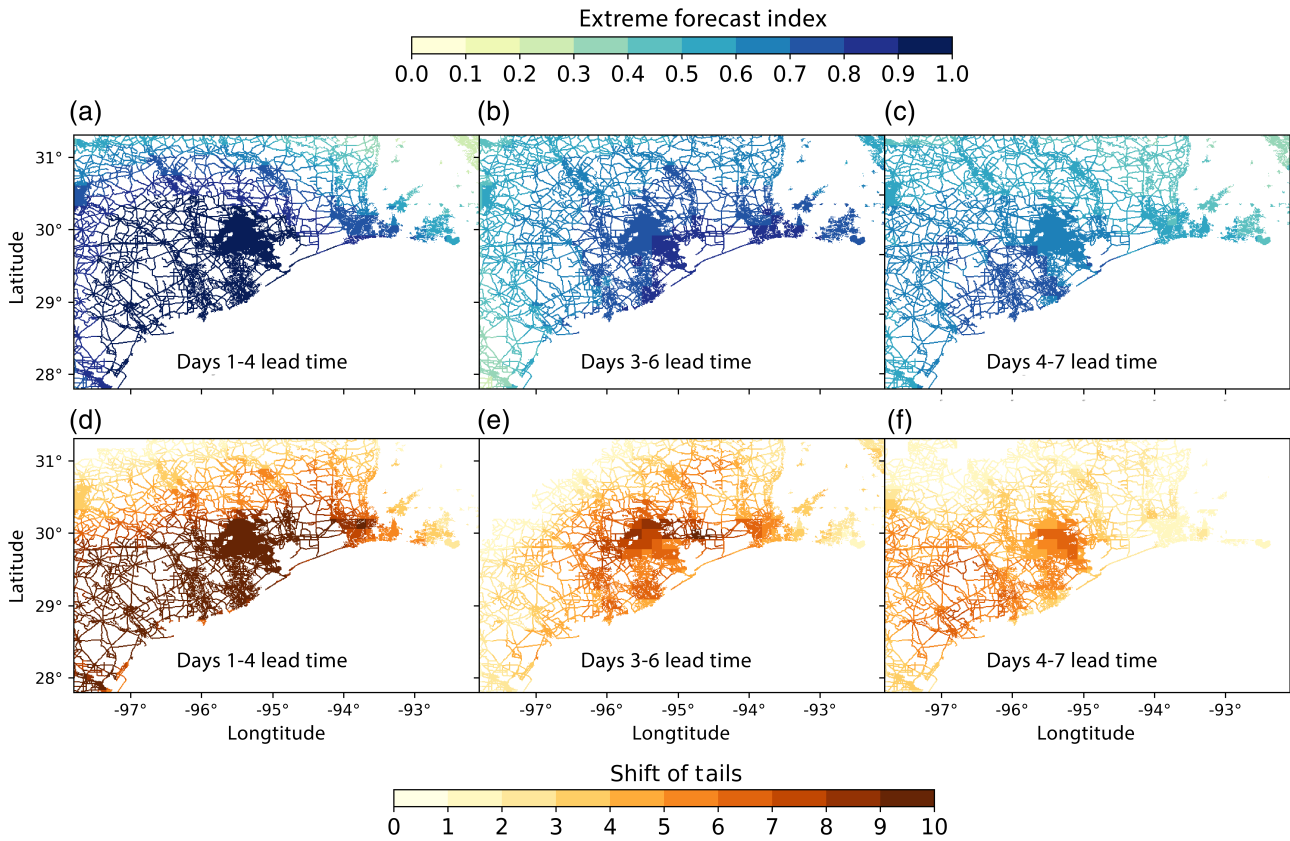
FIGURE 2 Time evolution of the flood hazard risk forecast index (FHRFI) for the Houston area of Texas. The FHRFI is based on three day total precipitation extreme forecast index (EFI) (a–c); and shift of tails (SoT) index (d–f) for the period August 26–29, 2017. Shaded areas highlight the local flood-generating land-surface information

or hit area counts. Only the buffer area that intersects the forecast area is accounted for.

- *False-alarm area* is defined as the sum of all forecasted areas not intersecting the buffered observation area.
- *Correct negatives area* is the study area minus all buffered observation areas and forecasted areas. This form of categorical verification can be prone to skill scores tending to zero when the base rate for an event occurrence is small (Ebert *et al.*, 2013). This issue was mitigated here by focusing upon a single-event case study, with 153 observations being available within a three day period.

