

This is the accepted manuscript version of the contribution published as:

Tripathy, S.S., **Hari, V.**, Karmakar, S., Ghosh, S. (2020):

Flood risk forecasting at weather to medium range incorporating weather model, topography, socio-economic information and land use exposure

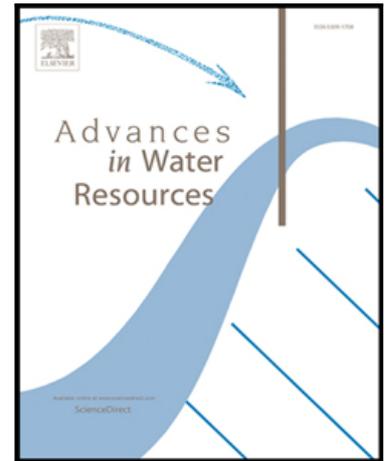
Adv. Water Resour. **146** , art. 103785

The publisher's version is available at:

<http://dx.doi.org/10.1016/j.adwatres.2020.103785>

Journal Pre-proof

Flood Risk Forecasting at weather to medium range incorporating Weather Model, Topography, Socio-economic Information and Land Use Exposure



Shrabani S. Tripathy , Vittal Hari , Subhankar Karmakar ,
Subimal Ghosh

PII: S0309-1708(20)30416-4
DOI: <https://doi.org/10.1016/j.advwatres.2020.103785>
Reference: ADWR 103785

To appear in: *Advances in Water Resources*

Received date: 29 April 2020
Revised date: 24 September 2020
Accepted date: 5 October 2020

Please cite this article as: Shrabani S. Tripathy , Vittal Hari , Subhankar Karmakar , Subimal Ghosh , Flood Risk Forecasting at weather to medium range incorporating Weather Model, Topography, Socio-economic Information and Land Use Exposure, *Advances in Water Resources* (2020), doi: <https://doi.org/10.1016/j.advwatres.2020.103785>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier Ltd.

Highlights

- Inadequacy of weather forecasts to identify the flood hotspots
- Development of a new framework for weather scale flood risk forecasting
- New definition and forecasts of Hazard at a weather scale using the hindcasts
- Incorporation of socio-economic, geomorphologic and exposure information in flood forecasting
- Forecasting and identification of high risk zones using the developed methodology

Journal Pre-proof

Flood Risk Forecasting at weather to medium range incorporating Weather Model, Topography, Socio- economic Information and Land Use Exposure

Shrabani S. Tripathy¹, Vittal Hari^{2,3}, Subhankar Karmakar^{1,2,4}, Subimal
Ghosh^{1,4,5,*}

¹Inter-Disciplinary Program in Climate Studies, Indian Institute of Technology Bombay,
Mumbai – 400076, India

²Environmental Science and Engineering Department, Indian Institute of Technology
Bombay, Mumbai – 400076, India

³UFZ–Helmholtz Centre for Environmental Research, Leipzig D-04318, Germany

⁴Centre for Urban Science and Engineering, Indian Institute of Technology Bombay, Mumbai
– 400076, India

⁵Department of Civil Engineering, Indian Institute of Technology Bombay, Mumbai –
400076, India

*Correspondence to Subimal Ghosh, Email: subimal@civil.iitb.ac.in

Keywords:

Extreme rainfall, Flood risk, Weather forecast, Hazard, Vulnerability, Exposure

Abstract

Non-structural mitigation measures to the globally increasing flood events include forecast based alert generation. However, the extreme rainfall forecasts are associated with low hit rate, high false alarm, and spatiotemporal bias; which makes it difficult to rely on them. Further, the losses due to flood in a region not only depend on rainfall severity but also on topography, socioeconomic conditions and exposure of the region to floods. Here, we introduce a new concept of spatial flood risk mapping and forecasting at weather to medium range based on forecasted hazard, embedded with vulnerability (topographic and socioeconomic) and exposure information. Here, we define hazard as the probability of extreme rainfall event during upcoming days given an available weather forecast for the same days. As hindcast is used for computation of probabilities, hazard contains prior information about the false alarm, hit rate and spatiotemporal bias of the forecast. Vulnerability is calculated by averaging the topographic and socioeconomic indicators, and exposure is calculated using a land use land cover map. Topographic vulnerability is computed with digital elevation model using Height Above the Nearest Drainage method, whereas Data Envelopment Analysis is performed to derive the socioeconomic vulnerability based on the demographic census data. For a specific region and a specific event, the relative flood risk maps are generated at an administrative level (e.g., district, subdistrict or village level for India) and the high-risk areas can be identified from those maps for mitigation. The methodology is demonstrated for a very recent extremely severe flood event that happened in Kerala, India in August, 2018. It is evident from the results that the high-risk areas forecasted well in advance (as high lead time as 15 days) match fairly well with the areas, which suffered maximum losses because of direct flood.

1. Introduction

Climate change has caused a considerable impact to the global water cycle which lead to changes in seasonal patterns as well as an increase in the frequency of extreme rainfall events [Oki and Kanae, 2006]. Intergovernmental Panel on Climate Change [IPCC, 2012] reports that this increase in extreme rainfall is statistically significant in many parts of the globe. It also states with medium confidence, that these changes in extremes are attributed to anthropogenic influences. In the Indian sub-continent as well, both empirical methods and model projections have shown an increase in the frequency and magnitude of extreme rainfall events [Goswami et al., 2006; Preethi et al., 2017; Roxy et al., 2017]. A decrease in the total rainfall and intensification of extremes in the tropics for 21st century has been projected in the IPCC reports [Seneviratne et al., 2012]. Irrespective of climate change, the importance of forecast and understanding its uncertainty have a very high societal relevance. Increasing extremes in a changing climate further increases its importance. Extreme rainfall events not only affect our day to day lives but also result in floods and landslides which cause huge loss of lives and property [Dottori et al., 2018; Fowler et al., 2010]. Every year a billion people are affected and thousands die because of these extreme events [IFRC/RCS, 2011].

According to the National Disaster Management Authority (NDMA, 2008), floods have become a cause of concern because of an increasing trend in flood related losses in India. Some examples of such extreme events include the heavy rainfall event in Mumbai, India on 26th July 2005, which recorded 944 mm rainfall in 24 hours [Jenamani et al., 2006]. The rainfall and resulting flood caused the death of almost 1000 people and an economic loss of about US\$100 million [Kumar et al., 2008]. Uttarakhand in India received almost 340 mm rainfall in a day (375% above daily mean) on 17th June 2013, which resulted in severe flash floods [Dube et al., 2014]. Chennai city experienced a terrible flood during November-

December 2015, which caused at least 400 deaths, economic loss of US\$ 1120 million [Narasimhan et al., 2016; Seenirajan et al., 2017].

A significant amount of research has been conducted in the past decades to understand these events. A large proportion of these events could not have been predicted accurately and hence resulted in devastation [Coumou and Rahmstorf, 2012]. The commonly used alert generation and warning system for flood are mostly based on the rainfall forecast. Hence, an accurate prediction of extreme rainfall at an administrative level is very important for the stakeholders and decision-makers. However, the complex multiscale atmospheric processes responsible for the occurrence of any extreme rainfall event and their inherent variability makes them difficult to predict [Fritsch et al., 1998]. Coarse resolution dynamical models often fail to predict these extreme rainfall events with accuracy as these have high false alarm, low hit rate and spatio-temporal biases [Březková et al., 2010; Khaladkar et al., 2007; Selvam, 2011; Shastri et al., 2017]. Prediction skills can be improved using regional models at a high spatiotemporal resolution, but these simulations are computationally very intensive thus are difficult to perform on real-time [Dodla and Ratna, 2010].

Very heavy rainfall in a short time span often results in floods when it exceeds the ground absorption capacity and the runoff exceeds the capacity of river system [Neuendorf et al., 2005]. This makes mitigation planning difficult. Flood risk associated with these events are difficult to predict as it not only depends on the complexity of processes related to extreme rainfall but also on the interaction between these events and the geography, population, infrastructure, the preparedness of the region [Balk et al., 2012].

Flood risk can be defined as a product of hazard, vulnerability and exposure [IPCC, 2012; Kron, 2005; Karmakar et al., 2010], which is used for climatological projections. In the risk framework for climate applications, hazard is defined in IPCC (2012) as "The potential

occurrence of a natural or human induced physical event that may cause loss of life, injury, or other health impacts, as well as damage and loss to property, infrastructure, livelihoods, service provision, and environmental resources". Calculating flood risk also requires vulnerability, which works as a proxy of human-environment relationship [Turner et al., 2003]. IPCC (2012) defines vulnerability as the "propensity or predisposition to be adversely affected". In order to get the overall vulnerability of a region, various classes of vulnerability are combined together [Karmakar et al., 2010]. The exposure [IPCC, 2012] component of risk is defined as "the presence (location) of people, livelihoods, environmental services and resources, infrastructure, or economic, social, or cultural assets in places that could be adversely affected by physical events and which, thereby, are subject to potential future harm, loss, or damage". In the changing climate, along with an increase in the amount and frequency of extreme rainfall events, the exposure of humans to flood is also increasing [Hirabayashi et al., 2013], thus is considered for flood risk quantification. However, this approach of defining climatological flood risk cannot be used for event range flood mitigation.

Here, we propose a new methodology to use this concept of flood risk at weather to medium range scale to generate event specific risk maps. Traditionally the concept of hazard is used at a climate scale, where it is defined as the probability of extreme rainfall above a specific threshold and computed from the long-term data. Here we have introduced a concept of hazard, which is computed at weather to medium range in the forecast system. Hazard is computed as the probability of occurrence of an extreme rainfall in any grid, given a forecast value. Flood risk maps are to be generated by combining hazard with socioeconomic and topographic vulnerability and, exposure of a region. To demonstrate and evaluate this methodology, it is applied for the flood event in Kerala, India, which occurred during August 2018. The following section discusses the case study area and the data used. The limitations of the state of art weather forecast system is discussed in the section 3. Sections 4 and 5 contains the methodology and results respectively and the paper is summarised in the last section.

2. Case study and Data

2.1. Case study description: Kerala flood of August, 2018

Kerala is a southern coastal state of India, spread over an area of 38,863 km² and is divided into 14 districts and 63 subdistricts (Figure 1a) [Apel et al., 2009]. It is one of the most densely populated Indian states with a population of over 33 million (860 people per square kilometre) and with a gross domestic product (GDP) of US\$120 billion (Department of Economics and Statistics, Government of Kerala, <http://www.ecostat.kerala.gov.in/index.php/economy>). Many other human development indicators for Kerala are at par with those of developed countries like literacy rate 93.11%, life expectancy 77 years and a sex ratio of 1,084 women per 1,000 men [Census, 2011]. It should also be noted that the state Kerala is a coastal state extending like a thin strip along the

west coast of India and the grids of state of the art weather forecasting model, such as the Global Ensemble Forecast System (GEFS) Reforecast Version 2, is too coarse (spatial resolution of $1^\circ \times 1^\circ$ for the present one, which is used in the study) to obtain district and state specific information (Figure 1b).

Between June 1 and August 18, 2018, the cumulative rainfall in Kerala was 42% in excess of the normal average which caused the worst flood (in August) in the state since 1924. During this period, maximum rainfall occurred on 15th August and the daily accumulated rainfall is shown in Figure 2a. Coastal parts of Kerala received extremely heavy rainfall during the event. From figure 2a, it is evident that Thrissur, Malappuram, Alappuzha, Pattanamtitta and Kollam districts received the rainfall amount more than 120 mm/day, whereas in the other parts of Kerala the intensity was more than 60 mm/day. 1,259 out of 1,664 villages spread across all 14 districts (Supplementary table S1) were affected [KPDNA, 2018]. Nearly 341 landslides were reported from 10 districts with Idukki being the worst hit district with 143 landslides. The seven most affected districts were Alappuzha, Ernakulam, Idukki, Kottayam, Pattanamtitta, Thrissur, and Wayanad. Malappuram and Palakkad suffered moderate losses in this flood. This affected 5.4 million people, displaced 1.4 million people, and took 433 lives (22 May–29 August 2018) [KPDNA, 2018]. The district wise total crop, transport, aquaculture, health and heritage losses (in million USD), as obtained from the KPDNA, 2018 reports are presented in Figure 2b. It is interesting to note that there is lack of consistency between the districts with high rainfall and the districts with high loss. It may also be possible that the heavy rainfall that occurred at different places during the 1st half of August 2018 is responsible for such an inconsistency. To explicitly show this, we have shaded the districts with high loss in Figure 2a. This is quite evident that very high loss cannot be attributed to the rainfall intensity only and there exists lot of other factors, related to vulnerability and

exposure. Here, in this study we focus on the same for the forecasts of weather to medium range flood risk map.

