



OPEN

# A network percolation-based contagion model of flood propagation and recession in urban road networks

Chao Fan<sup>1</sup>✉, Xiangqi Jiang<sup>2</sup> & Ali Mostafavi<sup>1</sup>✉

In this study, we propose a contagion model as a simple and powerful mathematical approach for predicting the spatial spread and temporal evolution of the onset and recession of floodwaters in urban road networks. A network of urban roads resilient to flooding events is essential for the provision of public services and for emergency response. The spread of floodwaters in urban networks is a complex spatial–temporal phenomenon. This study presents a mathematical contagion model to describe the spatial–temporal spread and recession process of floodwaters in urban road networks. The evolution of floods within networks can be captured based on three macroscopic characteristics—flood propagation rate ( $\beta$ ), flood incubation rate ( $\alpha$ ), and recovery rate ( $\mu$ )—in a system of ordinary differential equations analogous to the Susceptible-Exposed-Infected-Recovered (SEIR) model. We integrated the flood contagion model with the network percolation process in which the probability of flooding of a road segment depends on the degree to which the nearby road segments are flooded. The application of the proposed model is verified using high-resolution historical data of road flooding in Harris County during Hurricane Harvey in 2017. The results show that the model can monitor and predict the fraction of flooded roads over time. Additionally, the proposed model can achieve 90% precision and recall for the spatial spread of the flooded roads at the majority of tested time intervals. The findings suggest that the proposed mathematical contagion model offers great potential to support emergency managers, public officials, citizens, first responders, and other decision-makers for flood forecast in road networks.

Given the essential role transportation plays in emergency response, provision of essential services, and maintenance of economic well-being<sup>1</sup>, the resilience of urban road networks to natural hazards, especially flooding events, has received increasing attention<sup>2, 3</sup>. Floodwaters in urban networks propagate over time and space, inducing a great deal of spatial–temporal uncertainty vis-a-vis protective actions, such as evacuation, and rapid emergency response<sup>4</sup>. Developing effective prediction tools to forecast the characteristics of flooding events is critical to the enhancement of urban road network resilience<sup>5</sup>.

Multiple studies have explored the spatial–temporal properties of floods in urban networks, including impact evaluation of environmental stress<sup>6–8</sup> and cascading effects in road networks<sup>9, 10</sup>. In particular, empirical studies adopting remote sensors<sup>11</sup>, hydraulic data<sup>12</sup>, or satellite images<sup>13</sup> have attempted to capture the properties of urban flooding. Temporal evolution of flood status is driven by the time-dependent profile of environmental stress, such as the duration of rainfall in hurricanes<sup>12</sup>. This temporal information facilitates identification of the outbreak and inflection points for flooding in affected networks. Flooding also exhibits high spatial correlation<sup>14</sup> in which the co-located road segments are more likely to be flooded in the immediately succeeding time increments, indexed to digital timestamps on the model<sup>15</sup>. Specifically, a hydrologic study has shown that proximity to flooded areas is a significant predictor of regional flood frequency based on regionalized flood quantiles for 575 Austrian catchments<sup>16</sup>. While empirical studies illustrated the complex spatial–temporal dynamics of floods, their capabilities for flood prediction often rely on various types of hydro-geomorphological monitoring datasets and

<sup>1</sup>Zachry Department of Civil and Environmental Engineering, Texas A&M University, College Station, TX 77843, USA. <sup>2</sup>Department of Computer Science and Engineering, Texas A&M University, College Station, TX 77843, USA. ✉email: chfan@tamu.edu; amostafavi@civil.tamu.edu

intensive computation<sup>17</sup>. Due to delay and computational cost issues, the existing physics-based hydrodynamic models may not be conducive to providing timely and reliable predictions for the spatial–temporal spread of floods and the failure of road segments within short time periods<sup>18</sup>.

To overcome the limitations in empirical models, machine learning techniques have been proposed and tested for predicting the spread of floodwaters in urban areas<sup>19, 20</sup>. Compared to the hydrodynamic methods, machine learning models, such as multiple linear regressions<sup>21</sup>, deep neural networks<sup>22</sup>, and Bayesian forecasting models<sup>23</sup> require fewer input parameters, so that the models can be easily trained on historical flood event data. For example, Khosravi et al. tested four decision tree-based machine learning models—logistic model trees, reduced error pruning trees, naïve Bayes trees, and alternating decision trees—for flood mapping<sup>24</sup>. The results show that, with adequate training, these models can achieve greater than 80% accuracy for predicting the flooded locations. Youssef et al. integrated frequency ratio and logistic regression models to evaluate the correlation between flood occurrence and various potential factors and developed a model providing an acceptable prediction accuracy<sup>14</sup>. Although existing machine learning models can achieve good predictive performance to capture the flood propagation in urban networks, these models are limited due to their dependence on large sets of historical data for model training. In addition, existing machine learning models are designed to capture only the propagation of flood in urban areas. The flood recession process, which is also important for assessing the resilience of urban networks, is often ignored by the existing machine learning models.

Recognizing the limitations of existing models, there is a real need for mathematical models that can capture the spatial–temporal evolution of flooding without relying on a variety of input parameters and historical data such as the volume of waters and the width of the roads. Recent studies have demonstrated a surprisingly significant similarity among spreading processes in different systems, including the spread of traffic congestion in transportation, the contagion of infectious disease in populations, the diffusion of ideas in social networks<sup>25</sup>, as well as the evolution of flooding in urban road networks<sup>26, 27</sup>. Motivated by these studies, our goal in this research was to describe the floodwater spreading process using generalized mathematical contagion models, such as classical epidemic models<sup>28</sup>. Existing epidemic models offer an analytical and numerical framework to quantify and forecast multiple spread phenomena in a variety of contexts. In particular, the popular susceptible-infectious-recovered (SIR) model created the basic building blocks of epidemic modeling using infectious and recovery rates. These mathematical models have two fundamental hypotheses: compartmentalization, in which each entity is associated with a state or compartment; and homogenous mixing, in which each entity has the same chance of contacting an inflected entity<sup>29</sup>. In the context of flooding, each road cell is associated with a state, functional-flow or flooded. Hence, the flooding propagation problem fits this compartmentalization hypothesis. While these hypotheses simplify the modeling of contagion by eliminating the need to know the structure of the networks, the mathematical models can still capture the temporal evolution of the fraction of infected entities in the networks very well.

Flood risk prediction is a task that should take into account both the temporal and spatial natures of the floodwater in road networks. Urban flood risk characterization requires not only knowledge of the fraction of flooded roads at each timestamp, but also needs to identify the geographic locations of flooded roads as flooding unfolds. Hence, pure mathematical models are not able to satisfy these requirements. To this end, the network percolation process has gained attention recently because it enables to capture of the propagation process through the topological connectivity in networks<sup>30</sup>. As defined in percolation theory, the spread of infection relies on the probability of the infected neighbors, in which the heterogenous mixing assumption is held in local components of the networks<sup>31</sup>. Specifically, the infection spreads from an initial node along edges of the percolated network<sup>32</sup>. Hence, the percolation process reflects the “amplification” effect of neighbors and weakens global network interactions. An infection is more likely to be transmitted to those the node comes encounters. This characteristic is essential for flood propagation prediction in urban networks since it considers the spatial co-location and constraints of urban networks. Without the temporal information about the fraction of flooded roads, however, the percolation process would fail to capture the temporal evolution of flooding in urban networks.