Contingency tables were computed using the EFI_T ranging from 0 to 1 with 0.02 steps, and SoT_T ranging from 0 to 10 with 0.2 steps.

Probability of detection (or hit ratio—HR) is defined as:

$$HR = \frac{hit_{area}}{hit_{area} + miss_{area}} \quad (3)$$

Probability of false forecast (false-alarm ratio—FAR) is defined as:

$$FAR = \frac{false\ alarm_{area}}{false\ alarm_{area} + hit_{area}}. \quad (4)$$

Relative operating characteristic (ROC) curves, calculated for EFI thresholds ranging from 0 to 1 (for positive EFIs) and for SoT thresholds ranging from 0 to 10 (for positive SoTs, while the maximum for the case study is around 10), and associated area under the ROC curve or ROC score. The ROC score ranges from 0 to 1, with forecast being skilful if > 0.5.

The EFI_T and SoT_T values used as two indicators of urban flood forecasting occurrence in this study were calculated using critical success index (CSI) or threat score (TS), defined as:

$$CSI = \frac{hit_{area}}{hit_{area} + miss_{area} + false\ alarm_{area}}. \quad (5)$$

The CSI is useful when the event to be forecast occurs less frequently than the non-occurrence (Wilks, 2011), and ranges from 0 to 1 where the worst possible CSI is 0, and the best possible CSI is 1.

The two-sample Kolmogorov–Smirnov test was used to calculate the significance of the difference between two
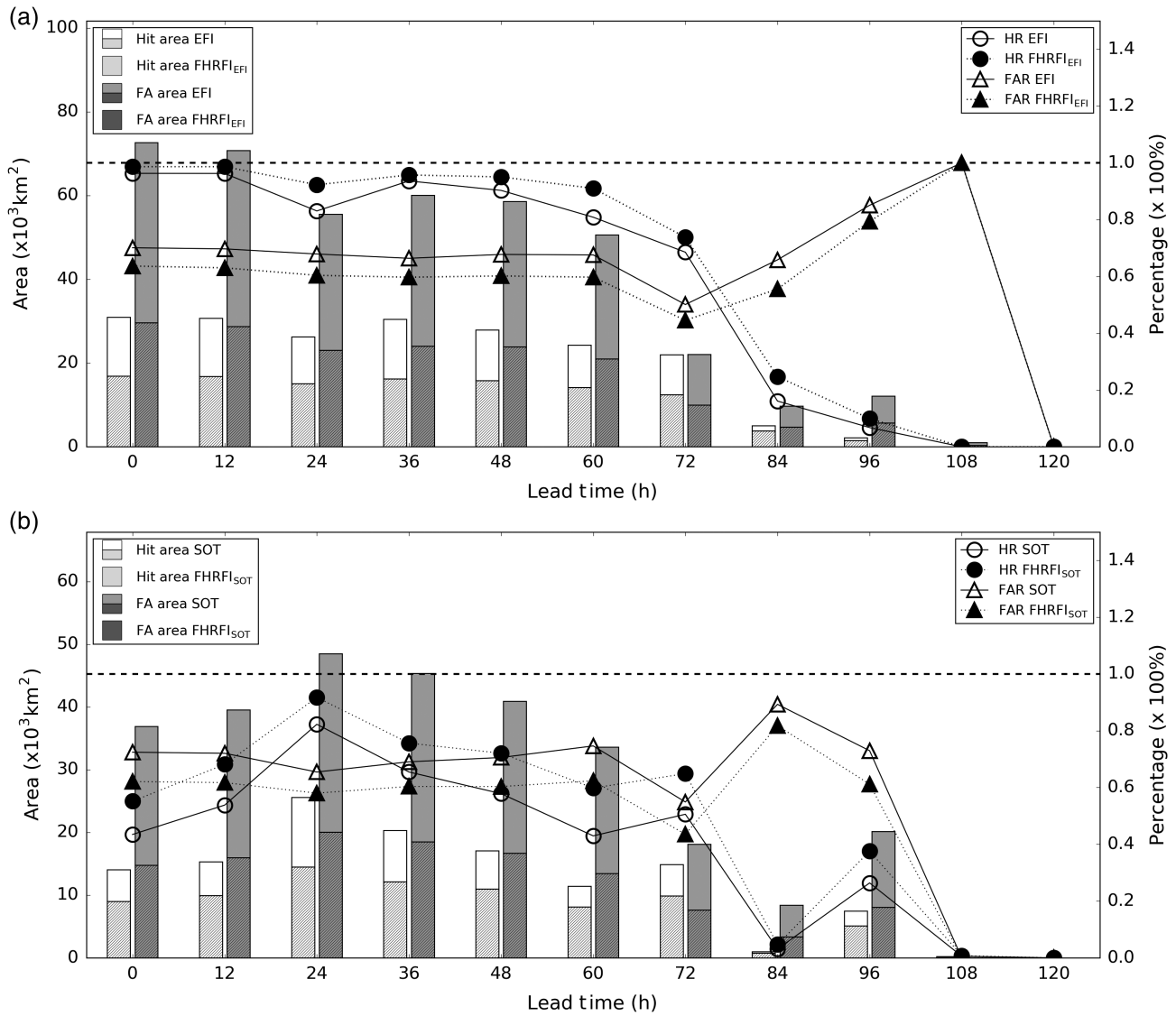
FIGURE 3 Model skill for Hurricane Harvey case. Hit ratio (HR) and false-alarm ratio (FAR) (lines and symbols) and hit and false-alarm areas (bars) for $FHRFI_{EFI}$, the $FHRFI_{SOT}$, extreme forecast index (EFI) and shift of tails (SoT) forecasts: (a) $EFI_{threshold} = 0.7$; and (b) $SoT_{threshold} = 5.0$

