Nat Hazards (2018) 93:1529–1546
https://doi.org/10.1007/s11069-018-3366-0

ORIGINAL PAPER

Crowdsourcing for forensic disaster investigations: Hurricane Harvey case study

Faxi Yuan[1] · Rui Liu[1]

Received: 16 June 2017 / Accepted: 22 May 2018 / Published online: 4 June 2018

Abstract A critical prerequisite of risk prevention measures for natural hazards is from the results of forensic disaster investigations (FDIs). The current studies of the FDIs are limited by data issues including data availability and data reliability. The applications of crowdsourcing method in natural disasters indicate the potential to provide data support for the FDIs. However, there is very limited existing research on the use of crowdsourcing data for the FDIs. Following the requirements published by the Integrated Research on Disaster Risk program for FDIs, this paper establishes the process map for conducting the FDIs by scenario analysis approach with the crowdsourcing and crowdsensor data. Hurricane Harvey is used as the case study to implement the process map. The results show that the use of crowdsourcing data for the FDIs is feasible. Though this paper takes practical measures for improving the reliability of crowdsourcing data (i.e., little data size) in the case study, future research can focus on the development of advanced algorithm for the crowdsourcing data quality validation.

Keywords Crowdsourcing · Crowdsensor · Forensic disaster investigations · Data-driven · Scenario · Hurricane Harvey

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11069-018-3366-0) contains supplementary material, which is available to authorized users.

✉ Rui Liu
liurui@ufl.edu

Faxi Yuan
faxi.yuan@ufl.edu

[1] M.E. Rinker, Sr. School of Construction Management, University of Florida, Gainesville, FL 32611, USA

1 Introduction

The occurrence of natural extreme events shows an increasing trend from 1980 to 2017 around the world (CRED 2018). Meanwhile, the dramatic increase in population worldwide and the associated changes of mobility preference result in more human exposure to the potential natural disasters (Kron 2005). The consequence is the increase in the damage and losses during natural disasters (e.g., Guha-Sapri and Santos 2012). Furthermore, Müller et al. (2011) reported the crucial prerequisites for the development of risk mitigation measures including the analysis of natural disaster events (e.g., flood), the associated damages and the causes of damages. These requirements are consistent with the contents of forensic investigation proposed in Integrated Research on Disaster Risk (IRDR) (2011) and De Groeve et al. (2013). However, the current case studies and projects for forensic disaster investigations (FDIs) indicate their data issues including data availability and data reliability (e.g., Gotangco et al. 2013; Huang et al. 2013).

To be specific, the FDIs were introduced by Burton (2010) to explore the root causes of growing disaster losses. Four complementary modes were proposed to conduct the FDIs, including critical cause analysis, meta-analysis, longitudinal analysis and scenarios of disaster (Burton 2010; IRDR 2011). Thereafter, a series of FORIN (FORensic INvestigations of disasters) projects and case studies under IRDR (2013) and other research agencies (e.g., Zurich insurance) were performed by using these four modes. For instance, German Committee for Disaster Reduction (2012) proposed the framework for conducting the critical (root) cause analysis of damages in Haiti Earthquake. Huang et al. (2013) performed the meta-analysis of the causes of damages to infrastructure and population in Typhoon Morakot. In addition, a longitudinal analysis for the summary of existing threats to physical, social, economic and health sectors was conducted by Gotangco et al. (2013) for the recurrent climate change events in Metro Manila. A scenario analysis was performed by Menoni et al. (2016a), to explore the causes of damages to physical (e.g., critical infrastructures and buildings) and social sectors (e.g., people) in the Umbria 2012 flood. However, the common limitation in German Committee for Disaster Reduction (2012) and Gotangco et al. (2013) is their use of expert interview/consultation to identify root causes and develop the FORIN narrative for four sectors (i.e., physical, social, economic and health), respectively. This process can introduce data bias due to selection of experts and their personal preferences (Dorussen et al. 2005). Meanwhile, Huang et al. (2013) performed a solid data-supported meta-analysis, while the critical limitation is the coverage of all the data sources. For example, the 'Online Journal Paper Databases' used by Huang et al. (2013) failed to include all the important journals focusing on natural hazards, e.g., ASCE Natural Hazards Review. In the scenario analysis performed by Menoni et al. (2016a), the flood depth (i.e., hazard factor) was not used for exploring the causes of physical damages on buildings and the damage data on population at individual scale are also not collected. In summary, the common issue in the previous case studies and projects for FDIs falls on the data, including data quality (i.e., bias), coverage of data source and lack of a specific data category.

