Manuscript

Urban Analytics and City Science

EPB: Urban Analytics and City Science
2022, Vol. 0(0) 1–19

DOI: 10.1177/23998083211069990
journals.sagepub.com/home/epb
SAGE

Anomalous human activity fluctuations from digital trace data signal flood inundation status

Hamed Farahmand
Zachry Department of Civil and Environmental Engineering, Texas A&M University, College Station, TX, USA

Wanqiu Wang
Department of Computer Science and Engineering, Texas A&M University, College Station, TX, USA

Ali Mostafavi
Zachry Department of Civil and Environmental Engineering, Texas A&M University, College Station, TX, USA

Mikel Maron
Community Team, Mapbox, Washington, DC, USA

Abstract

The emergence of mobile platforms equipped with Global Positioning System technology enables real-time data collection affording opportunities for mining data applicable to rapid flood inundation assessment. The collected data can be employed to complement existing methods for rapid flood inundation assessment, such as remote sensing, to enhance situational awareness. In particular, telemetry-based digital trace data related to human activity have intrinsic advantages to be used for inundation assessment. In this study, we investigate the use of Mapbox telemetry data, which provides human activity indices with high spatial and temporal resolutions, for application in rapid flood inundation assessment. Using data from Hurricane Harvey in 2017 in Harris County, Texas, we (1) study anomalous fluctuations in human activities and analyze the differences in activity level between inundated and non-inundated areas and (2) investigate changes in the concentration of human activity, to explore the disruption of human activity as an indicator of flood inundation. Results show that both analyses can provide valuable rapid insights regarding flood inundation status. Anomalous activities can be significantly higher/lower in flooded areas compared with non-flooded areas. Also, the concentration of human activity during the flood propagation period across affected watersheds can be observed. This study contributes to the state of knowledge in smart flood resilience by investigating the application of ubiquitous telemetry-based digital trace data to enhance rapid flood inundation assessment. Accordingly, the use of such digital trace data could

Corresponding author:
Hamed Farahmand, Zachry Department of Civil and Engineering, Texas A&M University, Engineering Building, 201 Dwight Look, College Station, TX 77843, USA.
Email: hamedfarahmand@tamu.edu

provide emergency managers and public officials with valuable insights to inform impact evaluation and response actions.

Keywords

big data, flooding, participatory sensing, statistical analysis, crowdsourcing

Introduction

An effective, rapid flood response requires timely assessment of flood propagation and recession, which enables community response to the rapidly evolving situation (Miguez and Veról, 2017; Rexiline Ragini et al., 2018). The information used for rapid flood inundation assessment is primarily gathered from flood gauges and networks of sensors that collect hydrological data. Although flood gauges and hydrological sensors provide important information of streamflow and the intensity of rainfall on the gauge locations, such information is not adequate for situational awareness of the population or of community actors. To address this limitation, other sources of data that can provide indications of the evolving situation of inundations during a flood event have been examined as a complement to the information collected from flood gauges (Assumpção et al., 2018; Fan et al., 2020; Hao and Wang, 2020). In particular, the emergence of mobile platforms, such as cell phones, equipped with Global Positioning System (GPS) technology enables the rapid data collection from social media and crowdsourced platforms (Chatzimilioudis et al., 2012; Erdelj et al., 2017; Greenwood et al., 2020; Jiménez-Jiménez et al., 2020). Studies of the application of social media and crowdsourced data for improving rapid flood inundation assessment (Li et al., 2018) often integrate user-generated social media data with standard flood-mapping techniques, such as aerial imagery, to provide flood maps for decision-makers, responders, and residents (Kryvasheyeu et al., 2016; Landwehr et al., 2016; Middleton et al., 2014; Pogrebnyakov and Maldonado, 2018; Rosser et al., 2017; Yuan and Liu, 2018). Nevertheless, social media and crowdsourced data have shortcomings, such as limitations related to geo-tagging of posts, which make these data inadequate for rapid flood impact assessment. Digital trace data obtained from cellphones provide geo-tagged data about human activity that shows the potential to complement the information gathered from residents during floods (Stier et al., 2020). The key premise of using digital trace data is that fluctuations in human activities could signal protective actions (e.g., evacuation) and damages to the built environment indicating flood inundation. In addition, digital trace data can provide evidence of human activity at high spatial and temporal resolutions. These characteristics of digital trace data overcome the main limitations of social media and crowdsourced data (such as the need to associate content with locations, limitation of the amount of geo-tagged data, and content validity) (Stieglitz et al., 2018; Yang et al., 2019; Zhang et al., 2019).

This study investigates the application of high-resolution spatial and temporal digital trace data as a source of situational information for the examination of human activity fluctuations for facilitating flood inundation assessment. The promise of this study is that the records of human activities can be a valuable source of information for decision-makers in emergency response; however, there is a need for investigation of the potential applications since (1) human activities are complex and dynamic behaviors that cannot be simply associated to certain parameters and (2) the available data has certain specification such as temporal and spatial aggregation, which might impact its application. More specifically, we investigated the changes in human activity from digital trace data. We used metrics of telemetry-based activity provided by Mapbox for rapid assessment of flood inundation. The data contains two distinct metrics, namely, driving and non-driving activity indexes. The data of human activities are collected by Mapbox and traces are classified based on users' behavior recognized by Mapbox into activities that occurred while users are driving and activities that are not representing driving users. The data of these classes of traces are aggregated and normalized distinctly, and thus, cannot be compared. We used data related to Harris County, Texas, in the context of Hurricane Harvey in August 2017 to examine anomalous changes in human activities, as well as changes in the concentration of human activities to evaluate the association between anomalous fluctuations in human activity and flood inundation status.

Background

This section briefly reviews the state of the literature and recent advancements in using crowd-sourced data for flood inundation assessment and the application of telemetry-based digital trace data for disaster response.

Crowdsourced data for flood inundation assessment

Flood inundation assessment refers to associated impacts across an affected area and the degree of flooding. Hydrodynamic modeling methods and data-driven methods are the most widely used for flood propagation modeling and further inundation assessment (Mustafa et al., 2020; Teng et al., 2017). Using the equations from physical laws, hydrodynamic models simulate the water propagation and overflow using input variables, such as the topography of the region, rainfall, and surface characteristics. (Balekelayi and Tesfamariam, 2019). On the other hand, data-driven methods use data collected in real-time or near real-time to map the extent of flooding. The data might be gathered through satellite imaging, flood gauges and sensors, and surveys, among other sources (Dong et al., 2020b; Ogie et al., 2017).

While hydrodynamic and data-driven techniques can generate effective and relatively accurate flood inundation maps, they present some limitations related to the timeliness of data, spatial coverage, and resolution of data. For example, running a hydrodynamic model to achieve an effective flood inundation map requires high computational power and a wide range of up-to-date data, which might not be feasible to acquire. Moreover, data-driven methods often encounter challenges in the data collection process, such as the spatial-resolution of satellite data, which makes detecting smaller inundated areas difficult; the unavailability of imagery in areas with heavy cloud coverage; and difficulties in the detection of inundation from the photographs taken during night time (McDougall, 2011). Compounding these restrictions (Zhong et al., 2016) is the fact that these techniques often fail to provide timely information required for crucial emergency response decisions (Hao and Wang, 2020).