distributions, the $FHRFI_{EFI\ (SOT)}$ and EFI (SoT). For the $FHRFI_{EFI\ (SOT)}$ and EFI (SoT), the CSI and FAR distributions were calculated, then compared.

3 | RESULTS

The flood event of August 26–28, 2017, over the urban area of Houston is a typical example of a devastating urban flooding hazard, with at least 70 confirmed deaths and estimated losses over US$150 billion (Emanuel, 2017), making it one of the costliest natural disaster ever in the United States (Golnaraghi *et al.*, 2014). Its disastrous impact was due to extreme precipitation resulting from a quasi-stationary system over the region (Emanuel, 2017). As Houston is one of most rapidly urbanizing areas in the United States, the

damage was immense, exemplifying the potential risks facing large urban areas across the world.

In a first step, the ENS forecast products EFI and SoT from the ECMWF were used alone. As the event lasted three days, indices were tested based on precipitation totals of one and three days to investigate whether indices based on a longer accumulation degraded the performance of the warning. Performance was measured based on a contingency table comparing issued warnings (when the forecast index exceeded a given threshold) and observed floods for the same date. Observed floods were identified from several flood reports, with a buffering area of 20 km around each report to allow for uncertainty in the associated location information. The area under the relative operating characteristic (AROC) suggests skilful predictions based on the EFI
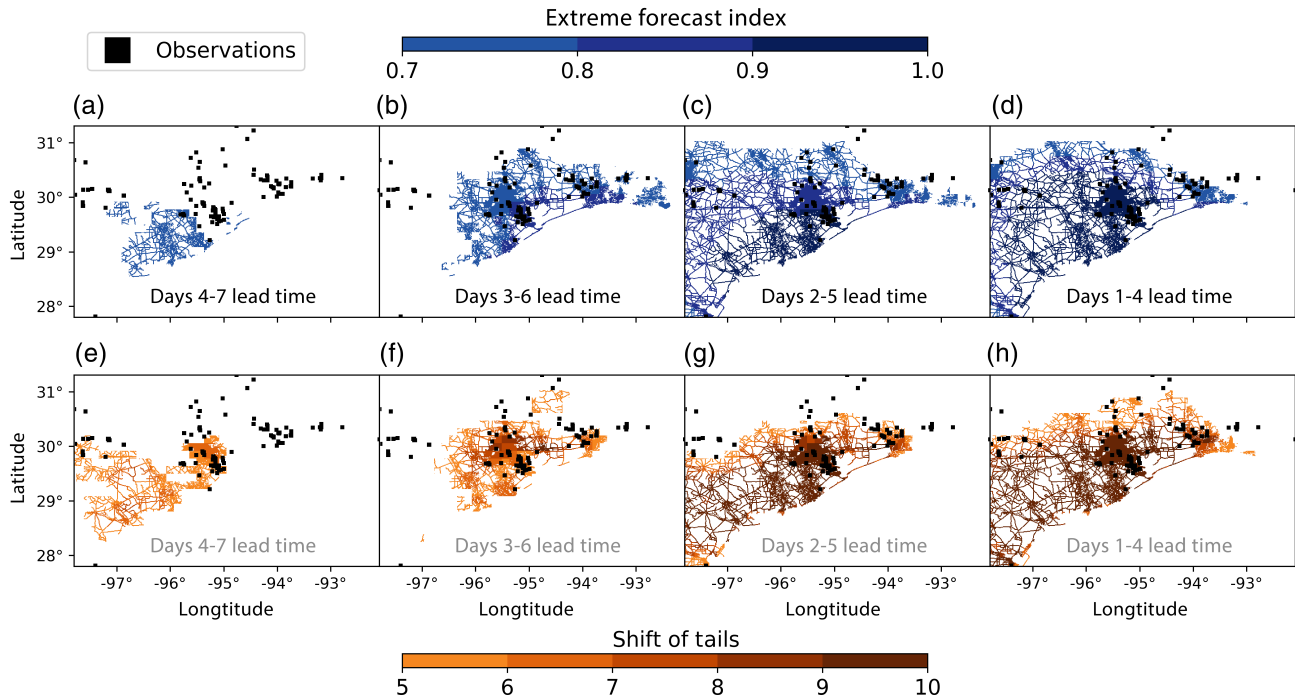
FIGURE 4 Evolution of the flood hazard risk forecast index (FHRFI) above the extreme forecast index (EFI) (shift of tails—SoT) thresholds for the Houston area of Texas for different forecast lead time. The FHRFI is based on the three day total precipitation EFI (a–d); and the SoT index (e–h) for the period August 26–29, 2017

up to 96 hr ahead (Figure 1b) (an AROC consistently above the 0.5 line is representative of a climatological forecast). Whilst skill varies with lead time and date of the forecasts, the AROC remains between 0.6 and 0.8 for all four tested indices; generally, skill is higher for the one than for the three day forecasts up to a 36 hr lead time, when the August 28 one day forecast deteriorated. Similar patterns were found for forecasts based on the SoT (data not shown). For simplicity, the rest of the analysis is reported for the three day indices as integrators of the whole event.