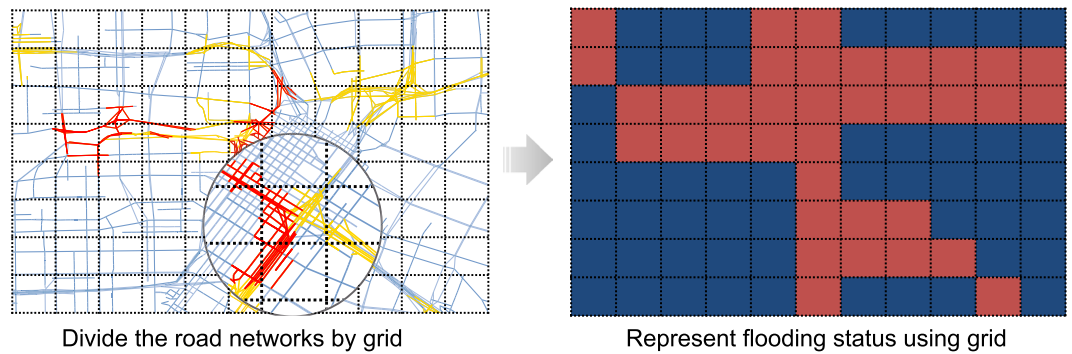
This study proposed and tested a network percolation-based contagion model that integrates the mathematical framework and network percolation process to predict the spatial propagation and temporal evolution of flooding in urban road networks. The mathematical framework fits the temporal dynamics of the flood situations, and the percolation process identifies locations of flooded road segments. To illustrate the performance of the model, we applied it to a case study of flood evolution in Harris County road networks during Hurricane Harvey. Potential control strategies are also identified based on the outcomes of the model in the case study.

## The network percolation-based contagion model

The proposed model is composed of four components: road network modeling, flood spread characterization, flood percolation process, and model evaluation. This epidemic-like model of flood spread process in urban road networks considers both global dynamics of flood scales and local probability of affecting other co-located roads within a set of neighbors.

**Road network modeling.** Road networks contain hundreds of thousands of road segments, most of which are quite short, about 100–800 m each. The traffic status information collected at the road segment level provides sufficient spatial resolution to precisely estimate the scale of flooding in road networks.

**Definition 1.** A **road segment** is a basic unit with a starting point and an ending point, which can be assembled into a whole road in the order of the points.



**Figure 1.** A schema of converting flooding status from a road network to grid.

Each point is associated with a longitude and a latitude so that it can be located on a geographical map. Then, a road segment can be represented as  $(lat_s, lng_s, lat_e, lng_e)$ , where  $lat_s$  and  $lng_s$  are the latitude and longitude of the starting point, while  $lat_e$  and  $lng_e$  are the latitude and longitude of the ending point.

Although road segments enable good resolution for understanding the situation in urban networks, flooded areas are usually not restricted in a banded road segment. Floodwaters tend to start from a point and spread in all directions. In addition, massive segments and their complex connections in the networks would also cause intensive computational cost. Hence, in modeling road networks, segment-to-segment modeling is limited to follow the nature of flood spread and achieve efficient computation. To this end, grid decomposition, a commonly used method<sup>33</sup>, is adopted in this study to generate equal-sized cells and divide the study area into small regions (see Fig. 1).

**Definition 2.** A road cell is a square over a rectangular projection of the geographical map.

The spatial boundary of the cell can be represented as a set of geo-coordinates  $(lat_{bl}, lng_{bl}, lat_{ur}, lng_{ur})$ , where  $lat_{bl}$  and  $lng_{bl}$  are the latitude and longitude of the bottom left corner, while  $lat_{ur}$  and  $lng_{ur}$  are the latitude and longitude of the upper right corner. To convert the road segments to grid, we apply the following criteria to the road network:

$$s_i \in grid_j, \text{ if } lat_{bl} \leq \frac{lat_s + lat_e}{2} \leq lat_{ur} \text{ and } lng_{bl} \leq \frac{lng_s + lng_e}{2} \leq lng_{ur} \quad (1)$$

After grid decomposition, the road network assembled from road segments could be described as a series of cells. In practice, multiple factors would influence the outcome of grid decomposition. For example, a large cell would include many segments, which might lead to losing spatial resolution. Accordingly, the model would further lose the capability of capturing the spatial spread of flood. On the other hand, a small grid that cannot cover at least one road segment will partition the network into discontinuous components, and subsequently increase the computational cost. Hence, grid decomposition requires pre-testing to ensure that a cell in the grid is able to maintain spatial resolution without unduly burdening computation.

Once the road segments are assigned to cells, we remove cells lacking segments to reduce computational cost. The remaining cells form a network in which the cells are considered as nodes, and their shared borderlines are considered as links. By doing so, we can construct an undirected grid network  $\mathcal{G}$  with average degree of  $k$  to represent the topology of the road networks. Average degree of a road cell is the average number of adjacent cells per cell in the road network. To model the flood propagation and process, we then associate a dynamical binary state variable  $x$  to each of the  $N$  cells (also called nodes) of the grid network  $\mathcal{G}$ , such that  $x_i(t) \in \{0, 1\}$  represents the flood status of node  $i$  at time  $t$ . Using a standard notation, we divide the cells into two classes, functional flow ( $F$ ) and flooded ( $C$ ), corresponding respectively to the values 0 and 1 of the status variable  $x$ . Functional flow is a state in which traffic can utilize a road (regardless of traffic level). In the context of flooding spread process, the status  $C$  represents the cells which have been flooded. At each time  $t$ , the macroscopic flooding situation is given by the fraction of flooded cells  $c(t) = \frac{1}{N} \sum_{i=1}^N x_i(t)$ .

**Flood spread characterization.** The flood propagation and recession process are temporally and spatially variant. To capture the temporal nature of flood evolution in urban road networks, the proposed model considers macroscopic characteristics to predict the temporal evolution of floodwater spread in urban networks.