To resolve the data limitation issues involved in performing the FDIs in the previous research, this paper introduces the use of crowdsourcing data. Crowdsourcing, a method based on the faith that the aggregated public wisdom can reach an expert level quality (Eskenazi et al. 2013), is popularly used in the disaster management (e.g., Goodchild and Glennon 2010; Yuan and Liu 2018a). To be specific, crowdsourcing data demonstrate its wide applications for damage assessment (e.g., Liang et al. 2017; Barrington et al. 2011) and rescue (e.g., Riccardi 2016; HOUSTON HARVEY RESCUE 2017) during natural disasters. Therefore, crowdsourcing data indicate its potential to provide both damage data (e.g., Wang et al. 2016) and vulnerability data (e.g., HOUSTON HARVEY RESCUE 2017), which can further support the FDIs by scenario method to analyze the causes of damages. However, the existing research using crowdsourcing data for conducting the FDIs is very limited. Hence, the goal of this paper is to conduct the FDIs with a solid and complete data support. Thereafter, this research proposes a hypothesis below:

Hypothesis Using crowdsourcing data for performing the FDIs is feasible.

Hurricane Harvey hitting Texas on August 25, 2017 and bringing huge losses to the state (The Balance 2017) is used as the case study to validate the hypothesis. Utilizing the crowdsourcing and crowdsensor data produced in Hurricane Harvey, this paper performs the scenario analysis for the FDIs with the concentration on the exposed evacuees. On the one side, the damages and vulnerability data of exposed evacuees are extracted from a 3-day-life crowdsourcing platform *HOUSTON HARVEY RESCUE* (Campoy 2017), including their locations, number of children and elderly, and health conditions. On the other side, a crowdsensor map of US Geological Survey (USGS) provides this study with flood parameters (i.e., water depth). Referring to the scenario analysis in Menoni et al. (2016a), we perform our FDIs based on the data-driven approach to analyze the causes of damages to the exposed evacuees in Hurricane Harvey flood. Following the bottom-up approach, this research summarizes the causes of damages to exposed evacuees at aggravated scale from the analysis results at individual scale.

2 Related research

2.1 Forensic disaster investigations (FDIs)

On the one side, following the principle of 'build back better after disasters' proposed in UNISDR's Sendai Framework for Disaster Risk Reduction (SFDRR) (2015), research scholars have to analyze how the damages occurred during natural disasters and what the root causes have been (Menoni et al. 2016a). This is consistent with the requirements of the forensic investigation proposed in IRDR (2011) and De Groeve et al. (2013). According to the requirements, various research agencies and projects focus on the forensic disaster investigations (FDIs), e.g., Zurich insurance, CEDIM of Karlsruhe Institute of Technology and IDEA project led by Politecnico di Milano, etc.

To be specific, Zurich insurance developed the Post-Event Review Capability (PERC) method for the assessment of large disasters (Zurich insurance 2013). The PERC method focuses on the investigation of reasons resulting in the devastating impacts of the current disasters and proposes the corresponding mitigation measures for the future (e.g., Zurich insurance 2016). The critical issue involved in PERC analysis is the lack of solid data support. For instance, in the PERC analysis of Flash floods in southern Germany 2016, Zurich insurance (2016) identified one of the causes of damages as people's preference to building their houses too close to the water, while data like the number or density of buildings close to the water are not used for deriving this conclusion. In terms of forensic disaster analysis (FDA) method proposed by CEDIM (2011), previous research indicates that CEDIM's FDA focuses mainly on the estimation of potential impacts during the disasters by using the high potential of the modern empirical and analytical methodologies (Wenzel et al. 2013). The CEDIM group conducted many case studies using their FDA approaches, such as Hurricane Sandy (Mühr et al. 2012), Hurricane Matthew (Mühr et al. 2016), Hurricane/Tropical Storm Harvey (Mühr et al. 2017a) and Hurricane Irma (Mühr et al. 2017b). Their case studies analyze the hazard variables from meteorological information (e.g., precipitation and wind) and collect the exposure/vulnerability information from census data (e.g., Mühr et al. 2017a) and impacts data from public media (e.g., Mühr et al. 2012). Thereafter, they apply this information to their CEDIM model to forecast the potential impacts/damages in the disasters, which is not in the scope of 'build back better after disasters.' In the IDEA project led by Politecnico di Milano, they have performed the forensic investigations for three case studies including Umbria floods in Italy (Menoni et al. 2016b), Lorca earthquake and Vall D'Aran floods in Spain (Garcia et al. 2016) and UK floods (Ogden et al. 2016). The key limitation in their FDIs is the assignment of scores by humans for the contribution of each factor (e.g., rainfall and lack of redundancy) to the damages on the affected sectors (e.g., people and residential buildings). As the scores represent the importance of the corresponding factors, the assignment mechanism for the scores is critical to guarantee the reliability of their FDIs. However, the score assignments can vary significantly by different scorers, while the three case studies lacked the discussion on this part.