According to the KPDNA (2018) report, though heavy rainfall was forecasted well in advance, a lack of planning caused the overflow of reservoirs. The late pre-emptying of reservoirs is often making flood damage worse, in India and elsewhere. These problems could have been avoided by the help of a reliable forecast system and proper mitigation plan. The method proposed here is applied to the Kerala flood of August 2018 for demonstration.

2.2. Data

The data obtained for different modules of the weather to medium range flood risk forecast models are mentioned in the following subsections.

2.2.1. Hazard

The 6 hourly rainfall data at a spatial resolution of $1^\circ \times 1^\circ$ for a period of 1985 to 2015 from Global Ensemble Forecast System (GEFS) Reforecast Version 2, is used as the reforecasted rainfall dataset [Hamill et al., 2013]. Reforecast (also called hindcast) data are retrospective forecasts for the past using the same model configuration employed for operational forecasts. The datasets are generated from an 11-member ensemble forecast, every day from 00 UTC for a period of December 1984 to present day; and is available at a 3-hourly time step for first 8 days (horizontal resolution is T254 (~50 km)) and 6-hourly time step for 8-16 days (T190 (~70 km)) at 42 vertical levels.

The daily gridded rainfall data for the period 1985 to 2015 from the India Meteorological Department (IMD) at a spatial resolution of $0.25^\circ \times 0.25^\circ$ is used as the observational dataset [Pai et al., 2015]. This dataset is developed based on ground observations from 6995 stations across India using an inverse distance weighing scheme and is available for the period of 1901-2015 [Pai

et al., 2015]. The data is re-gridded to a resolution of $1^\circ \times 1^\circ$ to match the resolution of forecast data. Figure 1b shows the grids considered in this study to generate the forecasted hazard for Kerala during August 2018 flood event.

To compare the model performance in forecasting the extreme rainfall event, Tropical Rainfall Measurement Mission (TRMM) 3 hourly $0.25^\circ \times 0.25^\circ$ rainfall product is used to generate daily accumulated rainfall for 15th August 2018, the day of maximum rainfall during the flood event. We have used the TRMM data, as the gridded data from IMD was not available for 2018.

2.2.2. Socioeconomic and topographic vulnerability

To calculate the socioeconomic vulnerability, the subdistrict level demographic and economic information is obtained from the Census of India, 2011. As per the availability of data from Census of India, a set of relevant indicators are chosen to appropriately represent the status of socio-economy of the region [Vittal et al., 2020]. Table 1 shows the list of indicators chosen along with corresponding justifications in support of their selection. For the topographic information, we procure digital elevation model (DEM) product [Farr et al., 2007] from NASA Version 3.0 Shuttle Radar Topography Mission (SRTM) at Global 1 arc second ($\sim 30\text{m}$) resolution, which is further used in the Height Above the Nearest Drainage (HAND) model to generate the topographic vulnerability.

2.2.3. Exposure

Here, the land use land cover (LULC) data is used as a proxy to quantify the exposure component of risk, as it directly affects the severity of flood. For example, but not limited to, the land use pattern specifically over urbanized region leads to increasing the severity of flood due to increase in an impervious fraction, which eventually increases the exposure of land use to flood over urban hotspots [Karmakar et al., 2010]. The decadal 100 m resolution

LULC dataset for 2005 from ORNL DAAC [Roy et al., 2016] is used for this purpose, which classifies the LULC into 19 classes, as provided in Supplementary table S2.

3. Limitations of State-of-the-Art Weather Forecasts

The state-of-the-art flood alert generation system uses the forecasted rainfall amount during extreme rainfall events. However, these forecasts are often associated with very low hit rate and high false alarm along with spatiotemporal biases. Here, we consider the weather to medium range forecasts of extreme rainfall event in Kerala on 15th August, 2018, at multiple lead times. Since, the meteorological forecasts have significant bias, we apply bias correction by scaling [Maraun, 2016], using the hindcast and observed data. The bias corrected outputs are presented in Figure 3. Following the conventional approach bias correction is done for the entire available hindcast dataset but the result is presented only for a single storm. Even at lead day 1, the model completely fails to produce the magnitude of the extreme rain (Figure 2a) over the state of Kerala. Further to this, at a lead time of 10 and 15 days, the forecasts are not at all showing indications of extremes over Kerala, and at a lead time of 2-3 days, it started showing a bit of indication of moderate rainfall. Clearly these are not sufficient for an evacuation action or taking any other precautionary measures. We also find that these forecasts are associated with very low hit rate and high false alarms. Here we define hit rate as the fraction of extremes predicted by the model successfully and false alarm as the fraction of time model predicts a non-extreme event as an extreme. Hit rate and false alarm are calculated as:

$$\text{Hit rate} = \frac{a}{a + c} \quad (1)$$

$$\text{False alarm} = \frac{b}{a + b} \quad (2)$$

where

a = number of intersects between A and B

b = number of intersects between A and \bar{B}

c = number of intersects between \bar{A} and B

A = Number of times model forecasts an extreme

B = Number of times observation shows an extreme

\bar{A} = Number of time model forecasts a non-extreme rainfall

\bar{B} = Number of time observation shows a non-extreme rainfall

The hit rate and false alarm associated with the forecasts are given in Supplementary figure S1. The hit rate always remains around 10% at any lead time with a very slight improvement at a lead time of 1 day. False alarm remains as high as 90%. It is evident that they are not enough for decision making towards disaster mitigation.

4. Method

Here, to address the above-mentioned problems associated with the uncertain and biased forecasts, we propose a novel approach of forecasting flood risk at weather to extended range scale. This is in contrast to the conventional approach of flood forecasting, which uses bias corrected weather forecasts. Conventionally, the risk to extremes is defined as the product of hazard, vulnerability and exposure as [Chen et al., 2015; Gusain et al., 2020; IPCC, 2012; Karmakar et al., 2010; Kron, 2005; Sahani et al., 2019] :

$$Risk = Hazard \times Vulnerability \times Exposure \quad (3)$$

The concept of risk has been widely used at a climate scale; however, we propose the same at a weather to medium range. The difference in the concept of risks between weather to medium range and climate scales lies in the definition of hazards. We define hazard as the probability of an extreme event (above a defined threshold) given the forecasts for the same day. The flood risk at a weather to medium range incorporates weather to medium range forecasts, which essentially makes the hazard component dynamic, whereas the remaining components of flood risk (vulnerability and exposure) are considered to be static. During extreme events, based on the forecasted risk maps generated at a weather to medium range, the high risk areas can be identified and targeted first. Hence, a location specific evacuation and flood mitigation can be done well in advance to reduce the losses. Figure 4 shows a complete flowchart of the methodology used to generate weather to medium range event specific flood risk maps. Further, these maps need to be generated at an administrative (eg. district, subdistrict or village for India) level depending upon the data availability for better efficiency in decision making. In the present study, flood risk is generated at subdistrict level based on the authoritarian decision process and availability of demographic data.

4.1. Hazard

Hazard is typically defined as the probability of an extreme event. This concept is traditionally used at a climate scale [Gusain et al., 2020; Sajjad et al., 2020]. As for example a hazard associated with 95th percentile of rainfall is 0.05. Here, we propose to define hazard at a weather to medium range and define it as the probability of getting an extreme rainfall in any grid, given a forecast value. This is a dependant of the forecasts, and hence with the change in lead time and subsequent forecasts, the hazard value at a location gets modified. At a location (for example for a grid), if we define extreme rainfall event as the rainfall above a threshold of 99th percentile, the hazard may be defined as:

$$\text{Hazard} = P(O \geq O_{99} | H = f) \quad (4)$$

where,

O = observed daily rainfall

O_{99} = 99th percentile of rainfall observations

H = forecasted (Hindcast) rainfall with its value denoted as f

In order to generate the conditional probability given in equation 4, we first obtain the joint probability of observed and hindcast (from the same model which is being used for forecast).

To generate the joint probability, copula is used which creates the multivariate distribution based on the individual marginal distributions [Dupuis, 2007; Ghosh, 2010]. Copula does not need the marginal distributions to follow a specific and same distribution. This makes copula advantageous over conventional multivariate parametric multivariate distributions like normal, log-normal.

The first step in applying the copula-based approach is to obtain the marginal distribution functions of associated variables, which are the observed rainfall (O) and hindcast rainfall (H). As both observational and hindcast rainfall data contain zero rainfall values, mixed marginal distributions are used. Gamma distributions are fitted to the non-zero values of rainfall. To model the probability mass function of zero and non-zero rainfall, we apply Bernouli Trials. The CDF of the variables (X , which stands for both O and H) is given by:

$$F_X(x) = \begin{cases} p + (1 - p) G_X(x) & , x > 0 \\ p & , x = 0 \end{cases} \quad (5)$$

where,

$$p = \text{probability of getting zero rainfall} = \frac{n_1}{n}$$

n = total no of days

n_1 = no of zero rainfall days

$G_X(x)$ = CDF of nonzero rainfall obtained by fitting a Gamma distribution

After generating the marginal distributions, Archimedean copula is used to generate the bivariate distribution. In this case, O and H are the two variables for which copula is to be fitted. By definition of copula, a two dimensional distribution function is given by:

$$F(o, h) = C(F_O(o), F_H(f)) \quad (6)$$

where,

C = Copula

F_O and F_H = marginal distribution functions of O and H

In the approach based on Copula, we consider two variables, $U = F_O(o)$ and $V = F_H(f)$ to be the CDF of O and H respectively, where U and V are uniformly distributed random variables with values u and v. Here a single parametric copula is used with parameter θ [Ghosh, 2010; Nelsen, 1999; Zhang & Singh, 2006]:

$$C_\theta(u, v) = \phi^{-1}\{\phi(u) + \phi(v)\} \quad (7)$$

where,

$\phi(\bullet)$ is a convex decreasing function (copula generator)

The parameter θ is generated using the relationship between Kendall's coefficient of correlation (τ) and $\phi(\bullet)$ [Ghosh, 2010; Karmakar and Simonovic, 2009]:

$$\tau = 1 + 4 \int_0^1 \frac{\phi(t)}{\phi'(t)} dt \quad (8)$$

where,

$t = u$ or v

Kendall's τ calculated using the following equation:

$$\tau_n = \binom{n}{2}^{-1} \sum_{i < j} \text{sign}[(x_i - x_j)(y_i - y_j)] \quad (9)$$

where,

$$\text{sign} = 1 \text{ for } [(x_i - x_j)(y_i - y_j)] > 0$$

$$\text{sign} = 0 \text{ for } [(x_i - x_j)(y_i - y_j)] \leq 0$$

$i, j = 1, 2, \dots, n$

Three types of Archimedean copulas are used namely Gumbel, Frank and Clayton. The relationship between τ and θ and the mathematical equations for each of these copulas are given in Supplementary table S3. The most important step in this approach is the choice of copula to best fit the distributions [Favre et al., 2004]. The copula having minimum Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) score is considered to be the best fitted copula. For correct representation of extreme events, tail dependence test is important for the selection of copula [Ghosh, 2010; Poulin et al., 2007]. As we are dealing with extreme rainfall events, copula having the highest upper tail dependence coefficient is desirable. Here, the upper tail dependence coefficient is computed using a nonparametric estimator, CFG (Capeeraa, Fougères, Genest) [Capéraà et al., 2000]. Based on AIC, BIC and upper tail dependence coefficient, the best copula is selected.

The other important criteria is the selection of grids for analyzing the forecasts. Ideally, the forecasted and observed rainfall should belong to the same grid. However, there is a high possibility of spatial bias in the forecasts that generates rainfall to a neighbouring place of the

area of interest. To overcome this, we perform three analyses: a) considering only the grid of interest for both observations and the forecasts, b) considering maximum of the observed and forecasted rainfall over a 3×3 box of 9 grids with the grid of interest at the centre; c) considering maximum of the observed and forecasted rainfall over a 5×5 box of 25 grids with the grid of interest at the centre. The hazard values thus generated using equation 4 are in gridded form which are converted into subdistrict level by area weighted average method.