In the first step, we define four flooding statuses for a cell: functional flow, exposed, flooded, and recovered, statuses by including the temporal attributes.  $C(t)$  represents the number of flooded cells in the network at time  $t$ ;  $F(t)$  represents the number of functional flow cells at time  $t$ ;  $E(t)$  represents the number of cells that are in flood incubation stage (i.e., roads in the path of approaching floodwater but on which traffic still moves) at time  $t$ ; and  $R(t)$  represents the number of cells that have recovered from flooding at time  $t$ . Given a grid network  $\mathcal{G}$  with an average degree of  $k$  and  $N$  nodes, each cell in the network has functional flow at time  $t = 0$ . That is,  $F(t = 0) = N$  and  $C(t = 0) = 0$ . The flood initially occurs at a set of nodes and then propagates throughout the network. From

a macroscopic perspective, a cell in the undirected network is on average connected to  $k$  other cells. The neighbors of a flooded cell are exposed to flood at a rate of  $\beta$ . In modeling the temporal evolution, the connections of the cells are assumed to be homogeneous, which forms the basis to formulate a general differential equation system. Then, the probability of a flooded cell being connected to a functional-flow link is  $F(t)/N$  at time  $t$ . Therefore, a flooded cell comes into contact with  $kF(t)/N$ . Since  $C(t)$  flooded cells have water flowing, each at a rate  $\beta$ , and the exposed cells become flooded at a fixed rate  $\alpha$ , the average number of new exposed cells  $dE(t)$  during a unit timeframe  $dt$  is determined as follows:

$$\frac{dE(t)}{dt} = \beta k \frac{C(t)(N - C(t) - E(t) - R(t))}{N} - \alpha E(t) \quad (2)$$

The exposed cells are not completely inundated, and traffic can still pass through, but they could be flooded within the next few timestamps. We call this stage, flood incubation stage. The defined four statuses of the road cells are the only statuses that a road cell could have. Hence, the sum of the four status variables should be  $N$ . That is,  $N = C(t) + E(t) + F(t) + R(t)$ . To capture the fraction of the cells in each status, we use the following variables:  $c(t)$  to represent the fraction of flooded cells in the grid network at time  $t$ ;  $f(t)$  to represent the fraction of functional-flow cells in which all road segments are accessible at time  $t$ ;  $r(t)$  to represent the fraction of recovered cells from flooding at time  $t$ ; and  $e(t)$  to represent the fraction of cells that are exposed to flooding but still in flood incubation at time  $t$ . Then, the differential equation for the change rate of exposed cells can be derived as follows:

$$\frac{de(t)}{dt} = \beta k c(t)(1 - c(t) - e(t) - r(t)) - \alpha e(t) \quad (3)$$

where, the product of  $\beta k$  is called transmission rate or transmissibility. The transmissibility can be used to measure the capability of floodwater to spread in an urban network under the same propagation rate. This also allows us to understand the effects of the topological structure of an urban network on the transmission of floodwaters.

Since a fraction of the functional-flow cells become exposed at each timestamp, the decreasing rate of the fraction of functional-flow cells can be represented by:

$$\frac{df(t)}{dt} = -\beta k c(t)(1 - c(t) - e(t) - r(t)) \quad (4)$$

Simultaneously, in a unit timeframe, a fraction of exposed nodes would be flooded at a rate  $\alpha$ , and some of the flooded cells recover at a rate  $\mu$ . Hence, the changing rate of the fraction of flooded cells and the fraction of recovered cells can be formulated as:

$$\frac{dc(t)}{dt} = -\mu c(t) + \alpha e(t) \quad (5)$$

$$\frac{dr(t)}{dt} = \mu c(t) \quad (6)$$

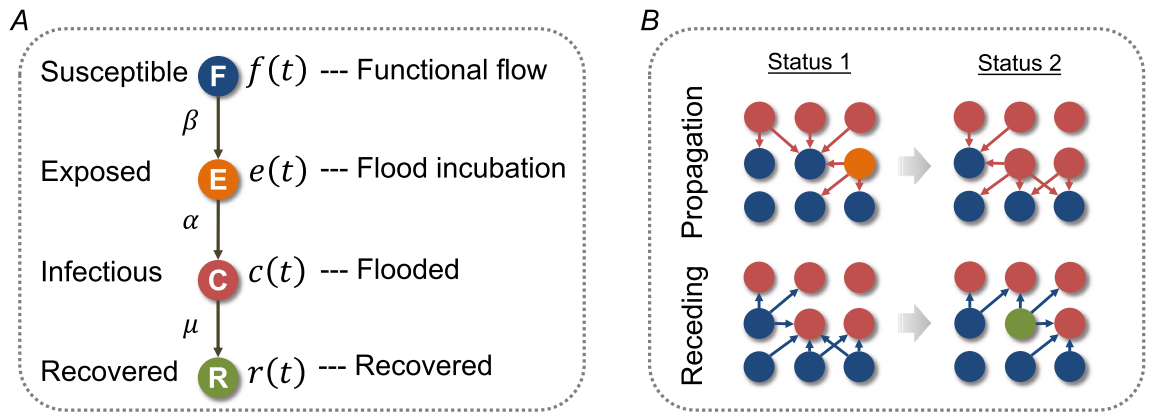
Evidently, in a large-scale network where  $N$  is large, the probability of a flooded cell being connected to a functional-flow cell could be close to zero. At the macroscopic scale, the assumption of homogeneous mixing makes the prediction more tractable and robust. It should also be noted that the model does not include mortality (i.e., significantly damaged roads that would not be functional after the floodwater recedes). Since it is not often the case that a flooded road cell is severely damaged, excluding the mortality is reasonable and realistic. Based on the former constructs, the model component derived for capturing the temporal dynamics of flooding scale using macroscopic characteristics is established (Fig. 2A).

**Network percolation process.** The spatial nature of floodwater spread in urban networks is modeled using a network percolation process. The network percolation process describes the contagion effects of a flooded cell on its network neighbors<sup>34</sup>. Specifically, the cells whose neighbors are flooded are more likely to be flooded than the cells whose neighbors are not flooded<sup>29</sup>. Similarly, floodwater is more likely to recede first from the cells whose neighbors are not flooded compared to the flooded cells whose neighbors are also flooded. To characterize this spatial aspect of flood spread, we define the probability of a node to be flooded or to be recovered in the next timestamp based on the number of flooded neighbors. The propagation and recession processes are modeled as described below, respectively (see Fig. 2B).

In the propagation process, the percolating cluster is a set of road cells which are flooded next to another bounded by functional-flow cells at a specific time interval. Since the extreme event, Hurricane Harvey in this study, affected a large area, multiple percolating clusters presented in the road network. They were spatially scattered across the network at the early stage of flooding. As the event progressed, some clusters broke through the bound and formed larger percolating clusters. We can obtain the number of flooded cells  $N_c^{(t)}$  at a unit timestamp  $t$  by the predicted fraction of flooded nodes  $c(t)$  and the total number of cells in the networks. The calculation can be formulated as:

$$N_c^{(t)} = N \cdot c(t) \quad (7)$$

The flooded cells at the unit timestamp  $t$  are composed of flooded cells  $N_c^{(t-1)}$  at the last timestamp, plus the additional flooded cells  $N_c^{(t)}$  at timestamp  $t$ , excluding the recovered cells. In reality, however, the fraction



**Figure 2.** The network percolation-based contagion model of flood propagation and recession in urban road networks. **(A)** The contagion model for capturing the rate of flooding cell; and **(B)** a schema of network percolation-based flood propagation and receding process.

of recovered cells is negligible during the propagation period before reaching the flooding peak. Hence, in our model, the number of additional flooded cells,  $N_p^{(t)}$ , in the current timestamp  $t$  is obtained by:

$$N_p^{(t)} = N_c^{(t)} - N_c^{(t-1)} \tag{8}$$

As discussed earlier, the spatial pattern of the propagation process ( $F \rightarrow E \rightarrow C$ ) is controlled by the fraction of flooded neighbors of a cell. Hence, we assign the probability of flooding (i.e., the fraction of flooded cells among all neighbors) to the unflooded cells. Each unflooded cell will be assigned a probability  $p_i^{(t)} \in \{p_1^{(t)}, p_2^{(t)}, \dots, p_k^{(t)}\}$ , where  $k \leq N$ . Then the cells are sorted based on their probabilities of flooding from high to low. The additional flooding cells at the timestamp  $t$  are identified from the cells with high probabilities of flooding, subject to the number of additional flooded cells,  $N_p^{(t)}$ . Hence, the percolation threshold is selected based on the number of additional flooded cells predicted by the proposed contagion model and the sorted probabilities of flooding to other cells.