To complement these methods for flood inundation mapping, various types of crowdsourced and social media data (such as Twitter posts) have been used to enhance the speed and spatial coverage of flood inundation estimation and to reduce the time lag between actual flood propagation and inundation estimation. Several recent studies mined content-based social media data, notably Twitter and Facebook posts, to extract information for flood inundation assessment. By analyzing crowdsourced data, information indicating flood inundation in the vicinity of the geo-location is extracted from user posts (Apel et al., 2004; Brouwer et al., 2017; De Moel et al., 2015; Deng et al., 2016; Fan et al., 2020; Hammond et al., 2015; Hao and Wang, 2020; Li, 2012; Nicklin et al., 2019). For example, in a comprehensive study, a web application has been developed to collect different citizen reports and social media content, analyze, and integrate the information gathered from each data stream into an interface to connect residents with emergency responders to inform each other regarding the dynamically changing situation during a flood event (Urbanrisklab, 2017). Natural language processing (NLP) methods are used for content analysis; for example, NLP has been used to extract the quantifier that indicates the water depth of a location in a Twitter post (Wang et al., 2018). Moreover, given the location and the time that the post is created, the spatial-temporal pattern of tweets can be integrated with other data sources to develop flood inundation status maps. For example, a framework has been proposed by Huang et al. (2018b) that integrates Twitter data into data gathered by remote sensing techniques and river water gauges to improve near real-time flood inundation maps. The Twitter activity data has also been shown to expedite the detection of flood inundation and flood-related events when combined with satellite flood signals (Jongman et al., 2015). Crowdsourced images during floods have been analyzed using deep convolutional neural networks to detect inundation estimate flood severity (Pereira et al., 2020). Volunteer-reported flash floods and geo-located tweets were integrated for real-time flood maps by generating probability index distribution layers and digital elevation models and assigning weights using image-extracted land surface showing the wetness of the area (Huang et al., 2018a). Another stream of research deals with identifying the regions affected by the flood and detecting emergency situations caused by flood inundation (Sarica et al., 2021; Wang and Taylor, 2019; Yin et al., 2020). Another topic of interest is the enhancement of credibility of information extracted from social media data regarding reported disruption events and damage. Some studies (such as (Fan et al., 2019)) have developed a graph-based method for credible disruption event detection from social media data, which can boost the situational awareness regarding disruption and damage in the community (Fan et al., 2019). Using social media data and remote sensing (Ahmad et al., 2019), researchers have developed a framework to estimate the passability of roads, taking into consideration disruptions and inundations that occur during floods.

Despite their growing use, there are limitations in social media data analysis. First, social media data might be biased by factors, such as distance to impacted areas, the popularity of the user, and demographic characteristics of users. Moreover, reliable analysis of social media content requires well-established and standard ontologies which are difficult to develop to classify posts, compare results, and validate findings. Finally, the number of geo-tagged social media posts is often limited compared to the digital trace data, which limits a comprehensive spatial and temporal coverage for disaster situational awareness applications. Due to the limitations of crowdsourced and social media data, there is a need for studies to evaluate the usefulness of digital trace data and other emerging data.

Digital trace data for improving disaster response

Digital trace information of locations and activities of cellphone users is obtained from telemetry-based data gathered from cell phones using GPS technology (Lopez and Ferreira, 2021; Ma et al., 2019). These data have been used to understand phenomena related to human activities. Digital trace data can reveal changes in human activity patterns. This characteristic is particularly useful in studying crises. Perturbations in communities cause fluctuations in human activities which in turn could signal hazard exposure and impacts. Digital trace data has also been used for investigating voters' behavior (Bach et al., 2019) and migration patterns (Marquez et al., 2019). In another study, human activity captured by human digital trace data was considered an early indication of the COVID-19 spread risk across the United States (Gao et al., 2020). Previous studies have relied upon digital trace data for investigating aspects of disaster response (Cumbane and Gidófalvi, 2019). For example, human mobility data recorded by GPS devices have been used to study human behavior during large-scale disasters, which shows that emergency behavior can be correlated with the mobility pattern extracted from digital trace data (Song et al., 2014). Digital trace data has also been used for near real-time investigation of population movement dynamics following earthquakes (Wilson et al., 2016); the study examined mobility patterns captured by mobile phones to estimate the population displacement triggered by hazard events (Wilson et al., 2016). Moreover, community-scale digital trace data has been used for quantifying the impacts of a winter storm in Texas to analyze disparities in storm-related impacts on different sub-populations (Lee et al., 2021).

Despite the growing attention to potential applications of digital trace data to enhance situational awareness in disasters, there is a dearth of studies focusing on the correlation of fluctuations in human activities obtained from digital trace data with hazard exposures and impacts. In particular, in the context of rapid flood inundation assessment, the potential of digital trace data remains under-investigated. There are very limited studies that focus on integrating the insight from the digital trace data into existing models to investigate the flood impact assessment. For example, to investigate the importance of different heterogeneous human activity features including human mobility, visits to points of interest, and social media posts, various machine learning models have been developed and fluctuation of feature importance in different flood phases have been analyzed (Yuan et al., 2021). Results indicate that daily changes in digital traces of human activity often have large importance for rapid flood impact assessment in the developed models. Therefore, the digital trace data would have the potential for providing inundation signals in near real-time inundation prediction and monitoring. It indicates that there is a need for further investigation that shows the association between inundation and the temporal and spatial fluctuation in digital traces of human activities. In this study, we investigate the potential use of temporal and spatial changes in human activity patterns derived from digital trace data to enable rapid assessment of flood inundation. This paper is organized as follows. In *Study area and context*, we introduce the characteristics of the study area and the flood event investigated in this study. In *Data collection and pre-processing*, we discuss data collection and processing steps. *Data analysis and results* presents the analysis framework, metrics, and methods devised for data analysis as well as the results of the data analysis. *Conclusion* presents study findings and conclusions and reviews limitations and future work directions.

Study area and context

Harris County, home to Houston, Texas, the fourth largest city in the United States, has experienced rapid population growth over the past decades (Qian, 2010). Harris County is among the most flood-prone counties in the United States due to its location in a coastal area, burgeoning urban development, and the lack of flood control infrastructure development in parallel with the urban development and population growth (Dong et al., 2020b). Accordingly, the county has experienced several severe flood events, including the Memorial Day Flood in 2015, the Tax Day Flood in 2016, and Hurricane Harvey in 2017. Each event caused extensive flooding with losses ranging from hundreds of millions to hundreds of billion dollars (Dong et al., 2019). Therefore, Harris County was selected as the testbed for this study. Hurricane Harvey, which made landfall in 2017, caused one of the most devastating floods experienced by Harris County. As a Category 4 storm landing on the Texas Gulf Coast on August 25, 2017, and dissipating inland on August 30, 2017, Hurricane Harvey caused extensive economic and social consequences (NOAA, 2017). In this study, we focused on areas with extensive flooding whose flood maps contain sufficient human activity tiles so that conclusions regarding potential associations between changes in human activity and flood inundation are valid. Of 22 watersheds in the county, we selected 8 watersheds in which (1) sufficient digital trace data for our analysis can be gathered and (2) flooding during Hurricane Harvey had clear impacts and caused considerable damage and disruptions.

Data collection and pre-processing

We obtained digital trace telemetry data from August 1 through August 30, 2017, from Mapbox. We chose Mapbox as the source of the telemetry data due to its ability to collect temporal and spatial telemetry-based human activity with a proper level of aggregation. Moreover, Mapbox provides different indexes based on the type of activity (i.e., driving vs non-driving), which allows better interpreting the activity level. The spatial unit of data aggregation is tile. The partition of tiles is based on Mapbox data format, which enables creating spatial-resolution grids. Human activity is collected, aggregated, and normalized by Mapbox based on the geography information updates of locations of users' devices (such as their cell phones) from applications that use Mapbox Software Development Kit (SDK). The more users located in a tile at time *t*, the greater the human activity index. Human Activity here refers to the density of digital traces recorded from user devices drawn from users of Mapbox SDK globally contributing to live location updates. The dataset captures significant driving and non-driving mobile device activity aggregated into geographic tiles. Tiles represent square geographic areas approximately 100 m per side, which varies depending on latitude. Mapbox provided a 4-hour temporal resolution as raw data. Data might not exist for all the spatial units, as data is derived from cell phone activity depends on the updates of the geography information of cell phone users. For example, for recreational facilities that are closed nights and weekends, it is possible that no update is generated for human activity. Moreover, due to privacy concerns and the data aggregation process, tiles with small numbers of users, traces are excluded. Mapbox provides two activity indexes, namely, driving activity index and, non-driving activity index. Each index is calculated for the aggregated traces for all tiles across time. Then the raw indices are normalized. Normalization is compartmented separately by month and type of the trace and yields a normalized activity index for each tile in each 4-hour time step. The normalized values range between 0 and 1. We used the Mapbox data for August 2017 to allow a comparison of activity levels during flood inundation with the baseline activity level. Using the collected Mapbox data, we performed data pre-processing, which consisted of normalization and aggregation of the human activity data into the time steps and spatial units for the analysis (Figure 1). To estimate flood inundation in the study area, we used estimated daily flood inundation maps. Estimated flood inundation maps were created based on gauge points obtained from the National Weather Service (NOAA National Weather Service, 2014). Water surface elevations were determined based on the stage readings. After data processing, water surface elevations were used to generate a triangular irregular networks (TIN) file that covers the affected area. The TIN file can be used to estimate the flood depth, identify flooded areas, and identify the areas with a possibility of flooding. We used the approximate flood extent maps for flood inundation estimation for this study.