4.2. Vulnerability

The vulnerability component of flood risk is calculated as the average of socioeconomic and topographic vulnerabilities.

4.2.1. Socioeconomic Vulnerability

Here, a measure of a region's susceptibility to flood damage is referred as flood vulnerability, which includes a portion of population susceptible to either emotional, mental or physical damage. In addition, the seriousness of current situation and previous experiences with disastrous event may further influence the vulnerability [Karmakar et al., 2010]. During a disastrous event, the socio-economic vulnerability mainly focuses on response, reaction and resistance for a population of a region, along with the damage caused to the economic sector. Here, the framework proposed by Sherly et al. (2015) and Vittal et al. (2020) is implemented to estimate the socio-economic vulnerability over the study region. The major steps followed, which include, judicious selection, standardization and aggregation of indicators and subsequent ranking of sub-districts have been shown in Figure 5. A step-by-step exposition has been provided in the following paragraphs.

Choice of indicator plays an important role in socioeconomic vulnerability quantification. The indicators chosen should be relevant, justifiable and a good representative of the social

and economic condition of the concerned region [Vittal et al., 2020]. The set of indicators selected for our study along with their justification are given in Table 1. Identifying an indicator as positive (sensitive) or negative (adaptive) has a significant influence on the overall vulnerability, and thus appropriate recognition of such indicators is crucial [Vittal et al., 2020; Sharma et al., 2020]. A positive (negative) indicator increases (decreases) the vulnerability, and consequently affect the flood risk. Here, the vulnerability indicators are standardized mainly to make the indicators dimensionless, which will allow us to compare the different indicators over the study region. The method of standardization for an indicator (Wu et al., 2002; Karmakar et al., 2010) is provided in the equation below:

$$V_i^{std} = \frac{V_i - V^{min}}{V^{max} - V^{min}} \quad \text{for negative (sensitive) indicators} \quad (10a)$$

$$V_i^{std} = \frac{V^{max} - V_i}{V^{max} - V^{min}} \quad \text{for positive (adaptive) indicators} \quad (10b)$$

where,

V_i^{std} = standardized vulnerability indicator of i^{th} subdistrict

V_i = vulnerability indicator of i^{th} subdistrict

V^{min} = minimum vulnerability indicator in all the subdistricts

V^{max} = maximum vulnerability indicator in all the subdistricts

These equations consider both maximum and minimum values in the expression, and ensure that the vulnerability values are within a [0, 1] interval (Wu et al., 2002) and always non-negative (Karmakar et al., 2010). During standardization, the indicators were adjusted for their sign, which indicates whether each indicator contributes positively (+) or negatively (−) to overall vulnerability.

The inherent subjectivity involved in the selection of threshold values for different classes or ranges makes it difficult to categorize the vulnerability [Holand et al., 2011; Mitchem, 2004; Uitto, 1998]. To address this issue, there have been different approaches adopted by the researchers ranging from simple averaging [Karmakar et al., 2010] to more complex cluster analysis [Kok et al., 2016; Sietz et al., 2011]. Although, these approaches are equally efficient for an elegant representation, these lacks in reducing the subjectivity in selection of weights [Rygel et al., 2006]. Contrarily, Data Envelopment Analysis (DEA) [Huang et al., 2011; Sherly et al., 2015; Vittal et al., 2020; Wei et al., 2004] does not require weight assignment, thereby reducing the subjectivity; and also introduces a new classification approach with minimal possibility of rank reversal [Huang et al., 2011; Wei et al., 2004]. In addition, DEA does not make any assumption on the form of the functions, as it is a non-parametric technique. Therefore, the present study implements DEA to aggregate the standardized indicators and subsequently rank each spatial unit, known as Decision Making Units (DMUs), i.e., subdistricts in the present study.

To use the DEA model efficiently, the indicators should have very low correlation. In order to decorrelate the highly correlated indicators and decrease the dimensionality of these indicators, Principal Component Analysis (PCA) is performed. Principal Components (PCs) explaining ~75% of the variability are considered as inputs for the DEA. Following the approach of Sherly et al. (2015) and Vittal et al. (2020), a dummy value 1 (unity) is assigned as output in the DEA, since it represents the state of the system prior to the occurrence of a hazard. Hence, the socioeconomic vulnerability of each DMU (subdistrict) is obtained by subtracting the relative efficiency of that subdistrict from unity. In our case, 2 PCs are considered as input and the Banker-Charnes-Cooper (BCC) model [Banker et al., 1984] of DEA is used to rank the DMUs by calculating their relative efficiencies. Lower the efficiency, lower is the rank and higher is the vulnerability.

4.2.2. Topographic vulnerability

To quantify the flood risk during extreme rainfall events, elevation of a region plays a very important role. Topographic vulnerability accounts for the elevation and HAND method [Rahmati et al., 2018; Rennó et al., 2008; Nobre et al., 2011, 2016] is used to quantify this parameter. The model normalises Digital Elevation Model (DEM) values by changing the elevation with respect to sea level into elevation with respect to nearest drainage. The advantage of using HAND values to calculate topographic vulnerability is that, grids with different elevation values, but same HAND value are considered equally vulnerable. A flow chart of the methodology followed to calculate topographic vulnerability at grid level is shown in Figure 6 and the steps followed are described considering an example DEM in Supplementary figure S2.

Firstly, the DEM with sinks is used as input for the HAND model. A sink (or depression) is an area or a point which has an elevation lower than all its neighbouring area or point [Rieger, 1998] (marked in red circles in Supplementary figure S2a). As it does not have any drainage outlet, flow network generation is not possible with DEM having sinks. Depression breaching method [Martz and Garbrecht, 1999] is used to connect these sinks and a hydrologically coherent DEM is generated (Supplementary figure S2c). The breaching method connects two neighbouring sinks by lowering elevation of some points in the shortest path connecting them. Flow direction is computed from this hydrologically coherent DEM using D8 method [Tarboron, 1997]. In this method flow from each grid is assigned to one of its eight neighbours having the steepest slope to generate a local drainage direction (LDD). The LDD generated using hydrologically coherent DEM is called Coherent LDD (Supplementary figure S2d). Next, a flow accumulation for each grid is computed by the adding the number of grids draining into that grid to get an accumulated area map

(Supplementary figure S2e). In order to generate a drainage network map, channel initiation is done by setting up a threshold of accumulated area (set as 10 in the example) in the map. Grids having an area above the threshold value are considered in the drainage network. This threshold can be applied manually or automatically, by using an accurate drainage network map, called mapped stream network (MSN) as input. Using the LDD with drainage network, nearest drainage map is generated and each grid is associated with the grid that it drains into, known as the drainage grid. An example is shown in Supplementary figure S2g, where each drainage grid is given a colour and the associated grids are represented as a lighter shade of the same colour. The relative height of each grid is calculated as the difference of height of that grid and its drainage grid to generate a HAND map (Supplementary figure S2h).

To calculate topographic vulnerability for Kerala, an automated geographic information system GIS tool of the HAND model developed by Rahmati et al.(2018) is used. Due to the unavailability of an accurate MSN, an appropriate threshold is to be selected manually by trial and error method as suggested by Rahmati et al. (2018). Here, 25000 is found to an appropriate threshold, as higher values fail to give any positive value to the grids in the low lying areas (coastal areas) and lower values result in a very high density drainage network. Using this drainage network, nearest area map is generated and HAND values calculated at a grid level (~30m). These gridded HAND values are converted to subdistrict level by weighted average method. Lower the HAND value, higher is the topographic vulnerability. Subdistrict level HAND values are standardised between 0 and 1 and are subtracted from 1 to obtain the topographic vulnerability.

4.3. Exposure

LULC of an area is used as indicator to quantify exposure of any area to flood. LULC is a primary characteristic of any region and decides the soil permeability, run off etc. Following

the methodology used by Karmakar et al. (2010) each LULC category is assigned with a Degree of Importance (DI). For example, the built-in area has more pavement and concrete surfaces which increases the run off and is more prone to losses during flood, and thus is assigned a higher DI value. On the other hand, grassland like open areas have very low DI values since they allow infiltration and decrease run off, so suffer lesser losses. DI values associated with each of the LULC type (as described by Karmakar et al. (2010)) are given in Table 2. Using the DI values and the fraction of area they covered by any LULC type in a subdistrict, exposure is calculated as:

$$E_i = \sum_{l=1}^n \left[DI_l \times \left(\frac{A_i^l}{A_i} \right) \right] \quad (11)$$

where,

E_i = exposure of subdistrict i

l = LULC type

DI_l = Degree of importance for LULC type l

A_i^l = Area occupied by LULC type l in subdistrict i

A_i = Area of subdistrict i

The subdistrict wise exposure values obtained are then standardised between 0 and 1 and exposure maps are generated.

5. Results

5.1. Hazard

In the proposed methodology hazard is calculated for different lead times, using the rainfall forecast, as the conditional probability of extreme rainfall event given the forecasts (equation 4). To get the conditional probability, joint probability needs to be calculated from the bivariate distribution of observed and forecasted (hindcasted) rainfall. A bivariate copula is used and three Archimedean copulas are fitted: Gumbel, Frank and Clayton in order to find the best fit. Among these three, Gumbel is found to be the best fitted, for all the grids and at all the lead times, as it has the minimum AIC and BIC values and the maximum upper tail dependence coefficient, calculated using the CFG-estimator. Hence, Gumbel copula is identified as the most suitable and is used to generate the conditional probability of extreme rainfall given the forecasts and to calculate the hazard. The daily accumulated rainfall forecasted at lead day 15 (forecasted on 1st August, 2018) to lead day 1 (forecast done on 15th August, 2018) is used in equation 4 to forecast hazard at respective lead days for the event that took place on 15th August, 2018.

Here, to consider the spatial bias, we propose to apply three approaches as mentioned in Section 4.1, by considering the rainfall data, (a) only at the grid of interest; (b) which is maximum over 3×3 grid boxes centred on the grid of interest and (c) which is maximum over 5×5 grid boxes centred on the grid of interest. First, we calculate the hazard for each grid using the observed and forecasted rainfall of the grid of interest at different lead times (Supplementary figure S3). For case (a), the higher hazard areas are not matching with the most affected areas (hatched areas) until lead day 2. Even on lead day 1, only some of the hatched parts are shown to be having a high forecasted hazard value. This may be attributed to the spatial bias present in the model which results in forecasting the rainfall in the

neighbouring grid. The other two analyses (case (b) and (c)) are done, by considering the maximum rainfall in a 3×3 box of 9 grids and 5×5 box of 25 grids with the grid of interest at the centre. The hazard maps generated using the 3×3 (Figure 7) and 5×5 boxes (Supplementary figure S4) shows high values at the areas that suffered maximum losses from lead day 15. This further proves that the rainfall forecast is likely to have a spatial bias in the model which can be overcome by also including the neighbouring grids in the analysis. Since the 3×3 and 5×5 cases (Cases (b) and (c)) do not show much disagreement, the maximum rainfall forecasted in a 3×3 box is used to generate the gridded hazard maps for each grid (Figure 7b (1-6)) further which is used to calculate risk. Hazard for the same event is generated using rainfall forecast from all the 11 ensembles present in the GEFS. It found that no major difference exists in the hazard values between different ensembles. This is because of the consideration of conditional probabilities from the hindcast.

In the gridded hazard maps from lead day 5 most of the grids start showing very high value, which implies our method is able to predict the extreme event. The hazard maps generated by this method are able to take into account the model's inability to forecast extreme rainfall magnitude correctly. We further use area weighted average method to convert these gridded values to subdistrict scale values (Supplementary figure S5). The subdistrict wise hazard values are standardised between 0 and 1 and divided in to five categories; very low (0 - 0.2), low (0.2 - 0.4), medium (0.4 - 0.6), high (0.6 - 0.8) and very high (0.8 - 1). Starting from day 15, till lead day 10, the standardized hazard values are showing a low to very low values in most of the affected areas. These values start to increase by lead day 5 and show high to very high hazard in most of the affected areas. As these subdistrict wise hazard values are obtained based on the method that considers 3×3 grids, these show a very little variability among adjacent subdistricts. Hence it is difficult to identify the high-risk zones in the area under consideration using these subdistrict wise hazard maps.