Like the propagation process, flood recession ( $C \rightarrow R$ ) can also be modeled in a way similar to the network percolation process. The percolating cluster is the set of road cells with functional flow next to another bounded by flooded cells. We first calculate the number of recovered cells  $N_r^{(t)}$  at the timestamp  $t$  based on the value of  $r(t)$  obtained from the flood dynamics model:

$$N_r^{(t)} = N \cdot r(t) \tag{9}$$

Floodwater starts receding after the peak of flooding; in actual experience, additional flooded cells usually do not occur. Hence, in the spatial prediction, the number of flooded cells  $N_c^{(t)}$  at the current timestamp  $t$  is equal to the number of flooded cells  $N_c^{(t-1)}$  at the last timestamp  $t - 1$ , minus the number of recovered cells  $N_r^{(t)}$  at the current timestamp  $t$ . The calculation can be formulated as:

$$N_c^{(t)} = N_c^{(t-1)} - N_r^{(t)} \tag{10}$$

In the next step, we assign the probabilities of flooding (i.e., the fraction of flooded cells among all neighbors) to the flooded cells. Each flooded cell is assigned a probability  $p_i^{(t)} \in \{p_1^{(t)}, p_2^{(t)}, \dots, p_k^{(t)}\}$ . That is, the probability of floodwater receding is  $1 - p_i^{(t)}$ . In a manner different from the propagation process (in which the cells are sorted based on their probabilities of flooding from high to low), we sort the cells based on their probabilities of flooding from low to high (i.e., probabilities of floodwater receding from high to low). The recovered cells at timestamp  $t$  are assigned a low probability of flooding.

Using the above calibration in the percolation process, we can mitigate the homogeneous mixing assumptions in the flood spread characterization model by adopting local heterogeneous flood probabilities and achieve high accuracy in predicting the spatial distribution of flooded cells in urban networks.

**Model evaluation.** We employed two metrics to evaluate the performance of the model for predicting the temporal evolution and spatial propagation of flooding spread in urban networks. The first component of the model is a system of differential equations to capture the magnitude of flooded cells in networks. The objective of this component is analogous to addressing a fitting curve. Hence, we use the root mean square error (RMSE)<sup>35</sup> to measure the error of the model:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \tag{11}$$

where,  $\hat{y}_i$  is the predicted value,  $y_i$  is the observed value, and  $n$  is the number of observations. Ignoring the division by  $n$  under the square root, the formula can be considered as a formula for Euclidean distance between the vectors of predicted values and observed values. Hence, the RMSE is a normalized distance between the predicted outcomes and the observations which can be used to evaluate the model accuracy. This can also serve as a heuristic for a training model, which will be used in a pattern search algorithm to obtain the numerical solution for the model (see “Results” section).

The spatial nature of the flooding propagation and receding is captured by the outcomes of the flood spread characterization model and the network percolation process. The intersection between the set of predicted flooded road cells and the set of observed flooded road cells indicates the precision and recall of the model<sup>36</sup>. They are formulated as follows:

$$\text{Precision} = \frac{\text{Truepositive}}{\text{Truepositive} + \text{Falsepositive}} \quad (12)$$

$$\text{Recall} = \frac{\text{Truepositive}}{\text{Truepositive} + \text{Falsenegative}} \quad (13)$$

where *true positive* is an outcome where the model correctly predicts the flooded class, *false positive* is an outcome where the model incorrectly predicts the flooded class, and *false negative* is an outcome where the model incorrectly predicts the unflooded class. The calculated precision and recall allow us to assess the performance of the model by identifying the specific correctly predicted road cells.

## Results

**Study context and data collection.** To illustrate the application and performance of the proposed network percolation-based contagion model for flood spread, we tested the model using high-resolution data related to flooded roads in Harris County during Hurricane Harvey in 2017. Hurricane Harvey was a Category 4 storm that made landfall in Houston (Harris County) on August 26, 2017, dissipated inland August 30, 2017<sup>37</sup>. The torrential rainfall brought by Harvey caused intensive flooding in Harris County, where the floodwaters damaged more than 290 roads and highways<sup>38</sup>. The flooding occurred August 27 through on September 4, 2017. We collected the traffic data for 19,712 roads in Harris County from INRIX, a private company providing location-based data and analytics. The INRIX traffic data covers all available road roads—from interstates to intersections, country roads to neighborhoods. The data includes the average speed for each road segment in 5-min intervals. The timeframe of data is August 1 to September 30, covering the entire flooding period. In this case study, we used squares with a length of 400 m for generating the grid network.

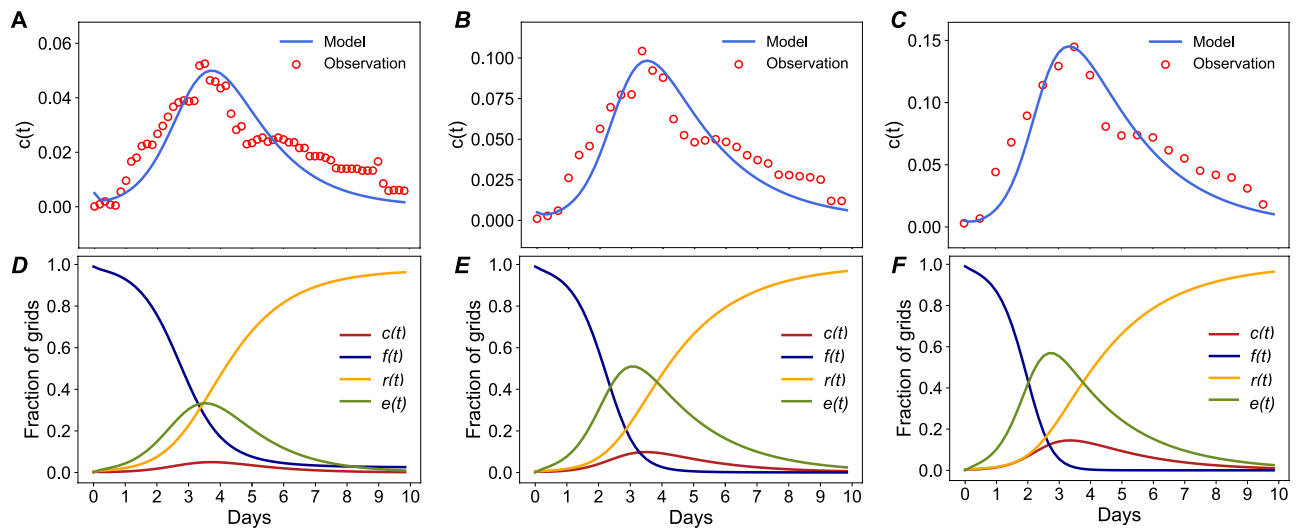
In the INRIX dataset, the flooded road segments can be identified by a designation of NULL for average speed, meaning there was no vehicle driving through the segment. Comparing the data before and after Hurricane Harvey, we found that the NULL average speed appears only during the flooding period. By cross-checking the flooded roads on government reports, most of the roads with NULL speed were flooded at that time. Although the average speed data was collected in 5-min intervals, the flood situation did not evolve significantly in such a short time period. To better capture the flood propagation and receding process, we aggregated the data at 4-, 8-, and 12-h intervals to test the performance of the model. Each interval was assigned a binary value of flooding status to the road segments based on their average speed. As documented in the model section, the value would be 0 if there is no NULL speed record, and the value would be 1 if there is a NULL speed record in the dataset (indicating functional flow status for the roads).