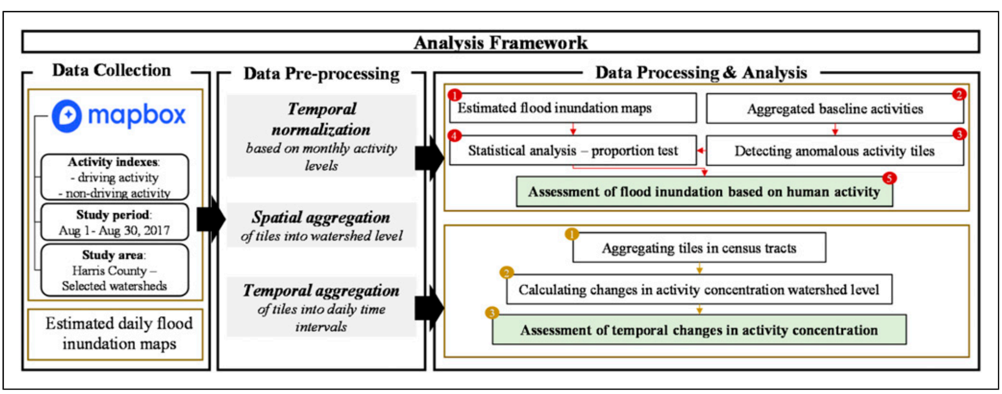

Figure 1. Analysis framework.

Data analysis and results

Analysis framework

Figure 1 shows the analysis framework in this study, including objectives as well as methods used for the analysis. Following the cleaning and pre-processing human activity data, we performed anomaly detection to determine the tiles with anomalous activity compare to the baseline activity, which is calculated based on the pre-Harvey human activity. Anomalous values in a dataset can be identified employing different techniques for defining anomalies (Prasad et al., 2009). One approach for defining anomalies is to associate them with the probability of occurrence. In this approach, a threshold in the cumulative distribution of the data points can be defined as the cut-off to determine the data points that have extreme values (Zhang et al., 2011). Many studies have considered 5 percentages of data at each tail as the set of data points with extreme values, which are identified as anomalous (Sun et al., 2015; Zhai et al., 2005; Zhang et al., 2011). Similarly, we define two types of anomalous activities based on the deviation of activity level in a spatial unit from the activity level in the baseline period to capture tiles with considerably lower or higher activity compared to the pre-Harvey activities. We considered activity levels lower than the 5th percentile (low anomaly) and higher than the 95th percentile (high anomaly) as anomalous activities. Then, we performed proportion tests to investigate the association between anomalous activities inside and outside flooded areas at the watershed level. In addition to activity indices of each tile, we calculated the Venables distance as a measure to quantify the agglomeration of human activity within each watershed. Then, we explored the temporal and spatial changes of the Venables distance to investigate how the human activity digital trace data may provide interpretable information of the flood inundation.

Flood inundation assessment using traces of human activity

To explore the use of telemetry-based digital trace data for rapid flood inundation assessment, we investigated the relationship between the changes in human activity and flood inundation status of the affected area. First, we selected the watersheds in which data availability and flood extent are reasonable for the validity of the analysis. Of the 22 watersheds in Harris County, we selected 8 in the west and southwest part of the county. These watersheds mostly fall inside the boundaries of Harris County; they experienced considerable inundation during Hurricane Harvey. Figure 2 shows the selected watersheds, as well as the 100-year and 500-year floodplains. The specifications of the watersheds can be seen in Table 1.

Our central hypothesis is that the areas with inundation would signal more anomalous telemetry-based human activities compared with the areas without inundation. To test this hypothesis, first we developed a baseline activity level to allow comparison and determination if the activity in a tile at a period is considered as an anomaly. We considered the activity of the tiles in each census tract during the pre-Harvey portion of August 2017 (August 1 to August 26) as the baseline activity. For each census tract, we batched all daytime (from 8:00 a.m., to 8:00 p.m.,) activities recorded from August 1 through August 26. We also distinguished between the weekdays and weekends for the calculation of the baseline since the non-driving activity for weekdays and weekends are different. Then, we considered the 5th and 95th percentile as the threshold for detecting low and high anomalous activities, respectively. In essence, a value of activity for tile *a* during time *t* during Harvey is considered as a high anomaly if its value is higher than the 95 percentiles of the values of the activities recorded in similar days (i.e., weekdays or weekends) in the same census tract during the pre-Harvey period in August 2017. Similarly, a low anomaly can be detected if its value is lower than the 5th percentile of the values of the activities recorded in similar days.

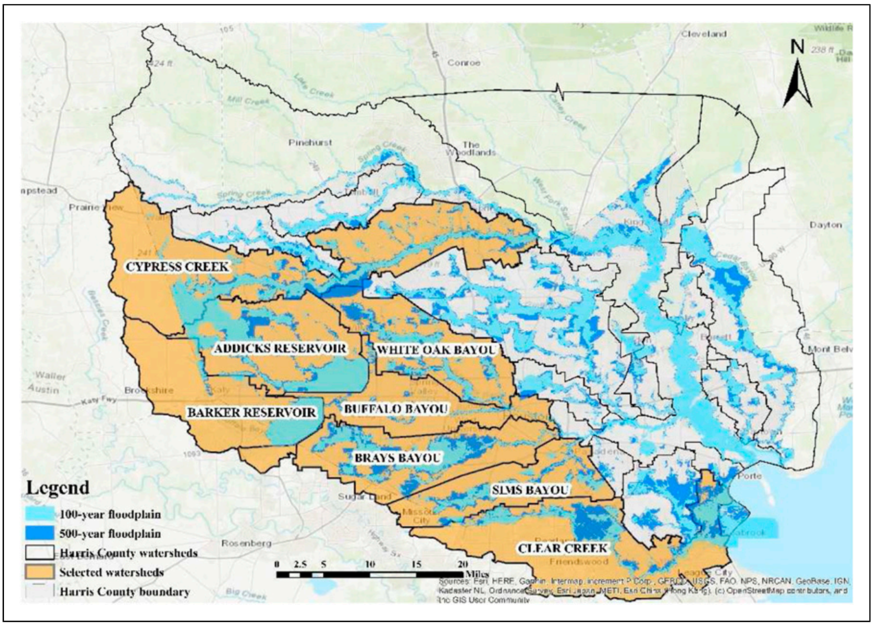

Figure 2. Study area and selected watersheds for flood inundation assessment task.

Table 1. Specification of selected watersheds.

Watershed name	Drainage area (sq. miles)	Open streams (miles)	Population (2010 US Census)
Addicks Reservoir	138	159	259,694
Sims Bayou	94	121	284,727
Buffalo Bayou	102	106	444,602
Barker Reservoir	126	69	88,895
White Oak Bayou	111	146	433,250
Cypress Creek	267	250	347,334
Clear Creek	202	128	36,878
Brays Creek	197	154	164,172

In the next step, we overlaid the tiles on the daily flood inundation maps for August 27 to August 30, 2017. Although Harvey affected Harris County in August and September, we focused on August activities. The main reason is that the baseline that is used for detection of anomalies needed to be in the same month as the event under study since Mapbox performs normalization process in a monthly basis. In the study period, tiles with low or high anomalies were detected for each day. We then counted the number of tiles with anomalous activities in each watershed for each day. Also, we determined the flooding status of the tile using the daily flood inundation maps. By doing so, we categorized tiles on a daily basis given their inundation status and whether they are signaling anomalous activities. For example, for the low anomaly activity, we categorized tiles for specific days during Hurricane Harvey as four categories (i.e., non-anomaly/non-flooded, non-anomaly/flooded, anomaly/flooded, and anomaly/non-flooded). Figure 3 shows the distribution of the tiles falling into four categories for August 27 to August 30.