The track of the tropical cyclone linked with Hurricane Harvey was well forecasted by the ENS, with a track error of around 300 km six days before the event and of only 200 km four days ahead (Figure 1a). As a result, total precipitation over the area was also well forecasted by the ENS.

In a second step, the FHRFI was calculated by overlapping the EFI (SoT) three day total precipitation forecasts with a local flood-generating land-surface map, resulting in the FHRFI$_{EFI}$ (FHRFI$_{SOT}$) indices (Figure 2). In the present case study, floodplain inundation-prone areas, primary roads, interstate network and urban areas were used as proxy information for areas susceptible to surface water flooding, effectively halving the forecast domain from 128,600 to 50,800 km^2 (Figure 2).

Compared with the benchmark EFI and SoT indices, the FHRFI shows comparable or higher skills (by around 5%), with a maximum AROC of 0.756 for the FHRFI$_{EFI}$ and of

0.760 for the FHRFI$_{SOT}$ for a 72 h lead time (data not shown). For each index, warning thresholds were identified using the CSI (Wilks, 2011) (or TS, CSI), which measures the proportion of correct forecasts issued ignoring correct negative forecasts. The FHRFI reached a CSI of 46% (43%) for an EFI = 0.7 (SoT = 5.0), whilst an EFI (SoT) alone achieved a CSI of 40% (31%). The distributions of the FHRFI$_{EFI}$ $_{(SOT)}$ and EFI (SoT) were statistically significantly different ($p < .0001$), hence the introduction of land-surface information in the FHRFI yields significant improvements in the forecast of surface flooding. Explicit decomposition of the forecast skill by HR and FAR shows that the increase of performance in the combined index is associated with an increase in the HR by 5–10% and a decrease in the FAR by 10% over different lead times for the EFI-based index (Figure 3), and an HR increase by up to 15% and a FAR decrease by 10–15% for an SoT-based index. In addition to higher overall skills, the FHRFI enables the refinement of warning areas, highlighting only those locations most at risk of flooding instead of issuing warnings over the whole region. This is shown in Figure 3, where an false alarm (FA) area drops when using the FHRFI. This can be critically important for emergency responders to target their efforts better, either by suggesting evacuation routes or by deploying assistance in targeted areas, specifically by issuing flood warnings according to the EFI/SoT used in the FHRFI forecasts.