**Pattern search for parameter estimation.** It is often the case that the analytical solution to the differential equation system cannot be generated. That is because the functions are usually not continuous or differentiable<sup>39</sup>. To efficiently estimate the parameters in the proposed model, we applied a global pattern search algorithm<sup>40</sup> as a derivative-free numerical optimization method to fit the curves for each variable (i.e.,  $f(t)$ ,  $e(t)$ ,  $c(t)$ , and  $r(t)$ )<sup>26</sup>. The objective function in this optimization process is the RMSE, which will be minimized by searching for the optimal propagation rate  $\beta$ , recovery rate  $\mu$ , and exposed rate  $\alpha$ . To start with this algorithm, we first specified initial values for the three parameters. The change of the parameters could be either increased or decreased in each step. The algorithm would compute the RMSE for each step until it finds an optimal point at which the current RMSE is smaller than the previous one. If the algorithm cannot find the optimal point using the current step size, it will use half of the current step size and repeat the computing process. Then all three parameters will move to the optimal values. The algorithm runs iteratively until one of the stopping criteria is met: the maximum number of iterations is reached or the step size is smaller than a certain threshold. In this study, we set the maximum number of iterations to be 2,000, and the threshold for the step size is 0.0001. The initial values for models with time intervals are the same,  $\beta = 1$ ,  $\alpha = 1$ , and  $\mu = 0$ . Through our tests, the selection of values around these initial values would yield similar results for the parameters.

Table 1 shows the estimated values for the parameters and the optimized RMSE for three models. The RMSE increases a little bit with the increase of the length of the time intervals. That is because the fraction of road cells that are flooded (i.e.,  $c(t)$ ) is greater in longer time intervals than in shorter time intervals. But, all three models fit the flood spread patterns in road cells well since the RMSE value accounts for only a very small proportion of the peak value of  $c(t)$ . In addition, the propagation rate  $\beta$  remains the same across three models. This result indicates that the length of the time interval does not affect the capability of the model to predict flooding propagation. The flood incubation rate  $\alpha$  and recovery rate  $\mu$ , however, decrease with the increase in the length of the time interval. That is because the number of flooded road cells increases with the increase in time intervals. Hence,

Length of time interval	RMSE	$\beta$	$\alpha$	$\mu$	$\beta/\mu$
4 h	0.020	0.9998	0.1267	0.8348	1.1977
8 h	0.024	0.9998	0.0830	0.4073	2.4547
12 h	0.027	0.9998	0.0780	0.2702	3.7002

**Table 1.** Estimated values for the parameters in the contagion model.



**Figure 3.** Four-state model of flood propagation and recession in the road network, analogous to the susceptible-exposed-infected-recovered model subject to a time-varying disaster profile. Day 0 is August 26, 2017; and Day 9 is September 4, 2017. (A,D) are the results from the model for 4-h time interval; (B,E) are the results from the model for 8-h time interval; and (C,F) are the results from the model for 12-h time interval.

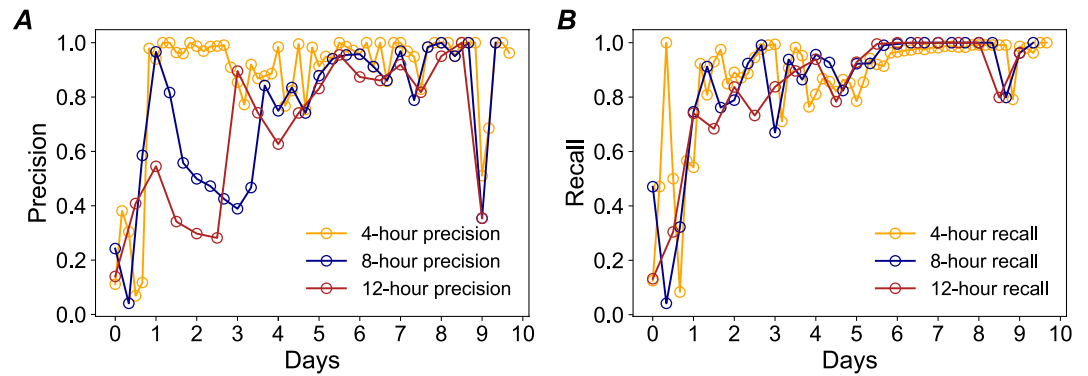
a great number of road cells would be in the flood incubation state, which leads to a low rate of  $\alpha$  (Fig. 3D–F). As such, the recovery rate must be small to represent a large number of flooded road cells. This result reveals that the length of the time interval has an influence on parameter estimation, but it does not significantly affect the fitting performance.

**Prediction results.** Once the optimal values for the parameters are obtained, we can capture the temporal evolution of flood propagation and recession in urban network by showing the spread curves for each variable (i.e.,  $f(t)$ ,  $e(t)$ ,  $c(t)$ , and  $r(t)$ ) (Fig. 3D–F).

Basic reproduction number  $R_0 = \beta/\mu$  is used as a measure of the number of secondary flooded cells generated by the first flooded cell over the course of the flooding unfolding<sup>41</sup>. Mathematically, when  $R_0$  is smaller than 1, the propagation will not occur as the recovery rate is greater than the propagation rate<sup>42</sup>. In the Hurricane Harvey case study, the estimated  $R_0$  is greater than 1 across all three models and increases with the length of time interval (see Table 1). This result explains how rapidly floodwaters spread. The closer the value of  $R_0$  to 1, the more stable the flooding situation is. In this case, we can observe that the flood situation would change more slightly if the time interval is shorter than 4 h, since  $R_0$  would be close to 1. When the time interval increases from 4 to 8 h, the value of  $R_0$  jumps from 1.20 to 2.45, meaning that in every 4- to 8-h increment, the flood situation would change more drastically.