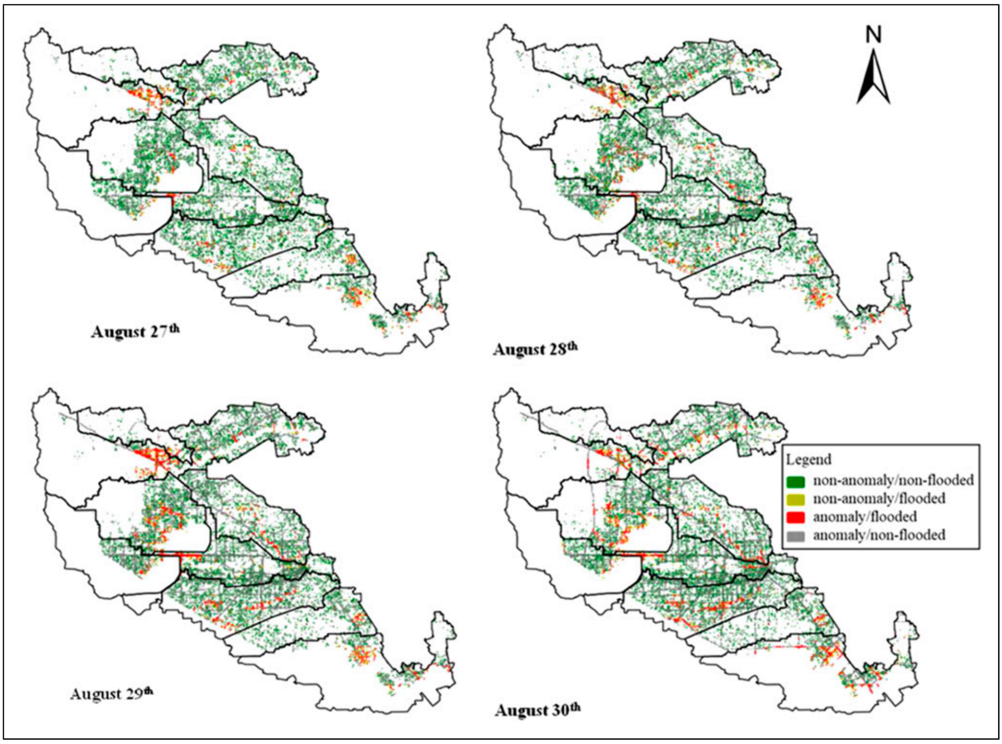

Figure 3. Distribution of human activity and status of inundation from August 27 through August 30.

In the next step, we used proportion tests for populations with different sample sizes. Given two sets of observations that can have two possible outcomes (i.e., success and failure), the test examines the null hypothesis, meaning the proportions of success in the two sets are the same. We used R software and the Stats package (Wilson and Norden, 2015) to perform the proportion test. For each day and each watershed, we divided tiles into two sets; inundated and non-inundated. Then we defined success as anomalous activity and failure as non-anomalous activity. Doing so we can define the null hypothesis (H_0) and the alternative hypothesis (H_1) as follows:

H_0: two sets from which the human activity tiles were drawn have the same proportion of tiles with anomalous activity.

H_1: this proportion is different for two sets.

Once we state our null hypothesis and alternative hypothesis, we calculate test statistics using equation (1) as follows:

$$z = \frac{(p_1 - p_2)}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \tag{1}$$

where p_1 and p_2 are proportions for set 1 and set 2, and n_1 and n_2 are sample sizes for set 1 and set 2, respectively. p is the average of proportions and is calculated using equation (2) as follows:

$$p = \frac{p_1 n_1 + p_2 n_2}{n_1 + n_2} \tag{2}$$

Then, using the z and the defined confidence interval, we can examine if we can reject the null hypothesis (Chow et al., 2017). Indeed, proportion tests calculated whether the proportion of tiles with anomalous activities was significantly different among the two groups using 10%, 1%, and 0.1% confidence intervals. The test was performed for anomalous high and anomalous low activities separately. Results are shown in Table 2.

As shown in Table 2, the result of the proportion test is statistically significant for some days and watersheds. For example, in Cypress Creek watershed, which is a heavily populated watershed and which experienced extensive flooding during Hurricane Harvey, inundated areas had a significantly lower proportion of tiles with anomalous activity levels from August 27 to August 30. In the Sims Bayou watershed, however, where flooding has been less extensive, we cannot see significant differences of the proportion of high anomalous activities. Similarly, we can see in Table 2 that there are significant differences in the proportion of low anomalous non-driving activities between inundated and non-inundated areas. In total, in 25 out of 64 cases, at least the proportion of tiles with either high or low anomalous activities are significantly different for inundated and non-inundated areas. This finding shows that the anomaly in tile activity can be a reliable signal for enhancing near rapid identification of inundated areas during a flood event.

Next, we explored the distribution of the tiles by focusing on the portion of tiles with anomalous activities that are not inside the inundated areas. To do so, we focused on three watersheds (i.e., Buffalo Bayou, Brays Bayou, and Sims Bayou. In Figure 4, we compare the distribution of anomalous activities versus the land use map in the study area. As we can see in box a, there is a concentration of the anomalous activities that are in a non-inundated area that forms around a main road in between a residential area. Also, comparing August 27th and August 28th, we can see that the concentration of such tiles is increasing. It can show that the activity is increasing in these areas significantly while people are using the road to move to safe zones as inundation starts. Also, in box b, which encompasses a safe residential area (i.e., area without inundation), the concentration of normal activity increases, which shows that there is no considerable disruption. Moreover, we can see that the density of tiles in Sims Bayou is relatively lower than Brays Bayou and Buffalo Bayou. On the other hand, we can see that there are few tiles in the floodplain in this watershed. Considering that the area experienced flooding in this time period, the limited data points could lead to the lack of significant results for proportion tests for this watershed (Table 2).

Table 2. Proportion tests for anomalous high activity and anomalous low activity.

Watershed name	Anomaly highs				Anomaly lows			
	Aug 27	Aug 28	Aug 29	Aug 30	Aug 27	Aug 28	Aug 29	Aug 30
Addicks Reservoir	.717	.287	.002**	.775	.002**	.358	.280	.336
Buffalo Bayou	.000***	.364	.012*	.087*	.000***	.001**	.000***	.050*
Barker Reservoir	.012*	.598	.402	.095*	.760	.000***	.137	.965
Sims Bayou	.908	.369	.742	.292	.440	.515	.060*	.932
White Oak Bayou	.524	.049*	.500	.235	.184	.070*	.015*	.001***
Cypress Creek	.000***	.006**	.012*	.003**	.212	.699	.188	.112
Clear Creek	.789	.395	.537	.431	.795	.110	.060*	.198
Brays Bayou	.129	.528	.786	.000***	.922	.507	.002**	.001***

* $p < 0.1$, ** $p < 0.01$., *** $p < 0.001$.