On the other side, Burton (2010) and IRDR (2011) concluded four main complementary modes to performing the FDIs, including critical cause analysis, meta-analysis, longitudinal analysis and scenarios of disaster. Meanwhile, a series of FORIN projects and case studies using these four modes were supported by Integrated Research on Disaster Risk (IRDR) and performed by various scholars such as Naruchaikusol et al. (2013), Huang et al. (2013), Gotangco et al. (2013) and Faustino-Eslava (2013). As mentioned in the introduction section, the common issue in the current FORIN FDIs falls in the data, including data quality (i.e., bias) and coverage of data source. In addition, the methods proposed in IRDR (2011) are also popularly used by other research scholars out of the FORIN project. An example is the use of scenario method by Menoni et al. (2016a) to explore the causes of damages in the 2012 Umbria floods. Their analysis of variables (i.e., hazard, exposure and vulnerability) to investigate the causes of damages to buildings lacked the critical flood parameter data (i.e., flood depth). In addition, the damage data to populations at individual scale were also not collected in their research.

Considering the data limitation issues identified in this section including data quality (i.e., bias), coverage of data source and lack of a specific data category, this paper introduces the use of crowdsourcing data for conducting the FDIs.

2.2 Crowdsourcing for disaster management

Crowdsourcing, combining the words crowd and outsourcing (Hirth et al. 2011), was firstly introduced by Howe (2006). It means 'the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, generally large group of people in the form of an open call' (Howe 2006). Meanwhile, crowdsourcing is widely used in the disaster management including the communication in the crisis (e.g., Ghahremanlou et al. 2014; Cool et al. 2015), information dissemination (e.g., Hughes and Palen 2009; Li and Rao 2010; Chatfield and Reddick 2015; Deng et al. 2016), damage assessment (e.g., Guan and Chen 2014; Kryvasheyeu et al. 2016; Yuan and Liu 2018b; Yuan et al. 2017) and rescue commitment (e.g., Yang et al. 2014). Their uses of crowdsourcing data particularly for damage assessment and rescue can provide the damage information in the affected areas (e.g., Cervone et al. 2016) and exposure/vulnerability information (e.g., location/healthy conditions) of the affected people (e.g., HOUSTON

HARVEY RESCUE 2017). The information derived from crowdsourcing data can be used as the data support for our FDIs.

In addition, CEDIM forensic disaster analysis (FDA) program has introduced the use of crowdsourcing data for FDA in their technical report on Hurricane Sandy 22–30 October 2012 (Mühr et al. 2012). They mainly used crowdsourcing data (i.e., Twitter data) for assessing the power outage during Hurricane Sandy. However, CEDIM mainly uses FDA approach for predicting the potential impacts during disasters, while the use of crowdsourcing data for the further analysis of the causes of damages (e.g., power outage) was not covered (e.g., Mühr et al. 2012).

This paper for the first time proposes the use of crowdsourcing data for performing the FDIs with following a principle of UNISDR's SFDRR (2015). However, the crowdsourcing data mainly provide the information of damages and exposure/vulnerability of exposed evacuees in Hurricane Harvey flood, while the hazard information (e.g., water depth) is not included. Therefore, the crowdsensor data (i.e., water depth at the peak stage during Hurricane Harvey) from USGS monitor system are introduced to fill in this gap. This research also for the first time integrates the crowdsourcing and crowdsensor data to construct the damage scenario for performing the FDIs.

3 Materials and methods

Aiming at conducting the FDIs with a solid and complete data support, this research employs the scenario analysis based on a data-driven approach. The process map for the implementation of the FDIs appears in Fig. 1. Each step contains the process of data. This section explains the three steps as the following.