In addition, the changes in  $c(t)$  allow us to predict the extent of flood in the networks and critical timestamps for breakout and peak (see Fig. 3A–C). The figures show the flood spread characterization in Eqs. (3)–(6) are fit to the empirical flooding data very well. We can observe that at the beginning of the hurricane (the hourly timestamps in day 1 and day 2), the number of flooded cells grew exponentially between every 4-, 8-, or 12-h increment, and reached the peak in the middle of day 3. After that timestamp on day 3, the flood receded gradually. Using the results, we can estimate the time that the urban road network needs to recover from flooding. In this case, evidently, the recovery rate was slow, which led to a long period for recovery at some severely flooded locations.

In the next step, we examined the predictive performance for both spatial spread and temporal evolution for the flooding, shown in Fig. 4. All models perform very well in terms of predicting the specific locations of flooded segments based on the flooding data at the last timestamp. The best performance of these models appears during the peak and receding period, while the performance is not particularly good at the beginning of the flooding. That is because the locations of the initial flooded cells depend on the rainfall magnitude and spatial patterns of precipitation in different areas. The emergence of initial flooded cells is difficult to be identified by the model without additional information about the precipitation pattern. After a few hours, however, the flood propagation



**Figure 4.** Prediction performance of the proposed model for flood propagation and receding. **(A)** Precision of the model for different time intervals; and **(B)** Recall of the model for different time intervals. Day 0 is August 26, 2017; and Day 9 is September 4, 2017.

and recession closely follow the percolation process since the majority of the initial flooded cells were identified correctly at that time. Hence, the model could predict the spatial spread of flooding after the locations of initial flooding are identified. In terms of the precision metric, the best model is the model based on a 4-h time interval, and with the increase of the length of the time interval, the performance of the model decreases a little bit. In terms of the recall metric, these three models show similar performance and show promise for special and temporal floodwater characteristics, especially during the peak and recession periods.

The three sets of figures in Fig. 5 show the true positive and false positive results for the road segments in the network. Here we converted the cells back to their segments. All segments in the flooded cells are considered flooded segments. We plotted three figures for each model at different timestamps: beginning, peak, and recession periods. As we observed in the figure, the flooding initially occurred at different locations in the timestamp of the beginning period (Fig. 5, left panel). With the continuous rainfall, the floodwater propagated from the initial flooded road segments to their neighbors (middle panel of Fig. 5). After Hurricane Harvey dissipated on August 30, the flood started to recede from the road segments at the edge of the flooded area (Fig. 5, right panel). The figures show that more than 90% of the flooded segments are captured by the proposed model, and the predictive performance decreases a bit with the increase of the length of the time interval. These findings support that the proposed model can accurately predict the spatial and temporal characteristics of flooding spread in urban networks.

**Adaptation of the model.** The proposed network percolation-based contagion model could be used to model flood spread in other regions and flooding events. Different extreme events have different attributes such as varying durations, intensities and geographical scales, which result in varying numbers and durations of flooded roads. To demonstrate the adaptation of the model in other contexts, we conducted further controlled experiments to examine the adaptation of the model to different topological properties of urban networks (average degree), the propagation rate which reflects the stress of rainfall, and flood incubation rate and recovery rate. We change the value of one parameter that was learned from the INRIX data and control other parameters to be constant as shown in Table 1. The results show the extent to which different parameters could affect the predicted outcomes. This experiment also provides evidence for supporting the generalizability of the proposed contagion model for different cities and various intensities of flooding.

First, we tested the effect of average degree of an urban network on the flood dynamics by keeping the same values for other parameters. Figure 6A shows the changes in model results for predicting the flood status in 4-h time intervals. With the increase of the average degree  $k$  of the networks, the growth rate of the flooded cells increases significantly during the propagation phase. That is because the network with a large  $k$  will allow the flooded road cells to infect more neighbors, which expedites the spread of flood even when the propagation rate remains the same. In addition, a large  $k$  would decrease the time to reach the peak of flooding since the flooded cell would have more connected neighbors. In contrast, a small  $k$  would lead to a longer period of flooding, although the fraction of flooded cells will not reach to the same flooding scale as the urban networks with a large  $k$ .

We further tested the effects of the three macroscopic parameters (i.e.,  $\beta$ ,  $\alpha$ , and  $\mu$ ) on the predicted flood spread (see Fig. 6B–D). On the one hand, similar to the impact of the average degree, the increase in the propagation rate  $\beta$  and the flood incubation rate  $\alpha$  would decrease the time to reach the peak of flooding and the number of flooded cells at the peak, but increase the time to recover to the normal situation. On the other hand, an increase in the recovery rate  $\mu$  significantly decreases the number of flooded cells at the peak, but the time to reach the peak and the time to recover to a normal situation are not changed significantly. The results of these experiments show that by adjusting parameters, the model can be adapted to different scenarios with various intensities of flood events and topological structures of urban networks.





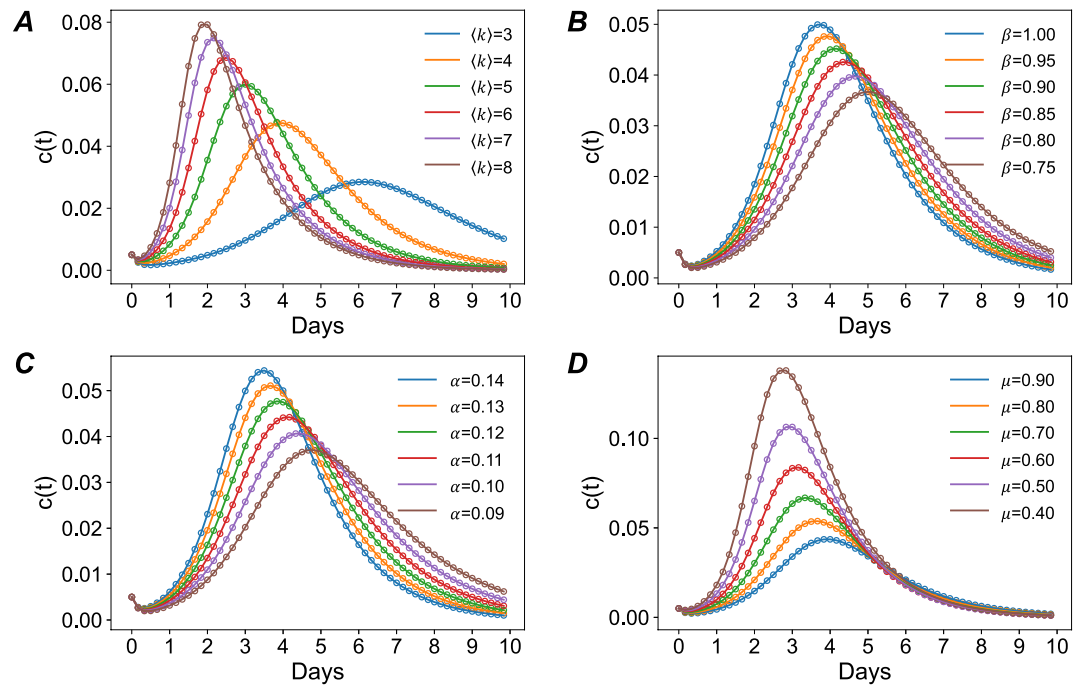
**Figure 5.** Examples of three prediction models for different phases of the flooding period (i.e., propagating, peak, and receding phases). (A) the model for predicting the situation in next 4 h; (B) the model for predicting the situation in next 8 h; and (C) the model for predicting the situation in next 12 h. The true positive (yellow) and false positive (red) road segments are all actual flooded road segments from the empirical INRIX data in a specific time interval.