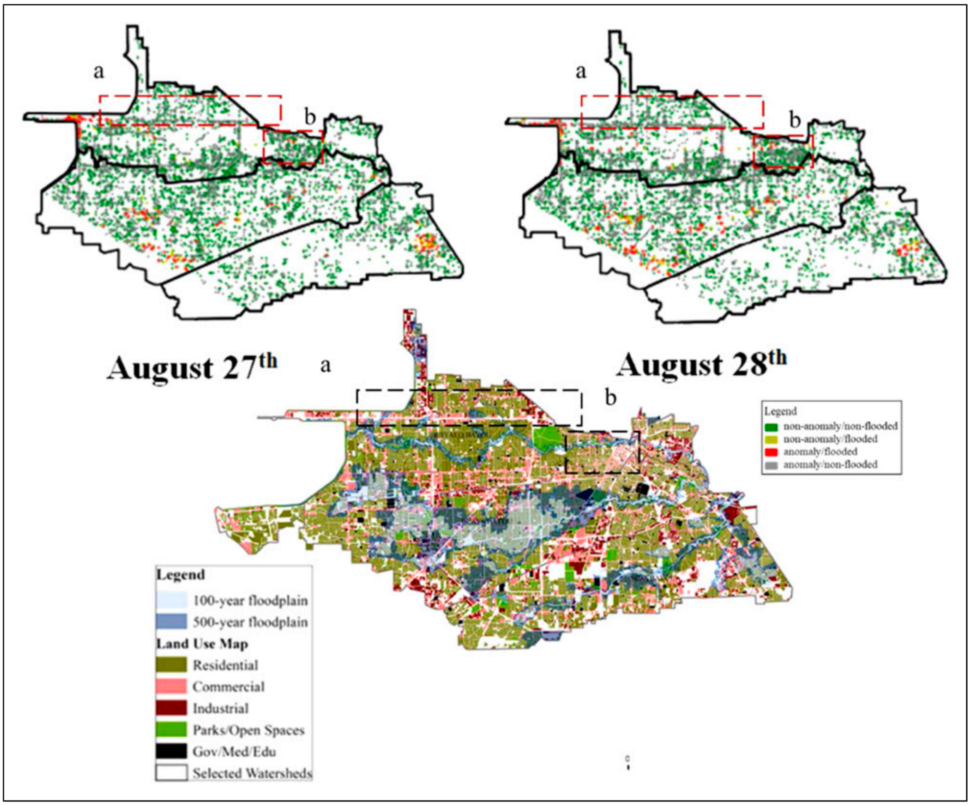

Figure 4. Distribution of anomalous activities versus land use in three watersheds for 27–28 August.

Exploring changes in venables distance based on telemetry-based digital trace data

To explore the spatial-temporal pattern of human activity, we investigated the changes of the Venables distance, D_v, which captures the spatial structure of human activities and could be an indicator of the concentration of human activities. The Venables distance captures the average distance (i.e., concentration) of human activities across a city, county, or any spatial unit (Louail et al., 2014). The D_v is calculated as follows:

$$D_v = \frac{\sum_{T_x \neq T_y} a_{T_x,t} \cdot a_{T_y,t} \cdot d_{T_x,T_y}}{\sum_{T_x \neq T_y} a_{T_x,t} \cdot a_{T_y,t}} \tag{3}$$

where $a_{T_x,t}$ and $a_{T_y,t}$ capture the value of activity in tile x and tile y, respectively, and d_{T_x,T_y} shows the distance between the centroids of tile x and tile y (Gao et al., 2021). The Venables Distance captures the intensity level of overall activities in a county or city. A higher value of D_v implies a lower concentration of human activity and higher distance among people (Gao et al., 2021). Figure 5 schematically illustrates the difference between a set of tiles with concentrated activity and a set of tiles with spatially distributed activity, where both sets of tiles have an equal total activity (values in cells show activity level). As can be seen, Venables distance can capture the difference in the extent to which activities are concentrated.

To study the temporal and spatial changes of D_v during Hurricane Harvey, we first needed to lower the computation cost of calculating D_v at the tile level. We aggregated the activity levels at the census tract level. We calculated the average activity level of each census tract in a watershed for 4-hour time periods.

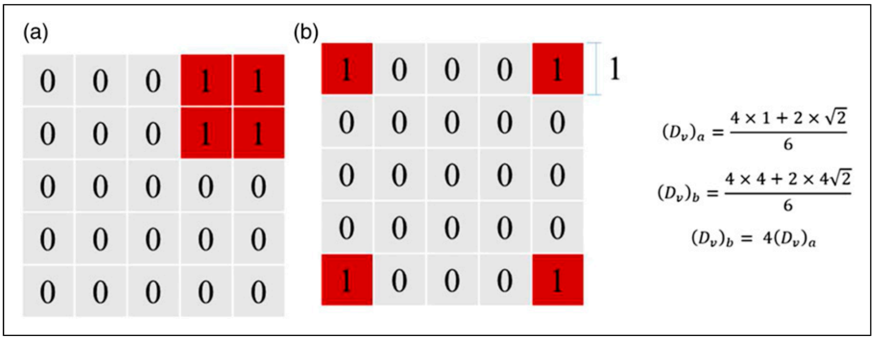

Figure 5. Illustration of (a) spatially concentrated and (b) spatially distributed activity in a tile grid.

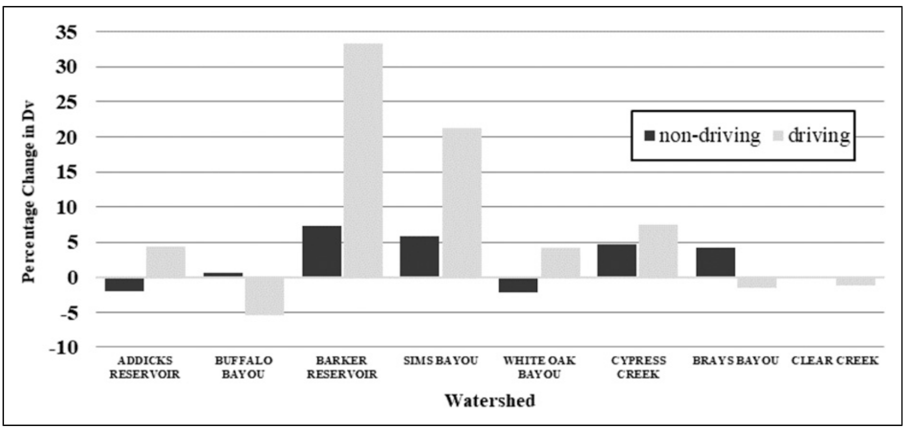

Figure 6. The percentage changes of the $\boldsymbol{D_v}$ for the selected watershed from August 27 to August 30.

Moreover, we calculated the distance between the centroids of pairs of census tracts. Then, the D_v was calculated for August 2017. Similar to the anomaly detection, we considered the average of the non-zero values of D_v between August 1 and August 26 as the baseline and also calculated the average of the D_v during the hurricane impact period of August 27 to August 30.

Figure 6 shows the percentage changes of the D_v for the selected watershed during the considered flooding period compared to the baseline. The Venables distance has seen considerable change in different watersheds during the inundation period. In general, the Venables distance calculated based on driving activity shows an increase, which implies a lower concentration of the driving activities. This result can be attributed to the disruption of daily activities that led to a reduction in driving activity in the areas that are points of concentration on normal days. Brays Bayou, Clear Creek, and Buffalo Bayou, however, show a decrease of Venables distance and an increase in the concentration of activities, which can be attributed to road inundation in a way that only specific roads in the non-inundated are passable and driving activities are concentrated on such areas. Moreover, the Venables distance for non-driving activity generally increased during the inundation period. As we can see, the increase in the Venables distance is considerable for watersheds with extensive inundation in the more populous areas such as Baker Reservoir, Cypress Creek, and Brays Bayou. The increase in the Venables distance during the flood inundation period implies a lower non-driving activity level, which shows that people have a lower presence in the areas of concentration during normal periods, possibly commercial areas.

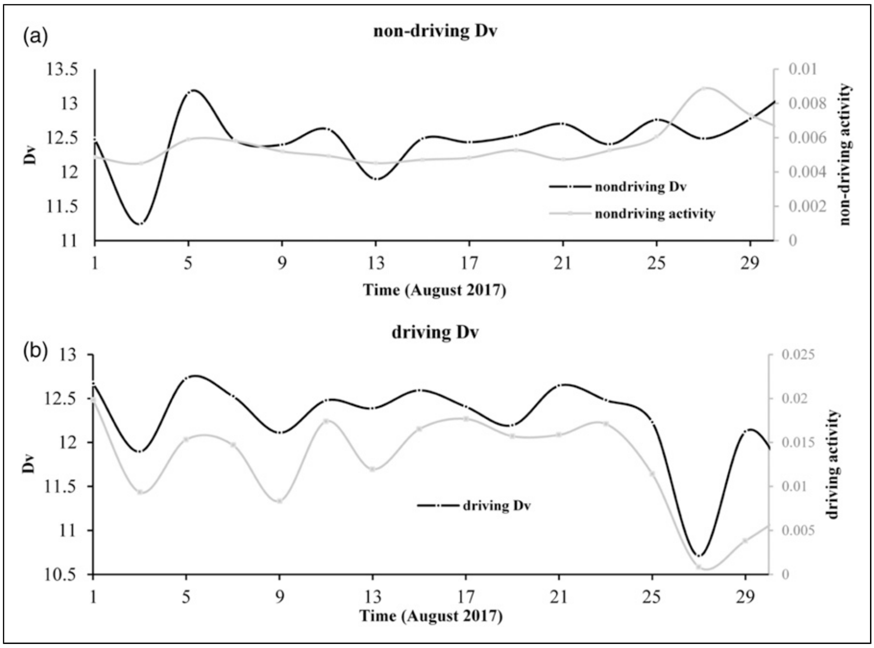

Figure 7. Changes of average daily activity and Venables distance for Brays Bayou from August 1 to August 30; (a) non-driving, and (b) driving activity.