To understand the applicability of the model for other extremes, we apply the same to six extreme events that took place during 1985-2015 in the study region (Supplementary table S4). We present the composites of hazard values with their band (from six events) at different lead days (Supplementary figure S6 (b-i)). It is quite interesting to note that within lead days of 10, for all the cases of extremes, the hazard value comes almost same and this is a good indication of identifying a correct threshold of hazard for a specific grid. Such an identification of threshold must be done probably with a higher number of extreme events corresponding to different high percentiles. We further compute the false alarm ratio by considering the days to have false alarms when they are not extreme days, but the model shows a higher hazard with respect to the threshold corresponding to the specific grid and lead day (Supplementary figure S7). We find still a huge false alarm ratio exists with our proposed post processing approach. The false alarm ratio drops a bit at a lead day of 1. Such a huge false alarm for the west coast of India during monsoon was also reported in Shastri et al., (2017). Reducing false alarm needs improvements in the weather models and such improvements are not possible using post-processing techniques alone for this specific region and the season. For a coastal region, it also needs finer resolution models to take care of the sea-land interface.

However, given the forecasts these are the best estimates and among the high hazard zones, the hotspots are identified with the help of vulnerability and exposure values. Hence, we introduce the concept of risk that considers all three aspects, hazard, vulnerability and exposure. The next subsections present the results obtained from the vulnerability analysis.

5.2. Vulnerability

5.2.1. Socioeconomic Vulnerability

Subdistrict wise socioeconomic vulnerability is computed using DEA approach, which calculates the relative efficiency of each subdistrict using various vulnerability indicators. Figure 8a shows the subdistrict wise socioeconomic vulnerability map of Kerala. To understand the importance of socio-economic vulnerability, we overlaid the economic loss map on the vulnerability map showing the highly affected districts with hatching. We find that most of the highly affected regions have high socio economic vulnerability. The results also show the correctness of the selection of the socioeconomic indicators for this present study.

5.2.2. Topographic vulnerability

The topographic vulnerability values, calculated using the HAND method and the vulnerability map is given in Figure 8b. Among the highly affected districts, Alappuzha, Kottayam and some parts of Pattanamtitta and Wayanad have very high topographic vulnerability but Idukki and eastern parts of Pattanamtitta show very low to low vulnerabilities though losses in these areas are high. This is because, Idukki and parts of Pattanamtitta lie in Western Ghats mountain ranges resulting in higher HAND values and thus lower topographic vulnerability.

5.3. Exposure

Exposure is calculated at subdistrict level as the sum of product of each the DI value of a LULC type with the fraction of area covered by each type using equation 11 and exposure maps are generated (Figure 8c). Subdistricts with more crop areas and built-in areas such as

Ernakulam, Kottayam and Wayanad are more exposed to flood risks because of a high DI value of these LULC types. Idukki and parts of Alappuzha show low exposure due to the presence of more forest, barren lands and, water bodies which are associated with low DI values.

5.4. Risk

Subdistrict wise relative risk maps are generated for the extreme rainfall event that took place on 15th August at different lead days, by combining hazard, vulnerability and exposure, using equation 3 (Figure 9). Risk is standardised and is categorised into 5 predefined categories, very low (< 20%), low (20-40%), medium (40-60%), high (60-80%) and very high (> 80%). It should be noted that “very low” here refers to very low relative risk compared to other sub-district, but at an absolute level, they may be still high. This risk map helps to prioritize sub-district level the action plans by selecting the region with highest relative risk. Subdistricts with relatively higher relative risk are identified in this process, which could have been prioritised for implementation of evacuation strategies, planning response and recovery practices and, other mitigation.

Comparing the generated relative risk maps (Figure 9) with the actual loss data (Figure 2b), it is evident that high relative risks in the districts like Wayanad, Thrissur, Ernakulam and Kottayam are forecasted by the model well in advance with a lead time of up to 15 days. Some subdistricts in Malapuram, which had reported moderate losses (Figure 2b), are also showing high to very high relative risk throughout the 15 days forecast. The predicted relative risk in Kannur region is higher in the extended range (15–8 days) forecasts because of the higher rainfall prediction in the region during that period. This relative risk eventually reduces with lead time as the rainfall forecast improves. The model failed to forecasts the high relative risk in some parts of Alappuzha district, because of the low exposure value

despite having a very high vulnerability value (Figure 8). The low exposure value in that area is attributed to the small water bodies, which cover a significant fraction of the total area, and these water bodies are assigned with a very low DI value (0.1). Losses in Idukki, Palakkad and eastern parts of Pattanamtitta are mostly attributed to landslides in the Western Ghats during the event. Our model does not consider landslides and hence, those areas are not identified by the relative *flood* risk maps as high risk zones. The future scope is to consider a landslide model in this framework to further improve the forecasting skill.

The loss data for the flood is available at district level whereas we are predicting risk at subdistrict level. It is evident from Figure 9 that in some of the affected districts, not all subdistricts are showing a high relative risk. This may be due to the fact that the reported loss in a district is dominated by some of the subdistricts in it. Overall, this method is successful in identifying most areas effected by direct flood as high-relative risk areas at a lead time of almost 15 days.

6. Summary

Extreme rainfall events show an increasing trend in the Indian subcontinent and so do the resultant flood events. The state-of-the-art rainfall forecast system is usually associated with a very high false alarm, low hit rate and spatiotemporal biases. In recent years there have been a number of cases where the prediction of extreme rainfall is either spatially or temporally inaccurate or the amount of rainfall predicted is wrong altogether. These forecasts were not good enough to be implemented in mitigation planning, which thus resulting in huge loss of life and property. Here a new methodology has been proposed which generates relative flood risk maps at weather to medium range as a product of hazard, vulnerability, and exposure. This methodology is applied to the August 2018 Kerala floods in India for demonstration. The concepts of vulnerability and exposure do not really add to the predictive information but

helps to identify the regions with high risk, well in advance. The flood event during 2018 over Kerala showed that all the areas with high flood losses did not necessarily experienced spatially highest amount of rainfall in the state. The losses were also governed by the vulnerability and exposure. Therefore this concept is brought into the forecast system with the proposed approach.

Hazard is defined as the probability of getting an extreme rainfall (above 99th percentile) in any grid given a forecast value. GEFS forecast data is used to generate hazard forecast up to 15 lead days. In order to find the conditional probability, a bivariate copula is fitted to the GEFS hindcast and IMD observed rainfall data for the period 1985 to 2015. CDF of observed and hindcasted rainfall data is found by fitting a mixed distribution where Gamma distribution is used for the non-zero values. For selection of the best copula, three Archimedean copulas are considered, namely Gumbel, Frank and Clayton. Gumbel is selected as it showed minimum AIC, BIC values and maximum tail dependence among the three Archimedean copulas. In order to take care of the spatial bias present in the rainfall forecast model, the maximum rainfall observed and forecasted in the neighbouring (3×3) box of 9 grids with the grid of interest in centre, is considered as the rainfall observed and forecasted respectively for a grid. The hazard generated is able to predict extreme rainfall by showing high values but fails to identify areas under high risk due to coarse spatial resolution. Hence, along with the hazard values other local parameters such as the social and economic conditions, topography and LULC are also incorporated in terms of vulnerability and exposure in order to find the relative risk of the units in the region of interest.

Vulnerability is defined as average of socioeconomic and topographic vulnerabilities. Socioeconomic vulnerability describes the socioeconomic conditions of an area. From the Census of India (2011) data, various positive and negative indicators are considered in order to quantify the socioeconomic vulnerability. PCA is performed to reduce the dimensionality

of these indicators and 2 PCs explaining 75% of the variability are used as input for the DEA framework. BCC model of DEA is used to calculate the relative efficiency of the subdistricts in order to rank them. Socioeconomic vulnerability is computed by subtracting the relative efficiency from 1. Topographic vulnerability is calculated using the HAND method which uses DEM output and generates HAND maps, where the relative height of each grid with respect to the flow path is considered. The smaller HAND values show larger vulnerabilities, so the value subtracted from 1 is used as the topographic vulnerability. Exposure for each subdistrict is calculated by assigning DI values to different LULC types and multiplying these with the area fraction covered by each type in that subdistrict.

To evaluate the performance of the method, it is applied to the devastating floods in Kerala during August, 2018. Relative risk maps are generated at a subdistrict level at different lead days by combining the hazard generated with corresponding lead days with socioeconomic and topographic vulnerability and, exposure. These maps are compared with the district wise loss data from KPDNA (2018) reports which is used as a proxy for severity of the flood. Most subdistricts in the highly affected districts like Wayanad, Thrissur, Ernakulam and Kottayam and moderately affected district Malappuram are identified as high relative risk areas starting from lead day 15. Though losses due to flood is very high in Alappuzha, the model fails to identify them in some of its subdistricts due to presence of water bodies with very low exposure value. The high relative risk areas identified in the risk maps could have been given priorities in flood mitigation and evacuation planning in order decrease the flood losses.

The risk model can be applied to any different case studies and may suitably be adjusted/modified depending on data availability. As for example, the major issue, a new case study area may face, are the problems associated with the availability of socio-economic data. The elevation data and the gridded rainfall data, as applied in the present manuscript, are mostly

available for majority of areas around the globe. The socio-economic vulnerability component in the present model is a flexible one to consider a lower availability of variables. The model is sensitive to the component used for vulnerability analysis and it does not consider intermodal uncertainty across different approaches used for computing vulnerability.

The proposed model assumes that only local rainfall feeds into the regional flood and does not consider either the upstream hydrology or the memory resulting from near past heavy rainfall. Such an assumption does not hold true for large watersheds. The units considered in the model are rather administrative depending on the availability of socio-economic data. Understanding of flood needs the consideration of hydrological units, such as watershed or sub-basin. Future scope of the work lies in addressing these limitations. Further improvement of this risk based forecast system can be done, by using in-situ station level or finer resolution rainfall data. By combining this method with hydro-economic models, flood loss can be predicted accurately which is very useful for reduction and assessment of losses during floods. Often extreme rainfall in hilly regions lead to landslides, and thus a landslide module can be added to the present model to further improve it for a multi-hazard system. Another limitation of this framework is combining multiple indicators like hazard, vulnerability and exposure using product-based combination/aggregation. Though this approach has been popularly used and are efficient for a quick and relative representation, it may not be the best metric. A solution could be a multi-criteria decision making approach which can be considered as a potential area of future research.

credit_statement

SG conceptualized the idea and designed the overall algorithms and methodology; ST performed the formal analysis. SK and VH performed the socio-economic vulnerability analysis. ST and SG performed investigation and analysis of results. ST performed data curation. SG and ST wrote the manuscript. SK and VH reviewed and edited the manuscript. SG and SK supervised the work. SG and SK did the acquisition of the financial support for the project leading to this publication.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Journal Pre-proof

Acknowledgements

The authors thank the Department of Science and Technology for the financial support through sponsored research project DST/CCP/CoE/140/2018. The authors acknowledge India Meteorological Department (IMD), Global Ensemble Forecast System (GEFS) reforecast and Tropical Rainfall Measuring Mission (TRMM) for the rainfall datasets. We thank Census of India, Government of India for the demographic dataset, NASA Version 3.0 Shuttle Radar Topography Mission (SRTM) for the DEM dataset and ORNL DAAC for the LULC dataset. We extend our gratitude to the Kerala Flood is provided by Kerala State Disaster Management Authority (KSDMA) for providing the Kerala Post Disaster Needs Assessment- August 2018 [KPDNA, 2018] report.