## Discussion and concluding remarks

We have presented a predictive model which integrates flood spread characterization and network percolation processes to forecast the propagation and recession of flood in urban road networks. The model is formulated as a system of ordinary differential equations relying on three characteristics: flood propagation rate  $\beta$ , flood incubation rate  $\alpha$ , and recovery rate  $\mu$ , analogous to the SEIR model. Using the output of the model, the network percolation process is obtained to model the spatial patterns of the flooded road cells over time. The study showed the application of the proposed model in an empirical case study of urban flooding in Harris County road networks during Hurricane Harvey in 2017. The application of the model using empirical data informs about some key findings and implications for urban flood risk prediction.

First, the extent to which flooding builds up in a network and how fast it recovers are shown to be dependent on the basic reproduction number  $R_0 = \beta/\mu$ , which can help us infer the effective time period in response to the flooding in road networks. In the case study, we found that the basic reproduction number changed significantly when the time interval increased from 4 to 8 h. This result indicates that flood situation evolves slightly in time intervals shorter than 4 h, but, the situation changes dramatically for time intervals that are longer than 4 h. Hence, to better monitor the flood in road networks, the status of roads should be observed in every 4 h. While urban networks in different cities may have different topological properties and experience different rainfall magnitudes, the proposed model can be adapted by adjusting the parameters. Results can inform decision-makers about the spatial propagation and temporal evolution of flooding.

Second, the spatial mechanisms of flood propagation and receding in urban networks is captured based on a network percolation process. This finding highlights the contagion effects of a flooded cell on its neighbors. Since the road cells are spatial networks, different from social contact networks, the spread of flood would be limited by geographical constraints. Hence, the local contagion from a cell to another cell through the link connecting



**Figure 6.** Adaptation of the model to different topological structures of urban road networks (A); different propagation rates (B); different incubation rates (C); and different recovery rates (D), by keeping the other parameters constant.

them would be the main spreading pattern. Practically, this finding provides an important implication for flood control in urban systems. One effective control strategy would be to increase drainage capacity and to build retention ponds around the roads with a high likelihood of flooding. This strategy can reduce the proportion of flooded cells among the unflooded cell neighbors, which can contribute to reducing the probability of flooding in the next timestamp.

Third, the model and its application show good performance in predicting the scale and locations of flooded road cells in disasters. In addition to the response strategies, this model could also support proactive strategies for coping with future flood events. In particular, the proposed model can be incorporated in an early warning system. The system can help officials and the public be aware of the flood situations in the coming hours so that they can perceive flooding risks surrounding them and make proactive preparations. For example, cars and buses drove through floodwater during Hurricane Harvey<sup>43</sup>. With the predicted outcomes of this contagion model, people can be informed about roads at a high risk of flooding in the next few hours. This predictive information could significantly contribute to reducing the economic loss of transportation agencies, and possible loss of life of residents.

Finally, the model we introduced in this paper is simple and robust and can be adapted to various phenomena. Unlike hydrodynamic models and machine learning models that rely on significant data and computational resources, the proposed mathematical model provides a simple but powerful tool for predicting the spatial-temporal evolution of flooding in urban networks. Also, the proposed network percolation-based contagion model can be adapted for modeling other network spread phenomena. Future studies can further investigate the proposed model in other general predictive tasks, such as the spread of traffic congestions in road networks<sup>44</sup>, infectious diseases in human contact networks<sup>45</sup>, and innovations in global communities<sup>46</sup>. This model also has some limitations; for instance, for initial flooded segments without flooded neighbors, it is usually difficult to predict these segments without other information. Future studies can focus on improving the model to precisely predict initial flooded road segments based on rainfall magnitude and capacity of urban drainage systems. In addition, the model considers the parameters such as propagation rate being constant throughout the process, in order to provide a simple but fairly accurate tool for flood prediction. Future studies could extend our model by considering the dynamics of the beta to improve the performance of the model.

### Data availability

The data that support the findings of this study are available from INRIX, but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available.

Received: 7 April 2020; Accepted: 30 July 2020  
Published online: 10 August 2020