To evaluate the temporal changes of the D_v, we focused on a single watershed, Brays Bayou, since it was impacted by Hurricane Harvey in terms of disruption of daily activities during the inundation period and its extent of inundation was relatively high. We also focused on examining the changes of activity level as they associated with the Venables distance. This approach promoted a better understanding of changes in human activity since it revealed both the number of activities and their spatial concentration. Figure 7 shows the changes of non-driving and driving average daily activities as well as the Venables distance calculated based on the daily activity levels deploying the same procedure as the previous section. As we can see in Figure 7(a), the non-driving activity level showed less change during the inundation period when compared to the baseline (August 1 to August 26). The Venables distances, however, show an increase during the inundation period. An increase in the Venables distance is often interpreted as a decline in the concentration of activities; therefore, the fact that the actual value of the activity has low variation, but the Venables distance increases imply that the people are not leaving the area in substantial numbers, but daily activity is disrupted. The disruption is reflected in the increase in the Venables distance. Figure 7(b) shows changes in the average daily driving activity in the watershed, as well as the changes of the Venables distance. From the results, it can be seen that the activity level experienced a considerable drop as a result of the inundation in the watershed. The decline in the D_v implies a disruption in driving activity. It can be also attributed to the fact that the flood resulted in the inundation of a proportion of road segments and forced driving to be concentrated where road segments are not inundated.

Conclusion

In this study, we evaluated the evidence of human activity derived from digital trace data for purposes of assessing rapid flood inundation. The core idea of the study is that spatial and temporal changes of human activity in a flood-impacted area can signal flood inundation. We examined telemetry-based digital trace data as a suitable source of human activity information that can provide high-resolution digital trace data and address intrinsic shortcomings of crowdsourced and social media data for flood situational awareness. In particular, we investigated the fluctuations in human activity indexes derived from Mapbox data that can be collected in near real-time. We found that the proportion of anomalous activities in flooded areas could be significantly higher/lower compared with non-flooded areas. Moreover, we observed that changes of watershed-level human activity agglomeration provide credible insight about flood inundation in the impacted area. The contributions of this study are twofold: first, this study reveals the promise of the use of a ubiquitous data source of credible human activity for harnessing big data for rapid flood inundation assessment. Second, it introduces and tests two different approaches for assessment of changes in human activity—assessment of anomalous activities and evaluation of activity agglomeration—to acquire interpretable information from digital trace data for flood inundation assessment. Hence, the findings of this study show the potential of community-scale big data (such as digital traces from cellphone activities) for enhancing situational awareness during urban flooding. These findings also have practical implications primarily for emergency managers and responders (Dong et al., 2020a). Notably, considering the importance of rapid flood inundation assessment for effective emergency response, our findings indicate that the analysis of human activity data provides complementary information for rapid identification of area of high inundation status. This study and its findings provide a basis for future studies to further investigate the characteristics (e.g., spatial biases) of telemetry-based digital trace data for flood situation awareness.

Despite the insights that the study of digital trace data provided for flood inundation and impact assessment, this study has limitations that need to be considered. First, the flood maps that are used for flood extent estimation and categorization of tiles into inundated and non-inundated areas do not have same shapes, and therefore, spatially comparing these two areas causes marginal errors, which can be relaxed if the human activity data is available in smaller tiles. Moreover, the pre-processing of digital trace data leads to removing tiles with small number of records, which may lead to missing pieces of information regarding the areas with a significant drop of human activity. Finally, the spatial aggregation used for reducing the computation cost of calculation of Venables distances might be relaxed using high-performance computers. Future studies are needed to develop pipelines to effectively collect and analyze digital trace data and evaluate human activity fluctuations for estimating inundation status and informing response actions by emergency responders and public officials.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The authors would like to acknowledge funding support from the National Science Foundation CRISP 2.0 Type 2 No. 1832662, "Anatomy of Coupled Human-Infrastructure Systems Resilience to Urban Flooding: Integrated Assessment of Social, Institutional, and Physical Networks."

Author contributions

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

Data availability

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

Disclaimer

Any opinions, findings, and conclusions or recommendations expressed in this research are those of the authors and do not necessarily reflect the views of the funding agencies.

ORCID iD

Hamed Farahmand https://orcid.org/0000-0003-2366-8161

References

Ahmad K, Pogorelov K, Riegler M, et al. (2019) Automatic detection of passable roads after floods in remote sensed and social media data. *Signal Processing: Image Communication*. Elsevier B.V 74: 110–118. DOI: 10.1016/j.image.2019.02.002.

Apel H, Thieken AH, Merz B, et al. (2004) Flood risk assessment and associated uncertainty. *Natural Hazards and Earth System Sciences*. European Geosciences Union 4(2): 295–308. DOI: 10.5194/nhess-4-295-2004.

Assumpção TH, Popescu I, Jonoski A, et al. (2018) Citizen observations contributing to flood modelling: Opportunities and challenges. *Hydrology and Earth System Sciences* 22(2): 1473–1489. DOI: 10.5194/hess-22-1473-2018.

Bach RL, Kern C, Amaya A, et al. (2019) Predicting voting behavior using digital trace data. *Social Science Computer Review*. SAGE Publications Inc 39: 862–883. DOI: 10.1177/0894439319882896.

Balekelayi N and Tesfamariam S (2019) Graph-theoretic surrogate measure to analyze reliability of water distribution system using bayesian belief network-based data fusion technique. *Journal of Water Resources Planning and Management*. American Society of Civil Engineers (ASCE) 145(8): 04019028. DOI: 10.1061/(ASCE)WR.1943-5452.0001087.

Brouwer T, Eilander D, Van Loenen A, et al. (2017). Probabilistic flood extent estimates from social media flood observations. *Natural Hazards and Earth System Sciences*. Copernicus GmbH 17: 735–747. DOI: 10.5194/nhess-17-735-2017.

Chatzimilioudis G, Konstantinidis A, Laoudias C, et al. (2012) Crowdsourcing with smartphones. *IEEE Internet Computing* 16(5): 36–44. DOI: 10.1109/MIC.2012.70.

Chow S-C, Shao J, Wang H, et al. (2017) *Sample Size Calculations in Clinical ResearchSample Size Calculations in Clinical Research*. 3rd edition. Boca Raton, Florida: Chapman and Hall/CRC. DOI: 10.1201/9781315183084.

Cumbane SP and Gidófalvi G (2019) Review of big data and processing frameworks for disaster response applications. *ISPRS International Journal of Geo-InformationMDPI AG* 8(9): 387. DOI: 10.3390/ijgi8090387.

De Moel H, Jongman B, Kreibich H, et al. (2015) Flood risk assessments at different spatial scales. *Mitigation and Adaptation Strategies for Global Change*. Kluwer Academic Publishers 20(6): 865–890. DOI: 10.1007/s11027-015-9654-z.

Deng Q, Liu Y, Zhang H, et al. (2016) A new crowdsourcing model to assess disaster using microblog data in typhoon Haiyan. *Natural Hazards*. Springer Netherlands 84(2): 1241–1256. DOI: 10.1007/s11069-016-2484-9.

Dong S, Li Q, Farahmand H, et al. (2020a) Institutional connectedness in resilience planning and management of interdependent infrastructure systems. *Journal of Management in Engineering. American Society of Civil Engineers (ASCE)* 36(6): 04020075. doi: 10.1061/(ASCE)ME.1943-5479.0000839.

Dong S, Yu T, Farahmand Het al. (2019) Bayesian modeling of flood control networks for failure cascade characterization and vulnerability assessment. *Computer-Aided Civil And Infrastructure Engineering* 35(6): 668–684. doi: 10.1111/mice.12527.