References

- Apel, H., Aronica, G. T., Kreibich, H., & Thielen, A. H. (2009). Flood risk analyses - How detailed do we need to be? *Natural Hazards*, 49(1), 79–98. <https://doi.org/10.1007/s11069-008-9277-8>
- Balk, D., Montgomery, M. R., & Liu, Z. (2012). Urbanization and Climate Change Hazards in Asia. *Population Association of America, 2013 Annual Meeting, New Orleans, LA*.
- Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, 30(9), 1078–1092. <https://doi.org/10.1287/mnsc.30.9.1078>
- Březková, L., Starý, M., & Doležal, P. (2010). The real-time stochastic flow forecast. *Soil and Water Research*, 5(2), 49–57.
- Capéraà, P., Fougères, A. L., & Genest, C. (2000). Bivariate Distributions with Given Extreme Value Attractor. *Journal of Multivariate Analysis*, 72(1), 30–49. <https://doi.org/10.1006/jmva.1999.1845>
- Census of India. (2011). *Provisional population totals, Census of India. New Delhi: Office of the Registrar General and Census Commissioner, India*.
- Chen, Y., Liu, R., Barrett, D., Gao, L., Zhou, M., Renzullo, L., & Emelyanova, I. (2015). A spatial assessment framework for evaluating flood risk under extreme climates. *Science of the Total Environment*, 538, 512–523. <https://doi.org/10.1016/j.scitotenv.2015.08.094>
- Coumou, D., & Rahmstorf, S. (2012). A decade of weather extremes. *Nature Climate Change*, 2(7), 491–496. <https://doi.org/10.1038/nclimate1452>

- Dodla, V. B. R., & Ratna, S. B. (2010). Mesoscale characteristics and prediction of an unusual extreme heavy precipitation event over India using a high resolution mesoscale model. *Atmospheric Research*, 95(2–3), 255–269. <https://doi.org/10.1016/j.atmosres.2009.10.004>
- Dottori, F., Szewczyk, W., Ciscar, J.-C., Zhao, F., Alfieri, L., Hirabayashi, Y., Bianchi, A., Mongelli, I., Frieler, K., Betts, R. A., & Feyen, L. (2018). Increased human and economic losses from river flooding with anthropogenic warming. *Nature Climate Change*. <https://doi.org/10.1038/s41558-018-0257-z>
- Dube, A., Ashrit, R., Ashish, A., Sharma, K., Iyengar, G. R., Rajagopal, E. N., & Basu, S. (2014). Forecasting the heavy rainfall during Himalayan flooding-June 2013. *Weather and Climate Extremes*, 4(June 2013), 22–34. <https://doi.org/10.1016/j.wace.2014.03.004>
- Dupuis, D. J. (2007). Using copulas in hydrology: Benefits, cautions, and issues. *Journal of Hydrologic Engineering*, 12(4), 381–393. [https://doi.org/10.1061/\(ASCE\)1084-0699\(2007\)12:4\(381\)](https://doi.org/10.1061/(ASCE)1084-0699(2007)12:4(381))
- Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., & Alsdorf, D. (2007). The Shuttle Radar Topography Mission. *Reviews of Geophysics*, 45(2). <https://doi.org/10.1029/2005RG000183>
- Favre, A. C., Adlouni, S. El, Perreault, L., Thiémondge, N., & Bobée, B. (2004). Multivariate hydrological frequency analysis using copulas. *Water Resources Research*, 40(1). <https://doi.org/10.1029/2003WR002456>
- Fowler, H. J., Cooley, D., & Sain, S. R. (2010). Detecting change in UK extreme precipitation using results from the climate prediction . net BBC climate change experiment. 241–267. <https://doi.org/10.1007/s10687-010-0101-y>

- Fritsch, J. M., Houze, R. A., Adler, R., Bluestein, H., Bosart, L., Brown, J., Carr, F., Davis, C., Johnson, R. H., Junker, N., Kuo, Y. H., Rutledge, S., Smith, J., Toth, Z., Wilson, J. W., Zipser, E., & Zrnich, D. (1998). Quantitative Precipitation Forecasting . In *Bulletin of the American Meteorological Society* 79(2), 285–299.
- Ghosh, S. (2010). Modelling bivariate rainfall distribution and generating bivariate correlated rainfall data in neighbouring meteorological subdivisions using copula. *Hydrological Processes*, 24(24), 3558–3567. <https://doi.org/10.1002/hyp.7785>
- Goswami, B. N., Venugopal, V., Sengupta, D., Madhusoodanan, M. S., & Xavier, P. K. (2006). Increasing Trend of Extreme Rain Events Over India in a Warming Environment. *Science*, 314(December), 1442–1445. <https://doi.org/10.1126/science.1132027>
- Gusain, A., Mohanty, M. P., Ghosh, S., Chatterjee, C., & Karmakar, S. (2020). Capturing transformation of flood hazard over a large River Basin under changing climate using a top-down approach. *Science of the Total Environment*, 726, 138600. <https://doi.org/10.1016/j.scitotenv.2020.138600>
- Hamill, T. M., Bates, G. T., Whitaker, J. S., Murray, D. R., Fiorino, M., Galarneau, T. J., Zhu, Y., & Lapenta, W. (2013). NOAA's Second-Generation Global Medium-Range Ensemble Reforecast Dataset. *Bulletin of the American Meteorological Society*, 94(10), 1553–1565. <https://doi.org/10.1175/BAMS-D-12-00014.1>
- Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., Kim, H., & Kanae, S. (2013). Global flood risk under climate change. *Nature Climate Change*, 3(9), 816–821. <https://doi.org/10.1038/nclimate1911>
- Holand, I. S., Lujala, P., & Rod, J. K. (2011). Social vulnerability assessment for Norway: A quantitative approach. *Norsk Geografisk Tidsskrift*, 65(1), 1–17.

<https://doi.org/10.1080/00291951.2010.550167>

Huang, J., Liu, Y., & Ma, L. (2011). Assessment of regional vulnerability to natural hazards in China using a DEA model. *International Journal of Disaster Risk Science*, 2(2), 41–48. <https://doi.org/10.1007/s13753-011-0010-y>

IFRC/RCS. (2011). *World disasters report 2011: focus on hunger and malnutrition*. Geneva, Switzerland: International Federation of the Red Cross and Red Crescent Societies.

IPCC. (2007): *Climate Change 2007: Impacts, Adaptation and Vulnerability*. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, M.L. Parry, O.F. Canziani, J.P. Palutikof, P.J. van der Linden and C.E. Hanson, Eds., Cambridge University Press, Cambridge, UK, 976.

IPCC. (2012). Glossary of Terms. In C. B. Field, V. Barros, T. F. Stocker, D. Qin, D. J. Dokken, K. L. Ebi, M. D. Mastrandrea, K. J. Mach, G.-K. Plattner, S. K. Allen, M. Tignor, & P. M. Midgley (Eds.), *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change (IPCC)* (pp. 555–564). Cambridge University Press.

Jenamani, R. K., Bhan, S. C., & Kalsi, S. R. (2006). Observational/forecasting aspects of the meteorological event that caused a record highest rainfall in Mumbai. *Current Science*, 90(10), 1344–1362.

Karmakar, S., & Simonovic, S. P. (2009). Bivariate flood frequency analysis. Part 2: A copula-based approach with mixed marginal distributions. *Journal of Flood Risk Management*, 2(1), 32–44. <https://doi.org/10.1111/j.1753-318X.2009.01020.x>

Karmakar, Subhankar, Simonovic, S. P., Peck, A., & Black, J. (2010). An Information

System for Risk-Vulnerability Assessment to Flood. *Journal of Geographic Information System, 02(03)*, 129–146. <https://doi.org/10.4236/jgis.2010.23020>

Khaladkar, R. M., Narkhedkar, S. G., Road, H. B., & Pune, P. (2007). *Performance of NCMRWF Models in Predicting High Rainfall Spells During SW Monsoon Season - A Study for Some Cases in July 2004 Indian Institute of Tropical Meteorology.*

Kok, G., Gottlieb, N. H., Peters, G. J. Y., Mullen, P. D., Parcel, G. S., Ruiter, R. A. C., Fernández, M. E., Markham, C., & Bartholomew, L. K. (2016). A taxonomy of behaviour change methods: an Intervention Mapping approach. *Health Psychology Review, 10(3)*, 297–312. <https://doi.org/10.1080/17437199.2015.1077155>

KPDNA. (2018). *Kerala Post Disaster Needs Assessment Floods and Landslides - August 2018 (Issue October).*

Kron, W. (2005). Flood risk = hazard • values • vulnerability. *Water International, 30(1)*, 58–68. <https://doi.org/10.1080/02508060508691837>

Kumar, A., Dudhia, J., Rotunno, R., Niyogi, D., & Mohanty, U. C. (2008). Analysis of the 26 July 2005 heavy rain event over Mumbai, India using the Weather Research and Forecasting (WRF) model. *Quarterly Journal Of The Royal Meteorological Society, 134(July 2005)*, 1897–1910. <https://doi.org/10.1002/qj.325>

Maraun, D. (2016). Bias Correcting Climate Change Simulations - a Critical Review. In *Current Climate Change Reports, 2(4)*, 211–220. Springer. <https://doi.org/10.1007/s40641-016-0050-x>

Martz, L. W., & Garbrecht, J. (1999). An outlet breaching algorithm for the treatment of closed depressions in a raster DEM. *Computers and Geosciences, 25(7)*, 835–844. [https://doi.org/10.1016/S0098-3004\(99\)00018-7](https://doi.org/10.1016/S0098-3004(99)00018-7)

- Mitchem, J. D. (2004). Place vulnerability to tornadoes in the United States: A multi-scale assessment. In *Hazards & Vulnerability Research Institute. Columbia, University of South Carolina*. <https://search.proquest.com/docview/305116578?pq-origsite=gscholar>
- Mukesh, S., Komal, C., & Alexander, K. (2017). Earth Science & Climatic Change Land Use / Cover and Vulnerability Mapping Through Remote Sensing and. *J. Earth Sci Clim Change*, 8(1), 1–6. <https://doi.org/10.4172/2157-7617.1000380>
- Narasimhan, B., Bhallamudi, S. M., Mondal, A., & Majumdar, P. (2016). Chennai Floods 2015 A Rapid Assessment. *Report, May*.
- National Disaster Management Authority. (2008). *National Disaster Management Guidelines Management of Floods*.
- Nelsen, R. B. (1999). *An Introduction to Copulas*, 139. Springer New York. <https://doi.org/10.1007/978-1-4757-3076-0>
- Neuendorf, K. K. E., Mehl, J.P., J., & Jackson, J.A., E. (2005). Glossary of Geology. *Glossary of Geology: American Geological Institute*, 799 .
- Nobre, A. D., Cuartas, L. A., Hodnett, M., Rennó, C. D., Rodrigues, G., Silveira, A., Waterloo, M., & Saleska, S. (2011). Height Above the Nearest Drainage - a hydrologically relevant new terrain model. *Journal of Hydrology*, 404(1–2), 13–29. <https://doi.org/10.1016/j.jhydrol.2011.03.051>
- Nobre, Antonio Donato, Cuartas, L. A., Momo, M. R., Severo, D. L., Pinheiro, A., & Nobre, C. A. (2016). HAND contour : a new proxy predictor of inundation extent. *Hydrological Processes*, 333(August 2015), 320–333. <https://doi.org/10.1002/hyp.10581>
- Oki, T., & Kanae, S. (2006). Global Hydrological Cycles and World Water Resources. *Science*, 313(5790), 1068 – 1072. <https://doi.org/10.1126/science.1128845>

- Pai, D. S., Sridhar, L., Badwaik, M. R., & Rajeevan, M. (2015). Analysis of the daily rainfall events over India using a new long period (1901–2010) high resolution ($0.25^\circ \times 0.25^\circ$) gridded rainfall data set. *Climate Dynamics*, 45(3–4), 755–776. <https://doi.org/10.1007/s00382-014-2307-1>
- Poulin, A., Huard, D., Favre, A. C., & Pugin, S. (2007). Importance of tail dependence in bivariate frequency analysis. *Journal of Hydrologic Engineering*, 12(4), 394–403. [https://doi.org/10.1061/\(ASCE\)1084-0699\(2007\)12:4\(394\)](https://doi.org/10.1061/(ASCE)1084-0699(2007)12:4(394))
- Preethi, B., Mujumdar, M., Kripalani, R. H., Prabhu, A., & Krishnan, R. (2017). Recent trends and tele-connections among South and East Asian summer monsoons in a warming environment. *Climate Dynamics*, 48(7–8), 2489–2505. <https://doi.org/10.1007/s00382-016-3218-0>
- Rahmati, O., Kornejady, A., Samadi, M., Nobre, A. D., & Melesse, A. M. (2018). Development of an automated GIS tool for reproducing the HAND terrain model. *Environmental Modelling and Software*, 102, 1–12. <https://doi.org/10.1016/j.envsoft.2018.01.004>
- Rennó, C. D., Nobre, A. D., Cuartas, L. A., Soares, J. V., Hodnett, M. G., Tomasella, J., & Waterloo, M. J. (2008). HAND, a new terrain descriptor using SRTM-DEM: Mapping terra-firme rainforest environments in Amazonia. *Remote Sensing of Environment*, 112(9), 3469–3481. <https://doi.org/10.1016/j.rse.2008.03.018>
- Rieger, W. (1998). A phenomenon-based approach to upslope contributing area and depressions in DEMs. *Hydrological Processes*, 12(6), 857–872. [https://doi.org/10.1002/\(SICI\)1099-1085\(199805\)12:6<857::AID-HYP659>3.0.CO;2-B](https://doi.org/10.1002/(SICI)1099-1085(199805)12:6<857::AID-HYP659>3.0.CO;2-B)
- Roxy, M. K., Ghosh, S., Pathak, A., Athulya, R., Mujumdar, M., Murtugudde, R., Terray, P., & Rajeevan, M. (2017). A threefold rise in widespread extreme rain events over central