4 | DISCUSSION

The results applied to the case study of flooding over the Houston urban area following the tropical cyclone Harvey have highlighted the benefits of using flood-generating land-surface information in a flood forecast index. The paper discusses some of the limitations associated with the data sources, verification strategies and application of the method which would need to be considered for a more global application.

4.1 | Verification data

One challenge when verifying predictions of surface water flooding (especially in an urban environment) is selecting a suitable observation data set. A total of 153 flood reports from the local NWS offices were used. The NWS offices collect and verify reports of flooded locations from sources including the media, public and emergency services. Alternative sources were considered but had to be discarded: the flood extent from observed satellite was not appropriate owing to cloud obscuration from the hurricane system and interference of the microwave signal by buildings; aerial imagery was not available to the authors; river gauge measurements (e.g. the severe hazards analysis and verification experiment [SHAVE] database; Gourley *et al.*, 2013) could not capture flooding away from river channels over urban surfaces; and short-range forecasting from the NWS, which could be treated as proxy observations of flooded areas, was not available to the authors at the time of the study.

4.2 | Numerical weather prediction (NWP) forecasts

Raw total precipitation forecasts from global NWP forecasts generally underestimate extreme rainfall totals at the local scale (Lavers and Villarini, 2013; Pillosu and Hewson, 2017), with locally derived exceedance thresholds criteria unlikely to be met. Warnings based on NWP reforecast climatologies, such as that used in the European runoff index based on climatology flash-flood indicator within the EFAS (Raynaud *et al.*, 2015), can be used to tackle the underestimation problem, but typically thresholds are derived from a single control reforecast member, hence ignoring uncertainty in the simulations. In contrast, the EFI and SoT methods used in the study explicitly use an ensemble reforecast series (in the case of the ENS, a 10 member 20 year reforecast) to produce a climatological reforecast distribution that can be compared against the equivalent ensemble forecast distribution. When the forecast distribution is located entirely beyond the distribution of the climatology (i.e. when EFI = 1.0), the complementary index of the SoT measures how far beyond the n-th climatological percentile the forecast distribution is shifted. The EFI and SoT thresholds can be used by forecasters during a storm to identify areas at risk of urban flooding, that is, EFI ≥ threshold. For this, the CSI was used to identify an optimum EFI/SoT in the FHRFI that optimizes the HR/FAR. This threshold is likely be region dependant as it depends on climatology and event extremity.

4.3 | Flood-generating land-surface information

By using the flood-generating land-surface data to create the FHRFI, a simple method to reduce the incidence of false-alarms is created, hence allowing the better identification of areas of greatest interest to emergency responders and civil protection agencies. If combined with additional exposure information, for example, hospitals, nursing homes and critical infrastructure such as electricity substations, it could further provide high-impact forecasts to emergency responders, for example, based on the rapid risk-mapping concept (Dottori *et al.*, 2017). Figure 4 shows a flow chart of the FHRFI area for different forecast lead times. The EFI and SoT start to predict the affected area very well from days 3–6 lead time, when the majority of urban flood reports are covered with FHRFI$_{EFI(SOT)}$. For the days 2–5 lead time, almost all reports are inside the FHRFI$_{EFI}$, while it is not the case for days 4–7 lead time.

Other approaches, such as a two-dimensional hydraulic flood-inundation models, driven off-line from the NWP forecasts, could also be used to account for these small-scale features, but such systems require intense computational expense, which cannot be easily scaled to cover large domains. It could be possible simply to use the flood-generating land-surface information as a static predictor of urban flooding hazards, that is, not including the dynamic NWP forecast information. However, this might further increase the number of false-alarms as it would not specify which particular areas were at risk. To test this, the CSI and FAR were calculated when the entire flood-generating land surface was used as the warning area. This achieved a CSI of 35%, which is 11% (8%) less than FHRFI$_{EFI}$ (FHRFI$_{SOT}$) at a three day lead time, and a FAR of 65%, which is 20% (21%) worse than FHRFI$_{EFI}$ (FHRFI$_{SOT}$) also for a three day lead time. This demonstrate the value of the additional information brought by the NWP forecasts.