Dong S, Yu T, Farahmand H, et al. (2020b) Probabilistic modeling of cascading failure risk in interdependent channel and road networks in urban flooding. *Sustainable Cities and Society*. Elsevier Ltd, 62, 102398. doi: 10.1016/j.scs.2020.102398.

Erdelj M, Król M and Natalizio E (2017) Wireless sensor networks and multi-UAV systems for natural disaster management. *Computer Networks*. Elsevier B.V 124: 72–86. DOI: 10.1016/j.comnet.2017.05.021.

Fan C, Mostafavi A, Yao W, et al. (2019) A Graph-based approach for detecting critical infrastructure disruptions on social media in disasters. In: 2019 52th Hawaii International Conference on System Sciences (HICSS), Maui, Hawaii, USA, 08–11 Jan 2019, pp. 1975–1984.

Fan C, Jiang Y and Mostafavi A (2020) Social sensing in disaster city digital twin: Integrated textual–visual–geo framework for situational awareness during built environment disruptions. *Journal of Management in Engineering*. American Society of Civil Engineers (ASCE) 36(3): 04020002. DOI: 10.1061/(ASCE)ME.1943-5479.0000745.

Gao X, Fan C, Yang Y, et al. (2020) Early indicators of COVID-19 spread risk using digital trace data of population activities. Available at: http://arxiv.org/abs/2009.09514 (accessed on 25 September 2020).

Gao X, Fan C, Yang Y, et al. (2021). Early indicators of human activity during COVID-19 period using digital trace data of population activities. *Frontiers in Built Environment*. Frontiers Media SA 6: 223. DOI: 10.3389/fbuil.2020.607961.

Greenwood F, Nelson EL and Greenough PG (2020) Flying into the hurricane: A case study of UAV use in damage assessment during the 2017 hurricanes in Texas and Florida. *PLOS One*. Public Library of Science. 15(2): e0227808. DOI: 10.1371/journal.pone.0227808.

Hammond MJ, Chen AS, Djordjević S, et al. (2015) Urban flood impact assessment: A state-of-the-art review. *Urban Water Journal*. Taylor and Francis Ltd 12(1): 14–29. DOI: 10.1080/1573062X.2013.857421.

Hao H and Wang Y (2020). Leveraging multimodal social media data for rapid disaster damage assessment. *International Journal of Disaster Risk Reduction*. Elsevier Ltd 51: 2212–4209. DOI: 10.1016/j.ijdrr.2020.101760.

Huang X, Wang C and Li Z (2018a) A near real-time flood-mapping approach by integrating social media and post-event satellite imagery. *Annals of GIS*. Taylor and Francis Ltd 24(2): 113–123. DOI: 10.1080/19475683.2018.1450787.

Huang X, Wang C and Li Z (2018b) Reconstructing flood inundation probability by enhancing near real-time imagery with real-time gauges and tweets. *IEEE Transactions on Geoscience and Remote Sensing* 56(8): 4691–4701. DOI: 10.1109/TGRS.2018.2835306.

Jiménez-Jiménez SI, Ojeda-Bustamante W, Ontiveros-Capurata RE, et al. (2020) Rapid urban flood damage assessment using high resolution remote sensing data and an object-based approach. *Geomatics, Natural Hazards and Risk*. Taylor and Francis Ltd 11(1): 906–927. DOI: 10.1080/19475705.2020.1760360.

Jongman B, Wagemaker J, Romero B, et al. (2015) Early flood detection for rapid humanitarian response: Harnessing near real-time satellite and twitter signals. *ISPRS International Journal of Geo-Information*. MDPI AG 4(4): 2246–2266. DOI: 10.3390/ijgi4042246.

Kryvasheyeu Y, Chen H, Obradovich N, et al. (2016) Rapid assessment of disaster damage using social media activity. *Science Advances*. American Association for the Advancement of Science 2(3): e1500779. DOI: 10.1126/sciadv.1500779.

Landwehr PM, Wei W, Kowalchuck M, et al. (2016) Using tweets to support disaster planning, warning and response. *Safety Science*. Elsevier B.V 90(3): 33–47. DOI: 10.1016/j.ssci.2016.04.012.

Lee C-C, Maron M and Mostafavi A (2021) *Community-scale Big Data Reveals Disparate Impacts of the Texas Winter Storm of 2021 and its Managed Power Outage*. arXiv, 2108.06046, Submitted for publication.

Li Y (2012) Assessment of damage risks to residential buildings and cost–benefit of mitigation strategies considering hurricane and earthquake hazards. *Journal of Performance of Constructed Facilities*. American Society of Civil Engineers 26(1): 7–16. DOI: 10.1061/(ASCE)CF.1943-5509.0000204.

Li Z, Wang C, Emrich CT, et al. (2018) A novel approach to leveraging social media for rapid flood mapping: A case study of the 2015 South Carolina floods. *Cartography and Geographic Information Science*. Taylor and Francis Inc 45(2): 1–14. DOI: 10.1080/15230406.2016.1271356.

Lopez RP and Ferreira J (2021) Identifying spatio-temporal hotspots of human activity that are popular non-work destinations. *Environment and Planning B: Urban Analytics and City Science*. SAGE Publications Ltd 48(3): 433–448. DOI: 10.1177/2399808320970209.

Louail T, Lenormand M, Cantu Ros OG, et al. (2014) From mobile phone data to the spatial structure of cities. *Scientific Reports* 4(1): 1–12. doi: 10.1038/srep05276.

Ma D, Omer I, Osaragi T, et al. (2019) Why topology matters in predicting human activities. *Environment and Planning B: Urban Analytics and City Science*. SAGE Publications Ltd 46(7): 1297–1313. DOI: 10.1177/2399808318792268.

Marquez N, Garimella K, Toomet O, et al. (2019) *Segregation and sentiment: Estimating refugee segregation and its effects using digital trace data. Guide to Mobile Data Analytics in Refugee Scenarios*. Springer International Publishing, 265–282. doi: 10.1007/978-3-030-12554-7_14.

McDougall K (2011) Using volunteered information to map the Queensland floods. In: Proceedings Of the 2011 Surveying And Spatial Sciences Conference (November). Surveying and Spatial Sciences Institute: 13–23. Available at: http://www.sssc2011.com (accessed on 23 September 2020).

Middleton SE, Middleton L and Modafferi S (2014) Real-time crisis mapping of natural disasters using social media. *IEEE Intelligent Systems* 29(2): 9–17. DOI: 10.1109/MIS.2013.126.

Miguez MG and Veról AP (2017) A catchment scale integrated flood resilience index to support decision making in urban flood control design. *Environment and Planning B: Urban Analytics and City Science*. SAGE Publications Ltd 44(5): 925–946. DOI: 10.1177/0265813516655799.

Mustafa A, Wei Zhang X, Aliaga DG, et al. (2020) Procedural generation of flood-sensitive urban layouts. *Environment and Planning B: Urban Analytics and City Science*. SAGE Publications Ltd 47(5): 889–911. DOI: 10.1177/2399808318812458.

Nicklin, Leicher, Dieperink C, et al. (2019) Understanding the costs of inaction–an assessment of pluvial flood damages in two European cities. *Water MDPI AG* 11(4): 801. DOI: 10.3390/w11040801.

NOAA (2017) *Hurricane Costs, Fst Facts*. Available at: https://coast.noaa.gov/states/fast-facts/hurricane-costs.html (accessed on 11 December 2019).

NOAA National Weather Service (2014) *AHPS Downloads*. Available at: https://water.weather.gov/ahps/download.php (accessed on 9 April 2021).

Ogie R, Holderness T, Dunbar M, et al. (2017) Spatio-topological network analysis of hydrological infrastructure as a decision support tool for flood mitigation in coastal mega-cities. *Environment and Planning B: Urban Analytics and City Science*. SAGE Publications Ltd 44(4): 718–739. DOI: 10.1177/0265813516637608.

Pereira J, Monteiro J, Silva J, et al. (2020) Assessing flood severity from crowd sourced social media photos with deep neural networks. *Multimedia Tools and Applications*. Springer 79(35): 35–36. DOI: 10.1007/s11042-020-09196-8.