- India. *Nature Communications*, 8(1), 1–11. <https://doi.org/10.1038/s41467-017-00744-9>
- Roy, P. S., Meiyappan, P., Joshi, P. K., Kale, M. P., Srivastav, V. K., Srivasatava, S. K., Behera, m. D., Roy, A., Sharma, Y., Ramachandran, R. M., Bhavani, P., Jain, A. K., & Krishnamurthy, Y. V. N. (2016). Decadal Land Use and Land Cover Classifications across India, 1985, 1995, 2005. *ORNL Distributed Active Archive Center*. <https://doi.org/10.3334/ORNLDAAC/1336>
- Rygel, L., O’Sullivan, D., & Yarnal, B. (2006). A method for constructing a social vulnerability index: An application to hurricane storm surges in a developed country. *Mitigation and Adaptation Strategies for Global Change*, 11(3), 741–764. <https://doi.org/10.1007/s11027-006-0265-6>
- Sajjad, M., Lin, N., & Chan, J. C. L. (2020). Spatial heterogeneities of current and future hurricane flood risk along the U.S. Atlantic and Gulf coasts. *Science of the Total Environment*, 713, 136704. <https://doi.org/10.1016/j.scitotenv.2020.136704>
- Seenirajan, M., Natarajan, M., Thangaraj, R., & Bagyaraj, M. (2017). Study and Analysis of Chennai Flood 2015 Using GIS and Multicriteria Technique. *Journal of Geographic Information System*, 09(02), 126–140. <https://doi.org/10.4236/jgis.2017.92009>
- Selvam, A. M. (2011). Signatures of universal characteristics of fractal fluctuations in global mean monthly temperature anomalies. *Journal of Systems Science and Complexity*, 24(1), 14–38. <https://doi.org/10.1007/s11424-011-9020-5>
- Seneviratne, S., Nicholls, N., Easterling, D., Goodess, C., Kanae, S., Kossin, J., Luo, Y., Marengo, J., McInnes, K., Rahimi, M., Reichstein, M., Sorteberg, A., Vera, C., & Zhang, X. (2012). Changes in climate extremes and their impacts on the natural physical environment. In: *Managing the Risk of Extreme Events and Disasters to Advance*

Climate Change Adaptation.[Field, C.B., V. Barros, T.F. Stocker, D. Qin, D.J. Dokken, K.L. Ebi, M.D. Mastra. *Managing the Risk of Extreme Events and Disasters to Advance Climate Change Adaptation.*[Field, C.B., V. Barros, T.F. Stocker, D. Qin, D.J. Dokken, K.L. Ebi, M.D. Mastrandrea, K.J. Mach, G.-K. Plattner, S.K. Allen, M. Tignor, and P.M. Midgley (Eds.)]. *A Sp*, 109–230. <https://doi.org/10.2134/jeq2008.0015br>

Sharma, T., Vittal, H., Karmakar, S., & Ghosh, S. (2020). Increasing agricultural risk to hydro-climatic extremes in India. *Environmental Research Letters*, 15(3), 034010.

Shastri, H., Ghosh, S., & Karmakar, S. (2017). Improving Global Forecast System of extreme precipitation events with regional statistical model: Application of quantile-based probabilistic forecasts Hiteshri. *Journal of Geophysical Research : Atmospheres*, 122, 1617–1634. <https://doi.org/10.1002/2016JD025489>

Sherly, M. A., Karmakar, S., Parthasarathy, D., Chan, T., & Rau, C. (2015). Disaster Vulnerability Mapping for a Densely Populated Coastal Urban Area: An Application to Mumbai, India. *Annals of the Association of American Geographers*, 105(6), 1198–1220. <https://doi.org/10.1080/00045608.2015.1072792>

Sietz, D., Lüdeke, M. K. B., & Walther, C. (2011). Categorisation of typical vulnerability patterns in global drylands. *Global Environmental Change*, 21(2), 431–440. <https://doi.org/10.1016/j.gloenvcha.2010.11.005>

Tarboron, G. (1997). A new method for the determination of flow directions and upslope areas in grid digital elevation models. *Water Resources Research*, 33(2), 309–319. <https://doi.org/10.1029/96WR03137>

Turner, B. L., Kasperson, R. E., Matson, P. A., McCarthy, J. J., Corell, R. W., Christensen, L., Eckley, N., Kasperson, J. X., Luers, A., Martello, M. L., Polsky, C., Pulsipher, A., & Schiller, A. (2003). A framework for vulnerability analysis in sustainability science.

Proceedings of the National Academy of Sciences, 100(14), 8074–8079.
<https://doi.org/10.1073/pnas.1231335100>

Uitto, J. I. (1998). The geography of disaster vulnerability in megacities: a theoretical framework. *Applied Geography*, 18(1), 7–16. [https://doi.org/10.1016/S0143-6228\(97\)00041-6](https://doi.org/10.1016/S0143-6228(97)00041-6)

Vittal, H., Karmakar, S., Ghosh, S., & Murtugudde, R. (2020). A comprehensive India-wide social vulnerability analysis: highlighting its influence on hydro-climatic risk. *Environmental Research Letters*, 15(1), 014005. <https://doi.org/10.1088/1748-9326/ab6499>

Wei, Y. M., Fan, Y., Lu, C., & Tsai, H. T. (2004). The assessment of vulnerability to natural disasters in China by using the DEA method. *Environmental Impact Assessment Review*, 24(4), 427–439. <https://doi.org/10.1016/j.eiar.2003.12.003>

Wu, S. Y., Yarnal, B., & Fisher, A. (2002). Vulnerability of coastal communities to sea-level rise: a case study of Cape May County, New Jersey, USA. *Climate Research*, 22(3), 255-270. doi:10.3354/cr022255

Zhang, L., & Singh, V. P. (2006). Bivariate flood frequency analysis using the copula method. *Journal of Hydrologic Engineering*, 11(2), 150–164. [https://doi.org/10.1061/\(ASCE\)1084-0699\(2006\)11:2\(150\)](https://doi.org/10.1061/(ASCE)1084-0699(2006)11:2(150))

Figure Captions

Figure. 1 (a) Map of Kerala, India showing districts and neighboring states. (b) Grids of weather forecast model, GEFS, superposed over the study region.

Figure. 2: (a) Observed daily rainfall from Satellite (TRMM) on 15th August 2018 (Hatched areas show the most affected districts as per the loss report). (b) District wise total crop, transport, aquaculture, health and heritage losses (in million USD) from KPDNA, 2018 report

Figure 3: (a-f) Bias corrected GEFS forecasted daily rainfall on lead days 15(1st Aug), 10(6th Aug), 5(11th Aug), 3(13th Aug), 2(14th Aug) and 1(15th Aug) for 15th August 2018

Figure. 4: Flowchart of the proposed methodology for event specific flood risk forecast at weather to medium range

Figure. 5: Flowchart of the methodology for calculating socioeconomic vulnerability

Figure 6: Flowchart of the methodology for calculating topographic vulnerability in terms of HAND values

Figure 7: (a) Demonstration of a 3×3 box of 9 grids considered (grey) with the grid of interest at the center (red) for the forecasted rainfall used to calculate hazard. (b1 - b6) Hazard forecasted for each grid considering the maximum rainfall in the neighboring 3×3 box of 9 grids with the grid of interest at the center. The hatched area shows the most affected districts (in terms of losses) by the flood.

Figure 8: subdistrict wise (a) socioeconomic vulnerability; (b) topographic vulnerability; and (c) exposure. The hatched area shows the most affected districts in terms of losses due to the flood.

Figure 9. Subdistrict wise relative flood risk map at a lead time (a) 15 days; (b) 10 days; (c) 5 days; (d) 3 days; (e) 2 days and (f) 1 day. The hatched area shows the most affected districts in the flood.

Journal Pre-proof

Tables

Table 1: List of socioeconomic vulnerability indicators and justification of each indicator

Positive Indicators	Justifications
Total number of houses	Increases potential loss
Total population	More difficult evacuation and more deaths
Total SC-ST population	Categorized as socially backward community by the government
Percentage of house with power	The electronic equipment increases the loss
Negative Indicators	
Main working population	More resilience against hazard
Female literates	The more the literate female the more they can save their families during hazard
Amenities	Helps during disaster management and emergency operations
Percentage of house with water	Resources to help during hazard

Table 2. DI values for different LULC

Type	DI values	Category
Deciduous broadleaf forest	0.3	Forest
Crop	0.8	Resources
Built-in	0.85	Urban area
Mixed Forest	0.3	Forest
Shrub land	0.3	Open area
Barren	0.4	Sparse crop land (Causes more loss than open area)
Fallow	0.4	No crop farm land (Causes more loss than open area)
Waste	0.2	Waste (Open area with less sensitivity to loss)
Water Bodies	0.1	Water body
Plantation	0.8	Resources
Aquaculture	0.8	Resources
Mangrove	0.3	Forest
Salt pan	0.8	Resources
Grass	0.4	Open area
Evergreen broadleaf forest	0.3	Forest
Deciduous needle leaf forest	0.3	Forest
Permanent wetland	0.5	
Snow and ice	0	

Evergreen needle leaf forest	0.3	Forest
------------------------------	-----	--------

Figures

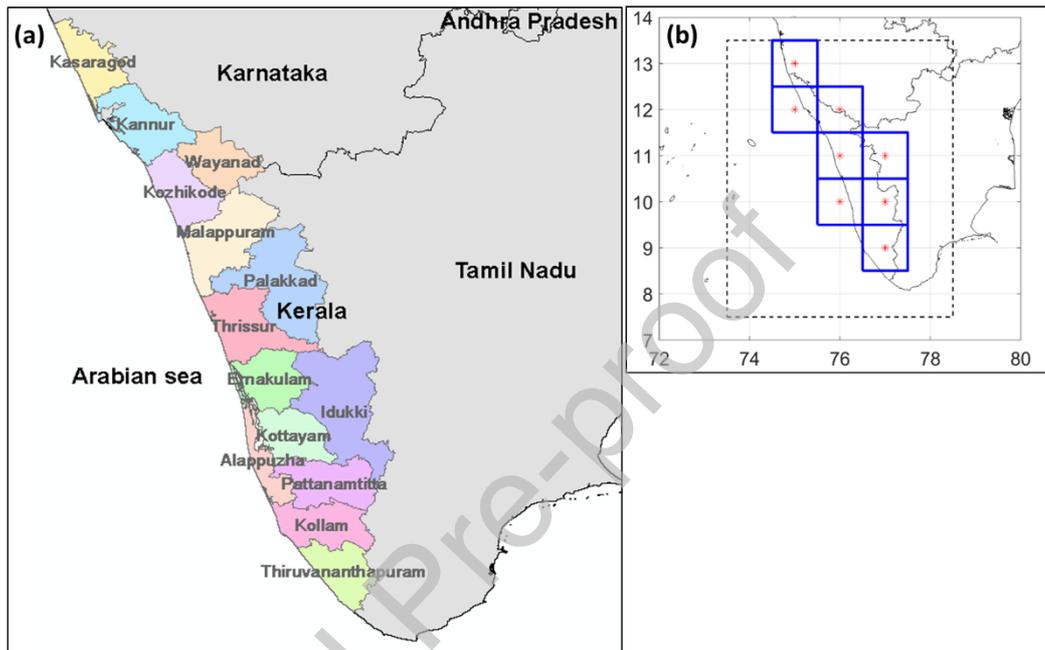


Figure. 1 (a) Map of Kerala, India showing districts and neighboring states. (b) Grids of weather forecast model, GEFS, superposed over the study region.