4.4 | Additional case study

As a second proof of concept, another case study was also conducted over the same Houston urban area. The event that occurred on July 4, 2018, lasted only one day and was less extreme than Hurricane Harvey. Results showed that the

FHRFI had a CSI of 13% for EFI = 0.5 (SoT = 1.0), whilst for the EFI (SoT) alone the CSI was 6% (10%). The increase in the performance of the FHRFI is associated with a decrease in the FAR by 3–6% over different lead times for the EFI-based index (Supporting Information Figure S1), and an FAR decrease up to 6% for the SoT-based index (see Supporting Information Figure S2 for the time evolution of the FHRFI). The HR remains the same in both cases.

5 | CONCLUSIONS

The present paper presents a flood hazard risk index specifically designed to generate surface flooding forecast information from the numerical weather prediction (NWP) for total precipitation forecasts. Applied to the flood event of August 26–28, 2017, over the Houston urban area of Texas, following Hurricane Harvey, it shows that (a) surface water flooding forecasts based on the global NWP can be skilful if precipitation extremes are well forecasted; and (b) the false alarm ratio (FAR) is significantly improved when flood-generation land-surface information is accounted for in the forecast. The method, demonstrated for a case study as a proof of concept, can be easily deployed at the continental or global scales, owing for the relevant land-surface information being available, but would require verification at a global scale and over a longer period of time. Despite improvements in hit ratio (HR) and FAR, the false-alarm area remains almost twice the hit area, high false-alarm rates being a known common challenge in flash-flood forecasting. However, the present research demonstrated that by including relevant land-surface information within the forecast products, one can proportionally reduce the false-alarm area (for the factor approximately 2.4) around 40% more than the hit area (factor approximately 1.7), hence increasing the forecast usefulness.

ORCID

Toni Jurlina https://orcid.org/0000-0002-7391-7341

REFERENCES

Alfieri, L., Berenguer, M., Knechtl, V., Liechti, K., Sempere-Torres, D. and Zappa, M. (2016) Flash flood forecasting based on rainfall thresholds. In: Duan, Q., Pappenberger, F., Thielen, J., Wood, A., Cloke, H.L. and Schaake, J.C. (Eds.) *Handbook of Hydrometeorological Ensemble Forecasting*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 1–38.

Alfieri, L., Pappenberger, F. and Wetterhall, F. (2014) *Natural Hazards and Earth System Sciences*, 14, 1505.

Alfieri, L., Salamon, P., Pappenberger, F., Wetterhall, F. and Thielen, J. (2012) *Environmental Science & Policy*, 21, 35–49.

Antonetti, M., Horat, C., Sideris, I.V. and Zappa, M. (2019) *Natural Hazards and Earth System Sciences*, 19, 19–40.

Calianno, M., Ruin, I. and Gourley, J.J. (2013) *Journal of Hydrology*, 477, 1–16.

Clark III, R.A., Flamig, Z.L., Vergara, H., Hong, Y., Gourley, J.J., Mandl, D.J., Frye, S., Handy, M. and Patterson, M. (2017) *Bulletin of the American Meteorological Society*, 98, 1697–1715.

Cole, S., Moore, R., Bell, V. and Jones, D. (2006) Issues in flood forecasting: ungauged basins, extreme floods and uncertainty. In: *Proceedings of the Frontiers in Flood Forecasting, 8th Kovacs Colloquium*. Paris: UNESCO.

Doswell, C.A., III, Brooks, H.E. and Maddox, R.A. (1996) *Weather and Forecasting*, 11, 560–581.

Dottori, F., Kalas, M., Salamon, P., Bianchi, A., Alfieri, L. and Feyen, L. (2017) *Natural Hazards and Earth System Sciences*, 17, 1111.

Ebert, E., Wilson, L., Weigel, A., Mittermaier, M., Nurmi, P., Gill, P., Göber, M., Joslyn, S., Brown, B. and Fowler, T. (2013) *Meteorological Applications*, 20, 130–139.