## References

- Ganin, A. A. *et al.* Resilience and efficiency in transportation networks. *Sci. Adv.* **3**, e1701079 (2017).
- Koks, E. E. *et al.* A global multi-hazard risk analysis of road and railway infrastructure assets. *Nat. Commun.* **10**, 2677 (2019).
- Paprotny, D., Sebastian, A., Morales-Nápoles, O. & Jonkman, S. N. Trends in flood losses in Europe over the past 150 years. *Nat. Commun.* **9**, 1985 (2018).
- Wang, W., Yang, S., Stanley, H. E. & Gao, J. Local floods induce large-scale abrupt failures of road networks. *Nat. Commun.* **10**, 2114 (2019).
- Fan, C., Zhang, C., Yahja, A. & Mostafavi, A. Disaster city digital twin: a vision for integrating artificial and human intelligence for disaster management. *Int. J. Inf. Manage.* <https://doi.org/10.1016/j.ijinfomgt.2019.102049> (2020).
- Lhomme, S., Serre, D., Diab, Y. & Laganier, R. Analyzing resilience of urban networks: A preliminary step towards more flood resilient cities. *Nat. Hazards Earth Syst. Sci.* **13**, 221–230 (2013).
- Pulcinella, J. A., Winguth, A. M. E., Allen, D. J. & Dasa Gangadhar, N. Analysis of flood vulnerability and transit availability with a changing climate in Harris County, Texas. *Transp. Res. Rec.* <https://doi.org/10.1177/0361198119839346> (2019).
- Serre, D., Barroca, B., Balsells, M. & Becue, V. Contributing to urban resilience to floods with neighbourhood design: The case of Am Sandtorkai/Dalmanndai in Hamburg. *J. Flood Risk Manag.* **11**, S69–S83 (2018).
- Lu, L. *et al.* Vulnerability of interdependent urban infrastructure networks: Equilibrium after failure propagation and cascading impacts. *Comput. Civ. Infrastruct. Eng.* **33**, 300–315 (2018).
- Guan, X. & Chen, C. General methodology for inferring failure-spreading dynamics in networks. *Proc. Natl. Acad. Sci.* <https://doi.org/10.1073/pnas.1722313115> (2018).
- Mousa, M., Zhang, X. & Claudel, C. Flash flood detection in urban cities using ultrasonic and infrared sensors. *IEEE Sens. J.* **16**, 7204–7216 (2016).
- Ramsey, E., Lu, Z., Suzuoki, Y., Rangoonwala, A. & Werle, D. Monitoring duration and extent of storm-surge and flooding in western coastal Louisiana marshes with Envisat ASAR Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **4**, 387–399 (2011).
- Dixon, T. H. *et al.* Subsidence and flooding in New Orleans. *Nature* **441**, 587–588 (2006).
- Youssef, A. M., Pradhan, B. & Sefry, S. A. Flash flood susceptibility assessment in Jeddah city (Kingdom of Saudi Arabia) using bivariate and multivariate statistical models. *Environ. Earth Sci.* **75**, 12 (2015).
- Douglas, E. M., Vogel, R. M. & Kroll, C. N. Trends in floods and low flows in the United States: impact of spatial correlation. *J. Hydrol.* **240**, 90–105 (2000).
- Merz, R. & Blöschl, G. Flood frequency regionalisation—Spatial proximity vs. catchment attributes. *J. Hydrol.* **302**, 283–306 (2005).
- Nayak, P. C., Sudheer, K. P., Rangan, D. M. & Ramasastri, K. S. Short-term flood forecasting with a neurofuzzy model. *Water Resour. Res.* **41**, W04004 (2005).
- Hossain, F., Katiyar, N., Hong, Y. & Wolf, A. The emerging role of satellite rainfall data in improving the hydro-political situation of flood monitoring in the under-developed regions of the world. *Nat. Hazards* **43**, 199–210 (2007).
- Mosavi, A., Ozturk, P. & Chau, K. Flood prediction using machine learning models: Literature review. *Water* **10**, 1536 (2018).
- Fan, C., Wu, F. & Mostafavi, A. A hybrid machine learning pipeline for automated mapping of events and locations from social media in disasters. *IEEE Access* **8**, 10478–10490 (2020).
- Tsakiri, K., Marsellos, A. & Kapetanakis, S. Artificial neural network and multiple linear regression for flood prediction in Mohawk River, New York. *Water* **10**, 1158 (2018).
- Sankaranarayanan, S. *et al.* Flood prediction based on weather parameters using deep learning. *J. Water Clim. Change* <https://doi.org/10.2166/wcc.2019.321> (2019).
- Dong, S., Yu, T., Farahmand, H. & Mostafavi, A. Bayesian modeling of flood control networks for failure cascade characterization and vulnerability assessment. *Comput. Civ. Infrastruct. Eng.* **35**, 668–684 (2020).
- Khosravi, K. *et al.* A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at Haraz watershed, northern Iran. *Sci. Total Environ.* **627**, 744–755 (2018).
- Fan, C., Jiang, Y. & Mostafavi, A. Emergent social cohesion for coping with community disruptions in disasters. *J. R. Soc. Interface* **17**, 20190778 (2020).
- Saberi, M. *et al.* A simple contagion process describes spreading of traffic jams in urban networks. *arXiv Prepr. arXiv:1906.00585* 1–10 (2019).
- Barabási, A.-L. Network science. *Philos. Trans. A. Math. Phys. Eng. Sci.* <https://doi.org/10.1098/rsta.2012.0375> (2013).
- McCluskey, C. C. Complete global stability for an SIR epidemic model with delay—Distributed or discrete. *Nonlinear Anal. Real World Appl.* **11**, 55–59 (2010).
- Barabási, A.-L. & Pósfai, M. *Network Science* (Cambridge University Press, Cambridge, 2016).
- Gao, J., Buldyrev, S. V., Stanley, H. E. & Havlin, S. Networks formed from interdependent networks. *Nat. Phys.* **8**, 40–48 (2012).
- Ball, F., Mollison, D. & Scalia-Tomba, G. Epidemics with two levels of mixing. *Ann. Appl. Probab.* **7**, 46–89 (1997).
- Miller, J. C. Percolation and epidemics in random clustered networks. *Phys. Rev. E* **80**, 20901 (2009).
- Zhou, Z. *et al.* A method for real-time trajectory monitoring to improve taxi service using GPS big data. *Inf. Manage.* **53**, 964–977 (2016).
- Teng, X., Pei, S., Morone, F. & Makse, H. A. Collective influence of multiple spreaders evaluated by tracing real information flow in large-scale social networks. *Sci. Rep.* **6**, 36043 (2016).
- Chai, T. & Draxler, R. R. Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature. *Geosci. Model Dev.* **7**, 1247–1250 (2014).
- Buckland, M. & Gey, F. The relationship between recall and precision. *J. Am. Soc. Inf. Sci.* **45**, 12–19 (1994).
- Sebastian, T. *et al.* Hurricane Harvey Report: A fact-finding effort in the direct aftermath of Hurricane Harvey in the Greater Houston Region. 1–103 (2017).
- Ibrahim, A. Texas officials see long road from Harvey for state transportation network|The Texas Tribune. *The Texas Tribune* 1–1 (2017). <https://www.texastribune.org/2017/09/04/texas-officials-harvey-transportation/>. Accessed 28th Mar 2020.
- Davidon, W. C. Variable metric method for minimization. *SIAM J. Optim.* **1**, 1–17 (1991).
- Raja, M. A. Z. Numerical treatment for boundary value problems of Pantograph functional differential equation using computational intelligence algorithms. *Appl. Soft Comput.* **24**, 806–821 (2014).
- Weitz, J. S. & Dushoff, J. Modeling post-death transmission of Ebola: Challenges for inference and opportunities for control. *Sci. Rep.* **5**, 8751 (2015).
- Chowell, G., Hengartner, N. W., Castillo-Chavez, C., Fenimore, P. W. & Hyman, J. M. The basic reproductive number of Ebola and the effects of public health measures: The cases of Congo and Uganda. *J. Theor. Biol.* **229**, 119–126 (2004).
- Bajaj, V., Ma, J. & Thompson, S. How Houston's growth created the perfect flood conditions. *The New York Times*
- Gehlot, H., Honnappa, H. & Ukkusuri, S. V. An optimal control approach to day-to-day congestion pricing for stochastic transportation networks. *Comput. Oper. Res.* **119**, 104929 (2020).
- Newman, M. E. J. Spread of epidemic disease on networks. *Phys. Rev. E Stat. Physics Plasmas Fluids Relat. Interdiscip. Top.* <https://doi.org/10.1103/PhysRevE.66.016128> (2002).
- Fan, C., Jiang, Y., Yang, Y., Zhang, C. & Mostafavi, A. Crowd or Hubs: information diffusion patterns in online social networks in disasters. *Int. J. Disaster Risk Reduct.* **46**, 101498 (2020).

## Acknowledgement

This material is based in part upon work supported by the National Science Foundation under Grant Number CMMI-1846069 (CAREER), CMMI-1832662 (CRISP 2.0 Type 2), the National Academies' Gulf Research Program Early-Career Research Fellowship, the Amazon Web Services (AWS) Machine Learning Award. The authors also would like to acknowledge the data shared by INRIX. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation and Amazon Web Services.

## Author contributions

C.F., X.J., and A.M. designed the study. X.J. and C.F. implemented the method and empirical case study. A.M. support with data acquisition. C.F. and A.M. wrote the main manuscript. All authors reviewed the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to C.F. or A.M.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020