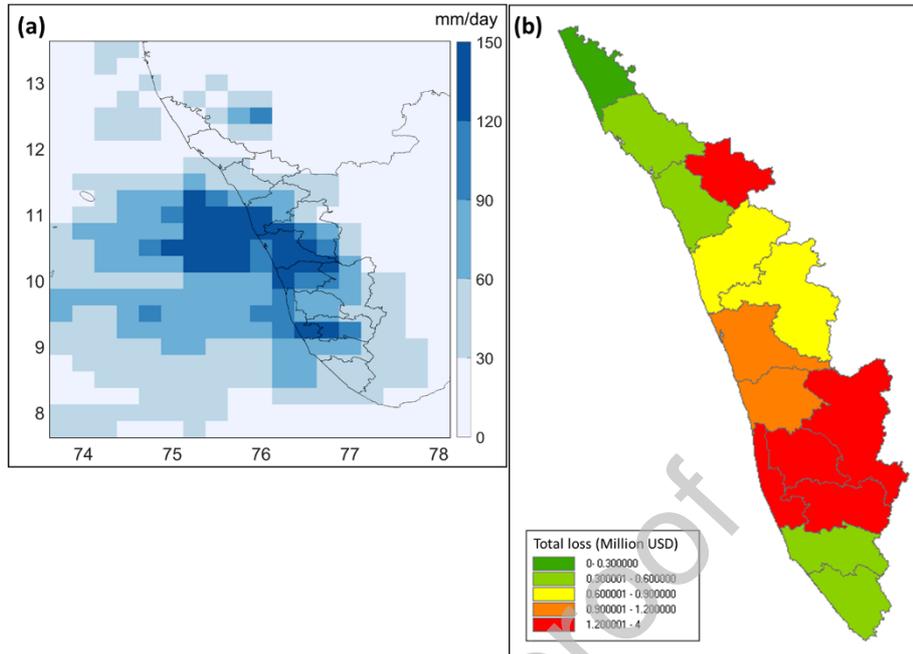


Figure. 2: (a) Observed daily rainfall from Satellite (TRMM) on 15th August 2018 (Hatched areas show the most affected districts as per the loss report). (b) District wise total crop, transport, aquaculture, health and heritage losses (in million USD) from KPDNA, 2018 report

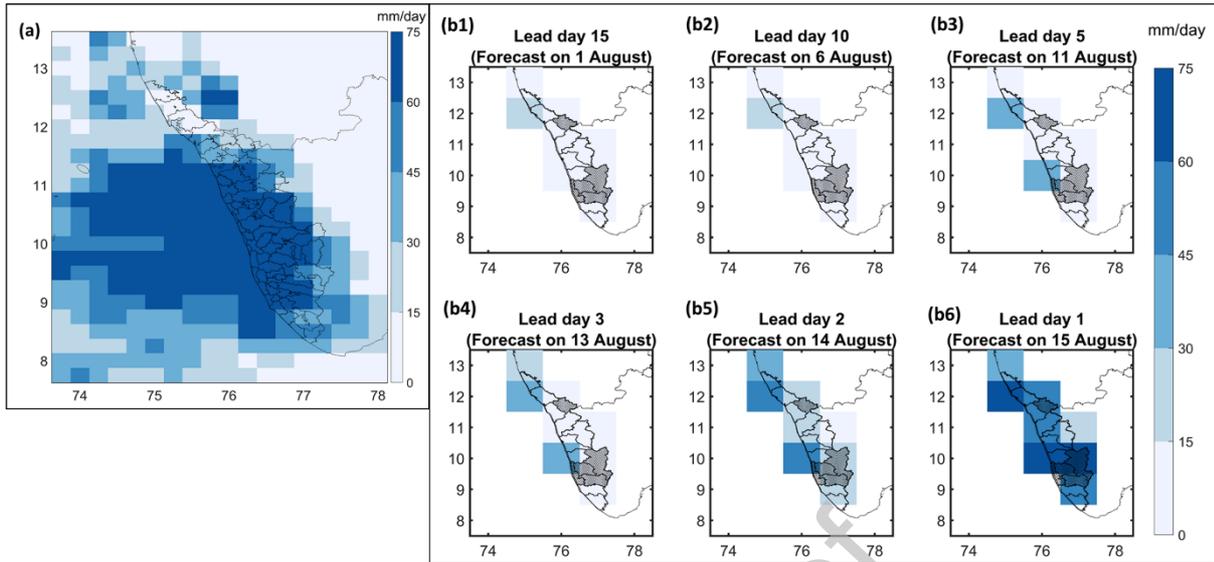


Figure 3: (a-f) Bias corrected GEFS forecasted daily rainfall on lead days 15(1st Aug), 10(6th Aug), 5(11th Aug), 3(13th Aug), 2(14th Aug) and 1(15th Aug) for 15th August 2018

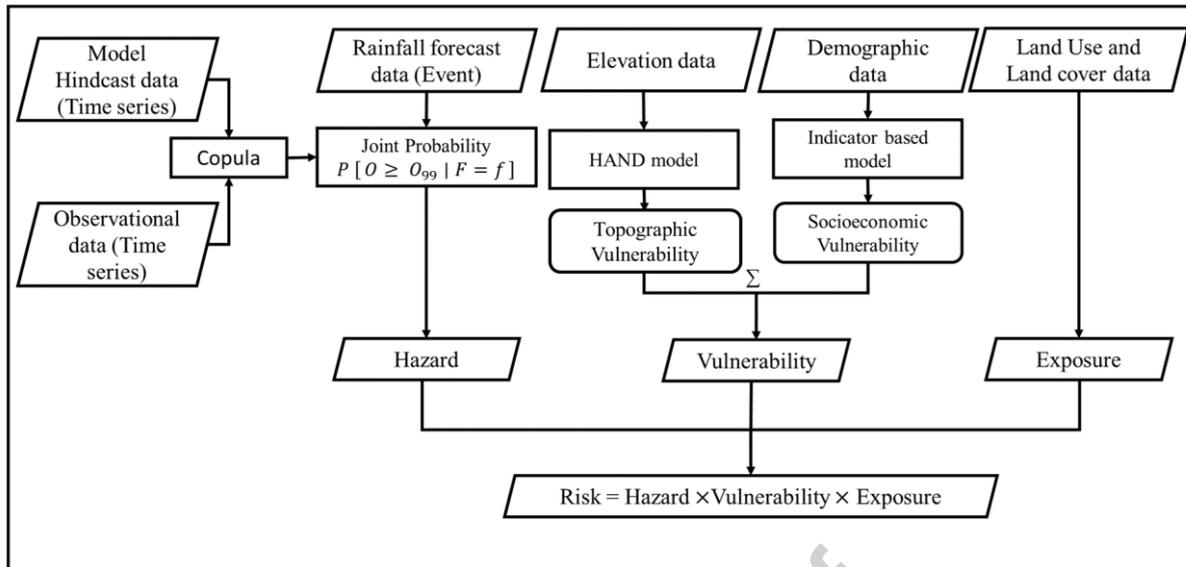


Figure. 4: Flowchart of the proposed methodology for event specific flood risk forecast at weather to medium range

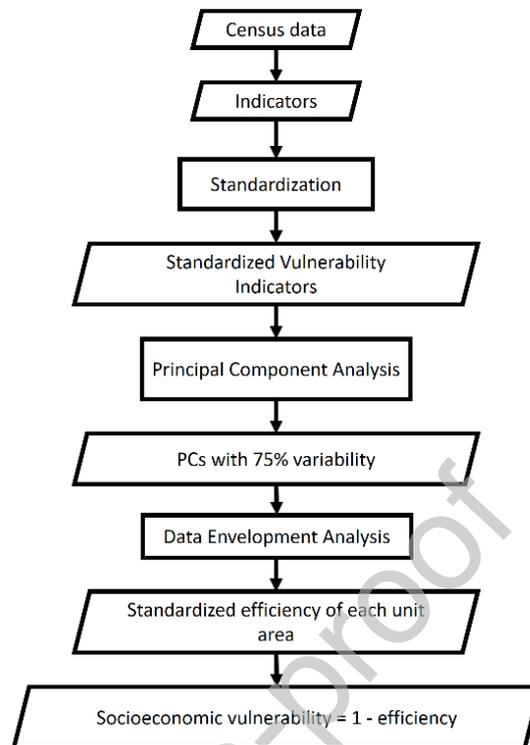


Figure. 5: Flowchart of the methodology for calculating socioeconomic vulnerability

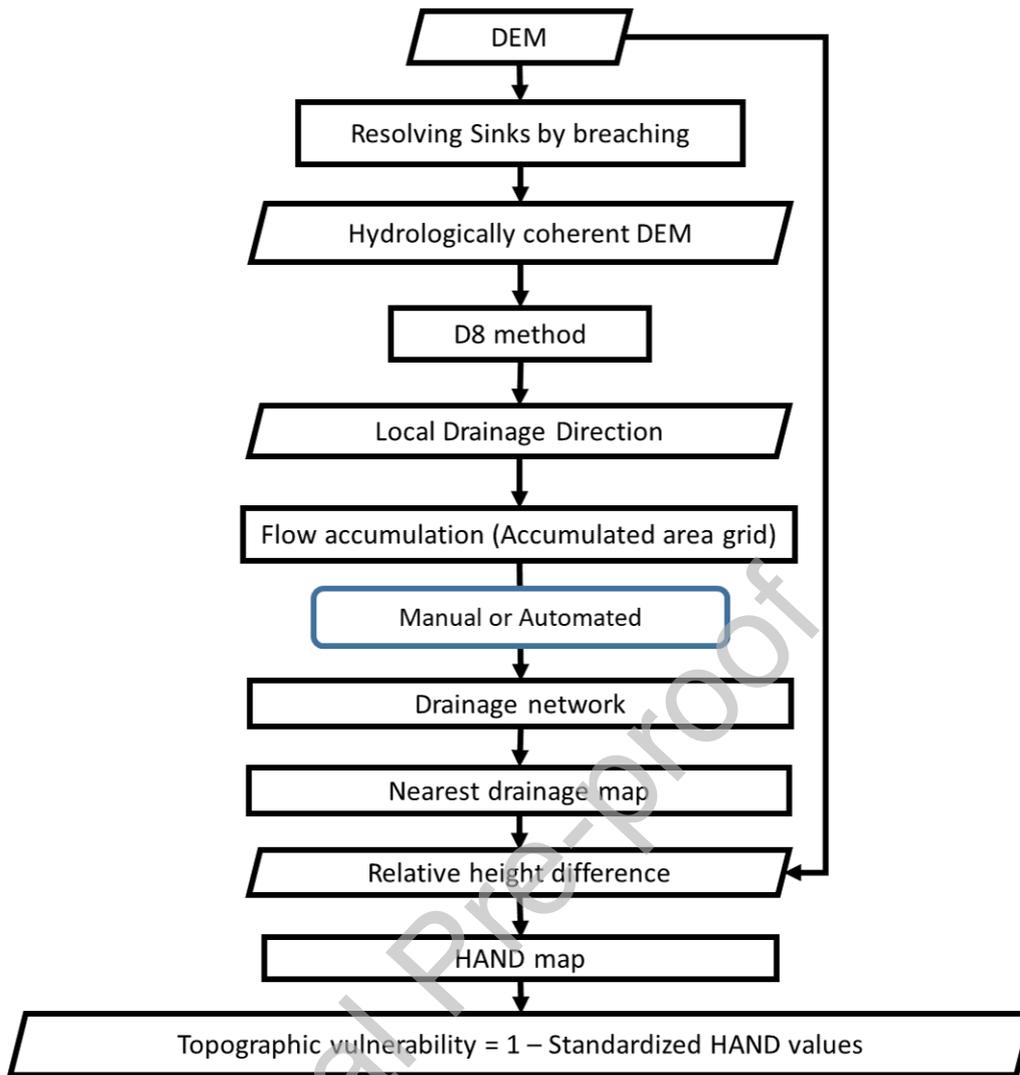


Figure 6: Flowchart of the methodology for calculating topographic vulnerability in terms of HAND values