Emanuel, K. 2017. *Proceedings of the National Academy of Sciences: 201716222.*

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. (2016) *Wiley Interdisciplinary Reviews: Water*, 3, 391–418.

GDAL/OGR Contributors (2019). GDAL/OGR Geospatial Data Abstraction software Library. Open Source Geospatial Foundation. Available at: https://gdal.org [Accessed 18 October 2015].

Georgakakos, K., Graham, R., Jubach, R., Modrick, T., Shamir, E., Spencer, C. and Sperfslage, J. (2013) *HRC Technical Report*, 9.

Golnaraghi, M., Etienne, C., Guha-Sapir, D. and Below, R. (2014) *Atlas of Mortality and Economic Losses from Weather, Climate, and Water Extremes (1970–2012)*. World Meteorological Organization.

Gourley, J.J., Erlingis, J.M., Hong, Y. and Wells, E.B. (2012) *Weather and Forecasting*, 27, 158–173.

Gourley, J.J., Flamig, Z.L., Vergara, H., Kirstetter, P.-E., Clark III, R.A., Argyle, E., Arthur, A., Martinaitis, S., Terti, G., Erlingis, J.M., Hong, Y. and Howard, K.W. (2017) *Bulletin of the American Meteorological Society*, 98, 361–372.

Gourley, J.J., Hong, Y., Flamig, Z.L., Arthur, A., Clark, R., Calianno, M., Ruin, I., Ortel, T., Wieczorek, M.E. and Kirstetter, P.-E. (2013) *Bulletin of the American Meteorological Society*, 94, 799–805.

Hapuarachchi, H.A.P., Wang, Q.J. and Pagano, T.C. (2011) *Hydrological Processes*, 25, 2771–2784.

Hapuarachchi, H. P. and Wang, Q. 2008. *A review of methods and systems available for flash flood forecasting*, Report for the Bureau of Meteorology, Australia.

Hewson, T. and Tsonevsky, I. (2016) *ECMWF Newsletter*, 150, 2–3.

Hurford, P., Priest, S.J., Parker, D.J. and Lumbroso, D.M. (2012) *International Journal of Climatology*, 32, 1768–1774.

Jain, A. and Srinivasulu, S. (2004) *Water Resources Research*, 40.

Javelle, P., Fouchier, C., Arnaud, P. and Lavabre, J. (2010) *Journal of Hydrology*, 394, 267–274.

Kraas, F., Aggarwal, S., Coy, M. and Mertins, G. (2013) *Megacities: Our Global Urban Future*. Springer Science & Business Media.

Lalaurette, F. (2003) *Quarterly Journal of the Royal Meteorological Society*, 129, 3037–3057.

Lavers, D.A. and Villarini, G. (2013) *Geophysical Research Letters*, 40, 6405–6410.

Norbiato, D., Borga, M. and Dinale, R. (2009) *Meteorological Applications*, 16, 65–75.

Ntelekos, A.A., Georgakakos, K.P. and Krajewski, W.F. (2006) *Journal of Hydrometeorology*, 7, 896–915.

Panziera, L., Gabella, M., Zanini, S., Hering, A., Germann, U. and Berne, A. (2016) *Hydrology and Earth System Sciences*, 20, 2317–2332.

Penna, D., Borga, M. and Zoccatelli, D. 2013.

Pillosu, F. and Hewson, T. (2017) *ECMWF Newsletter*, 153, 2–3.

Ramachandra, T. and Kumar, U. (2008) *Electronic Green Journal*, 1.

Raynaud, D., Thielen, J., Salamon, P., Burek, P., Anquetin, S. and Alfieri, L. (2015) *Meteorological Applications*, 22, 410–418.

Thielen, J., Bartholmes, J., Ramos, M.H. and de Roo, A. (2009) *Hydrology and Earth System Sciences*, 13, 125–140.

Thirumalaiah, K. and Deo, M. (1998) *Journal of Hydrologic Engineering*, 3, 26–32.

Wilks, D.S. (2011) *Statistical Methods in the Atmospheric Sciences*. Academic Press.

Wu, H., Adler, R.F., Tian, Y., Huffman, G.J., Li, H. and Wang, J. (2014) *Water Resources Research*, 50, 2693–2717.

Zsótér, E. (2006) *ECMWF Newsletter*, 107, 8–17.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.