Pogrebnyakov N and Maldonado E (2018). Didn't roger that: Social media message complexity and situational awareness of emergency responders. *International Journal of Information Management*. Elsevier Ltd 40: 166–174. DOI: 10.1016/j.ijinfomgt.2018.02.004.

Prasad NR, Almanza-Garcia S and Lu TT (2009). Anomaly detection computers, materials and continua. *ACM PUB27* 14: 1–22. DOI: 10.1145/1541880.1541882.

Qian Z (2010) Without zoning: Urban development and land use controls in Houston. *CitiesPergamon* 27(1): 31–41. DOI: 10.1016/j.cities.2009.11.006.

Rexiline Ragini J, Rubesh Anand PM and Bhaskar V (2018). Mining crisis information: A strategic approach for detection of people at risk through social media analysis. *International Journal of Disaster Risk Reduction*. Elsevier Ltd 27: 556–566. DOI: 10.1016/j.ijdrr.2017.12.002.

Rosser JF, Leibovici DG and Jackson MJ (2017) Rapid flood inundation mapping using social media, remote sensing and topographic data. *Natural HazardsSpringer Netherlands* 87(1): 103–120. DOI: 10.1007/s11069-017-2755-0.

Sarica GM, Zhu T, Jian W, et al. (2021) Spatio-temporal dynamics of flood exposure in shenzhen from present to future. *Environment and Planning B: Urban Analytics and City Science*. SAGE Publications Ltd 48: 1011–1024. DOI: 10.1177/2399808321991540.

Song X, Zhang Q, Sekimoto Y, et al. (2014) Prediction of human emergency behavior and their mobility following large-scale disaster. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY: Association for Computing Machinery, pp. 5–14. DOI: 10.1145/2623330.2623628.

Stieglitz S, Mirbabaie M, Ross B, et al. (2018) Social media analytics – challenges in topic discovery, data collection, and data preparation. *International Journal of Information Management*. Elsevier Ltd 39: 156–168. DOI: 10.1016/j.ijinfomgt.2017.12.002.

Stier S, Breuer J, Siegers P, et al. (2020) Integrating survey data and digital trace data: Key issues in developing an emerging field. *Social Science Computer Review*. SAGE Publications Inc 38(5): 503–516. DOI: 10.1177/0894439319843669.

Sun Q, Miao C, Duan Q, et al. (2015). Temperature and precipitation changes over the loess plateau between 1961 and 2011, based on high-density gauge observations. *Global and Planetary Change*. Elsevier 132: 1–10. DOI: 10.1016/J.GLOPLACHA.2015.05.011.

Teng JJ, Jakeman AJ, Vaze J, et al. (2017) Flood inundation modelling: A review of methods, recent advances and uncertainty analysis. *Environmental Modelling and Software* 90: 201–216. DOI: 10.1016/j.envsoft.2017.01.006.

Urbanrisklab (2017) *Urban Risk Lab*. Available at: https://urbanrisklab.org/work#/riskmap/ (accessed on 26 September 2021).

Wang RQ, Mao H, Wang Y, et al. (2018). Hyper-resolution monitoring of urban flooding with social media and crowdsourcing data. *Computers and Geosciences*. Elsevier Ltd 111: 139–147. DOI: 10.1016/j.cageo.2017.11.008.

Wang Y and Taylor JE (2019) DUET: Data-driven approach based on latent dirichlet allocation topic modeling. *Journal of Computing in Civil Engineering*. American Society of Civil Engineers (ASCE) 33(3): 04019023. DOI: 10.1061/(ASCE)CP.1943-5487.0000819.

Wilson A and Norden N (2015) The R project for statistical computing. Available at: https://www.r-project.org/ (accessed on 6 October 2020).

Wilson R, Zu Erbach-Schoenberg E, Albert M, et al. (2016) Rapid and near real-time assessments of population displacement using mobile phone data following disasters: The 2015 Nepal earthquake. *PLoS Currents (Disasters). Public Library of Science* 8. doi: 10.1371/currents.dis.d073fbece328e4c39087bc086d694b5c.

Yang Y, Zhang C, Fan C, et al. (2019). Exploring the emergence of influential users on social media during natural disasters. *International Journal of Disaster Risk Reduction*. Elsevier Ltd 38: 101204. DOI: 10.1016/j.ijdrr.2019.101204.

Yin J, Yu D and Liao B (2020) A city-scale assessment of emergency response accessibility to vulnerable populations and facilities under normal and pluvial flood conditions for Shanghai, China. *Environment And Planning B: Urban Analytics And City Science*. SAGE Publications Ltd., 48.8, 2239–2253.

Yuan F and Liu R (2018). Feasibility study of using crowdsourcing to identify critical affected areas for rapid damage assessment: Hurricane Matthew case study. *International Journal of Disaster Risk Reduction*. Elsevier Ltd 28: 758–767. DOI: 10.1016/j.ijdrr.2018.02.003.

Yuan F, Yang Y, Li Q, et al. (2021) Unraveling the temporal importance of community-scale human activity features for rapid assessment of flood impacts. *IEEE Access* 10: 1. Available at: http://arxiv.org/abs/2106.08370 (accessed on 14 August 2021).

Zhai P, Zhang X, Wan H, et al. (2005) Trends in total precipitation and frequency of daily precipitation extremes over China. *Journal of ClimateAmerican Meteorological Society* 18(7): 1096–1108. DOI: 10.1175/JCLI-3318.1.

Zhang C, Fan C, Yao W, et al (2019) Social media for intelligent public information and warning in disasters: An interdisciplinary review. *International Journal of* Information Management. Elsevier Ltd 49:190-207. DOI: 10.1016/j.ijinfomgt.2019.04.004.

Zhang X, Alexander L, Hegerl GC, et al. (2011) Indices for monitoring changes in extremes based on daily temperature and precipitation data. *Wiley Interdisciplinary Reviews: Climate Change*. John Wiley & Sons, Ltd 2(6): 851–870. DOI: 10.1002/WCC.147.

Zhong X, Duckham M, Chong D, et al. (2016) Real-time estimation of wildfire perimeters from curated crowdsourcing. *Scientific Reports. Nature Publishing Group* 6: 1–10. doi: 10.1038/srep24206.

Author biography

Hamed Farahmand is a Ph.D. student major in construction engineering and management in Zachry department of civil and environmental engineering at Texas A&M University. He is currently a research assistant in Urban Resilience.AI Lab. Hamed's work mainly focuses on characterization of flood inundation and impact on urban regions assessment of infrastructure vulnerability due to flood risk. He is interested in applying big data analytics to better understand the application of various crowdsourced and digital trace data for flood inundation assessment and developing data integration pipelines for early flood warning systems.

Wanqiu Wang is a master student major in computer engineering in Texas A&M University. She received her B.S in Electronic and Information Engineering at Dalian Maritime University. She worked as a research assistant in Urban Resilience.AI Lab, her research mainly focused on using JAVA to implement model on Anylogic and using Python to explore Harvey's influence on neighborhoods and characterization of flooding due to Hurricane Harvey using digital trace data. Her interests in machine learning applications, software development and distributed system desing.

Ali Mostafavi received the Master of Science degree in industrial administration from the Krannert School of Management and the Ph.D. degree in civil engineering from Purdue University, West Lafayette, IN, USA, in 2013. He is currently an Associate Professor with the Zachry Department of Civil and Environmental Engineering, Texas A&M University, College Station, TX, USA. He is also the Director of the Urban Resilience.AI Lab. His research interest includes creating transformative solutions for addressing the grand challenges pertaining to the nexus of humans, disasters, and infrastructure systems.

Mikel Maron is a programmer and geographer working for impactful community and humanitarian uses of open source and open data, most recently on Mapbox's COVID-19 response efforts. Mikel is a co-founder of Humanitarian OpenStreetMap Team (HOT) and leads the Community team at Mapbox. Mikel also serves on the OpenStreetMap Foundation Board. He is co-founder of Ground Truth Initiative, and of the Map Kibera project. He's previously served as Chair of Voting Members for HOT, and as President of the Board of the Humanitarian OpenStreetMap Team.