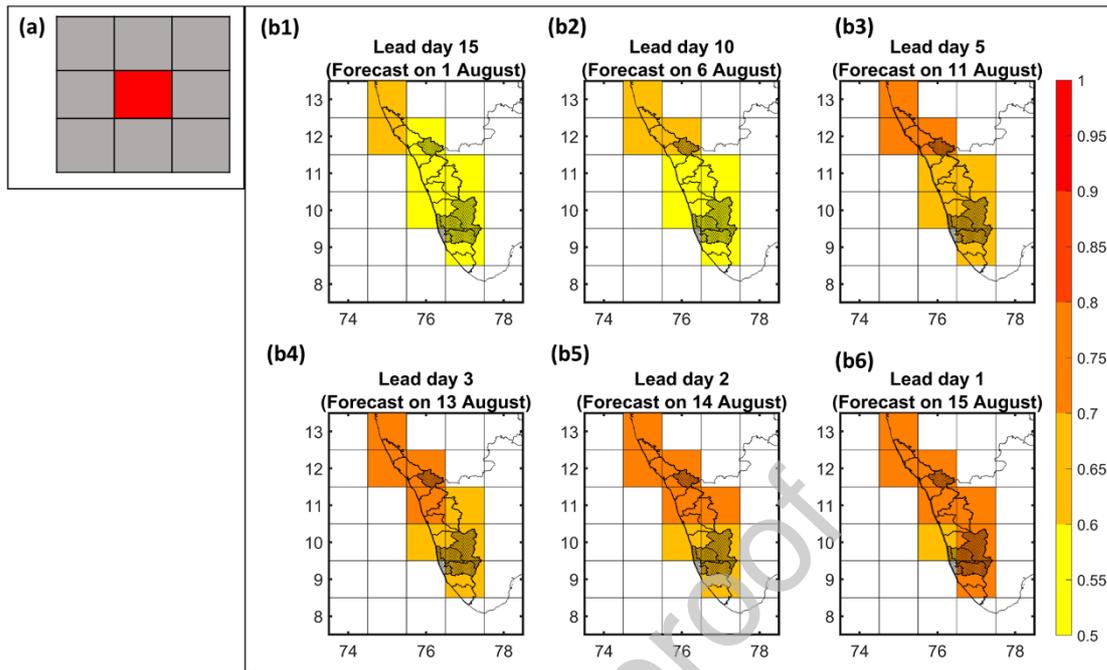


Figure 7: (a) Demonstration of a 3×3 box of 9 grids considered (grey) with the grid of interest at the center (red) for the forecasted rainfall used to calculate hazard. (b1 - b6) Hazard forecasted for each grid considering the maximum rainfall in the neighboring 3×3 box of 9 grids with the grid of interest at the center. The hatched area shows the most affected districts (in terms of losses) by the flood.

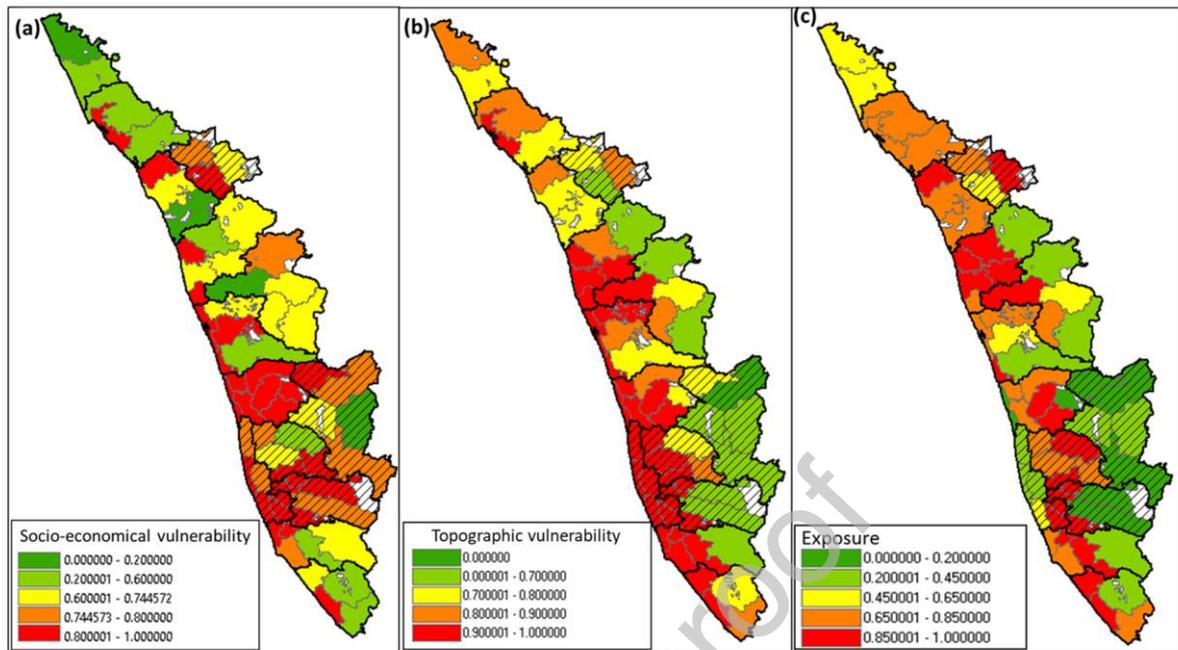


Figure 8: subdistrict wise (a) socioeconomic vulnerability; (b) topographic vulnerability; and (c) exposure. The hatched area shows the most affected districts in terms of losses due to the flood.

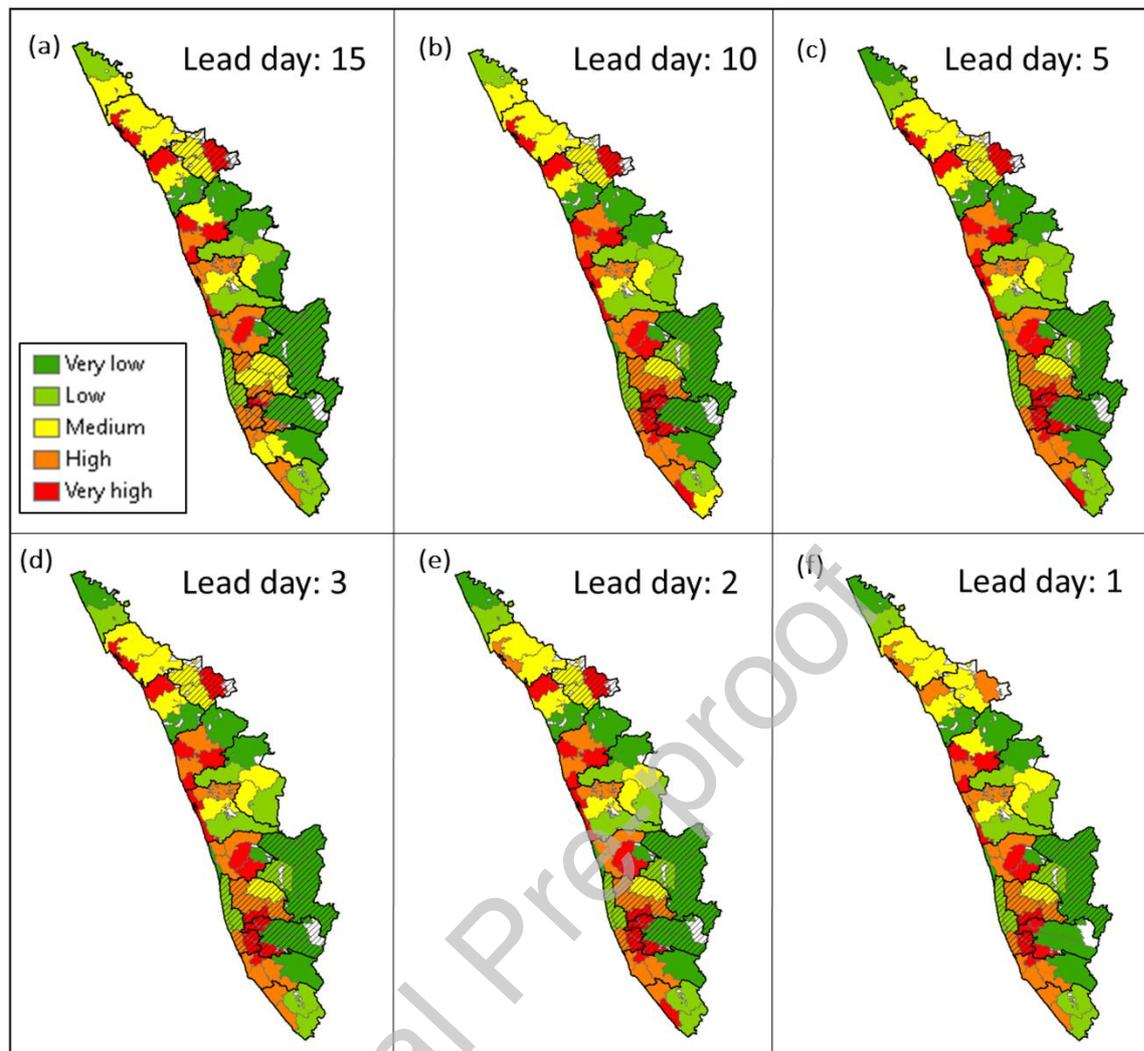
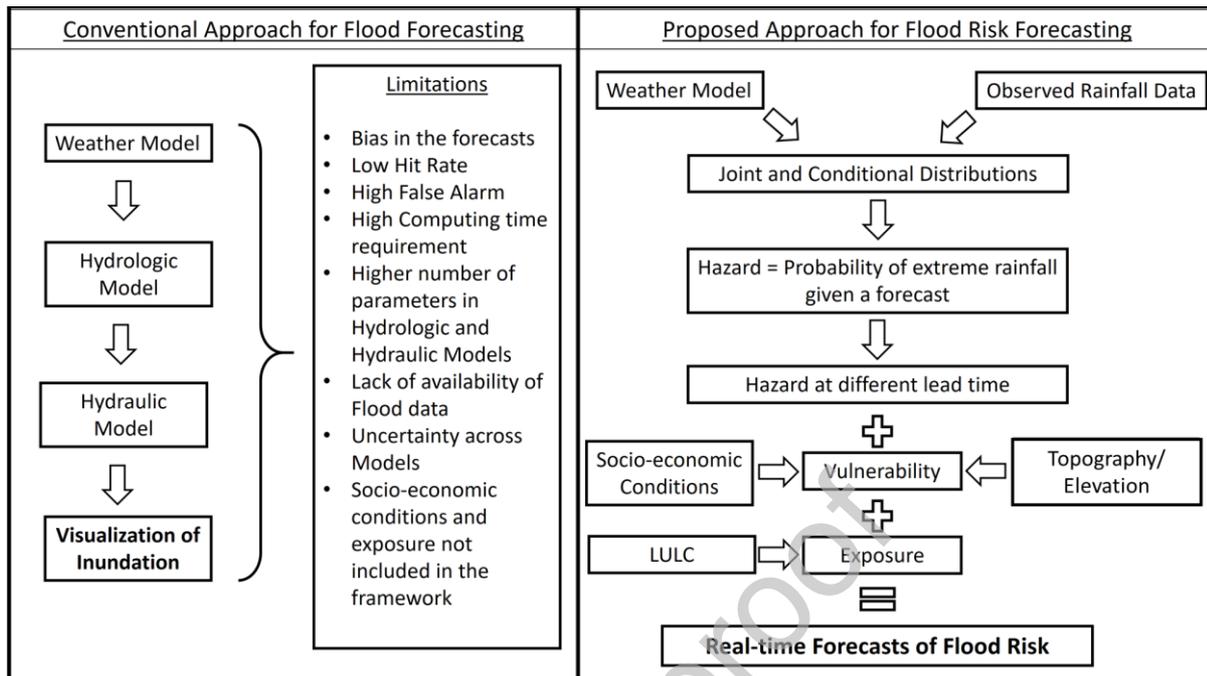


Figure 9. Subdistrict wise relative flood risk map at a lead time (a) 15 days; (b) 10 days; (c) 5 days; (d) 3 days; (e) 2 days and (f) 1 day. The hatched area shows the most affected districts in the flood.

Graphical



Journal Pre-proof