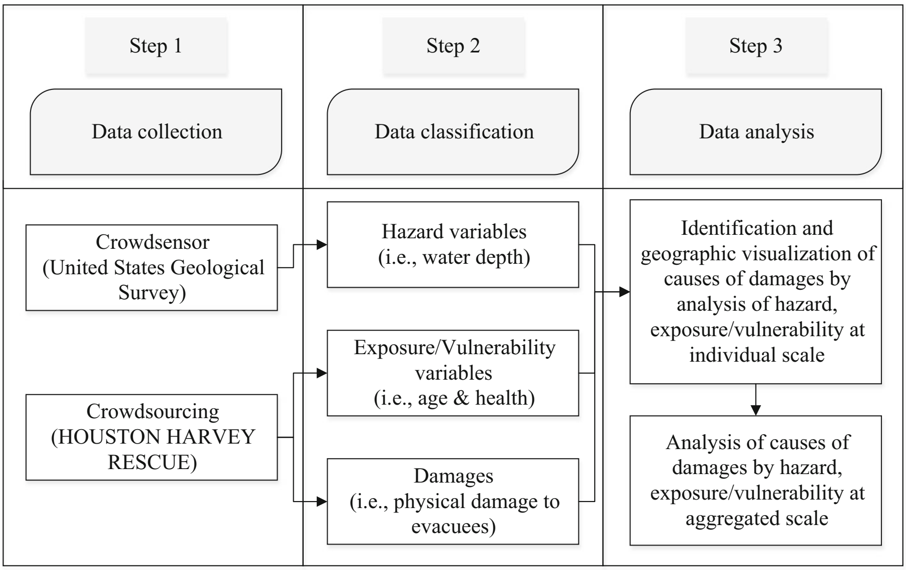

Fig. 1 Process map for the data-driven-based forensic disaster investigations

3.1 Step 1: data collection

The first step explores two sources for the data collection. The first one is a crowdsensor map of US Geological Survey (USGS). The USGS crowdsensor map provides the water depth above the ground in ft at peak stage during Hurricane Harvey (i.e., 26/08/2017–02/09/2017), and this research transfers original water depth unit from ft to meter. As the water depth is only available at the sensor locations (see Fig. 2 for the distribution of USGS sensors), we employ the inverse distance weighted (IDW) interpolation tool of ArcMap to define the water depth for the sensors' surrounding areas in Houston. The second data source is a voluntary crowdsourcing platform, *HOUSTON HARVEY RESCUE*. This platform was established to support volunteering rescue (Campoy 2017), and 7852 people were marked as SAFE thanks to 8000 voluntary rescuers through this platform (HOUSTON HARVEY RESCUE 2017). From this platform, this study extracts evacuees' reports from 1524 various locations on August 29, 2017, including their names, contacts (i.e., cell phone number), addresses, the number of elderly and children, health conditions and rescue needs. Moreover, each location may have more than one evacuee. This dataset of *HOUSTON HARVEY RESCUE* enables us to derive both exposure (e.g., location in the floods)/vulnerability (e.g., age factor by elderly and children) and damage (e.g., trapped in the floods) information of the exposed evacuees in Hurricane Harvey flood.

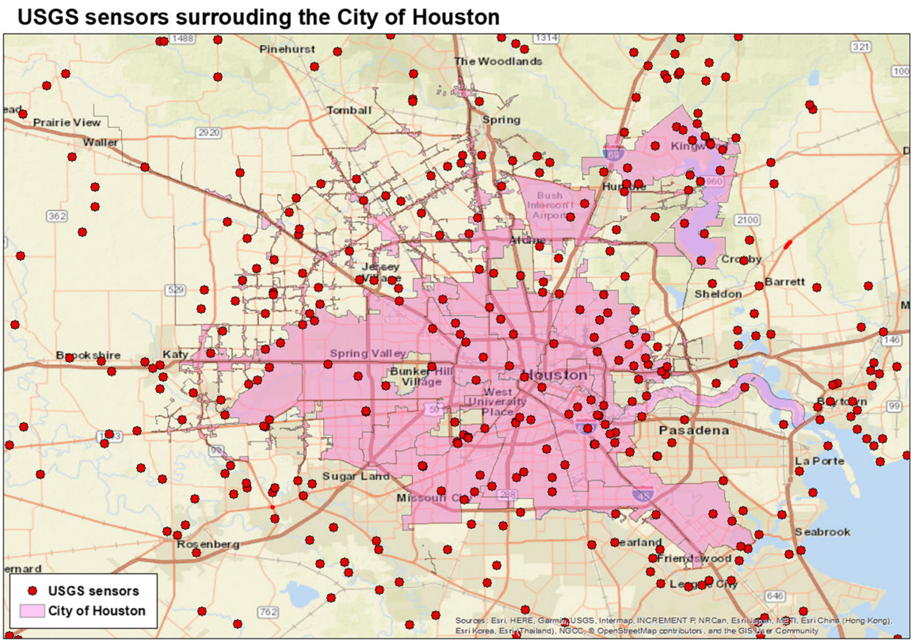

Fig. 2 The distribution of USGS sensors surrounding the City of Houston. The basemap was provided by ArcMap. The shapefile of Houston was taken from City of Houston GIS Open Data Portal. The scale of this map is 1: 420,000. The data source of basemap and shapefile of Houston and the scale applies to Figs. 3, 4, 5 and 6

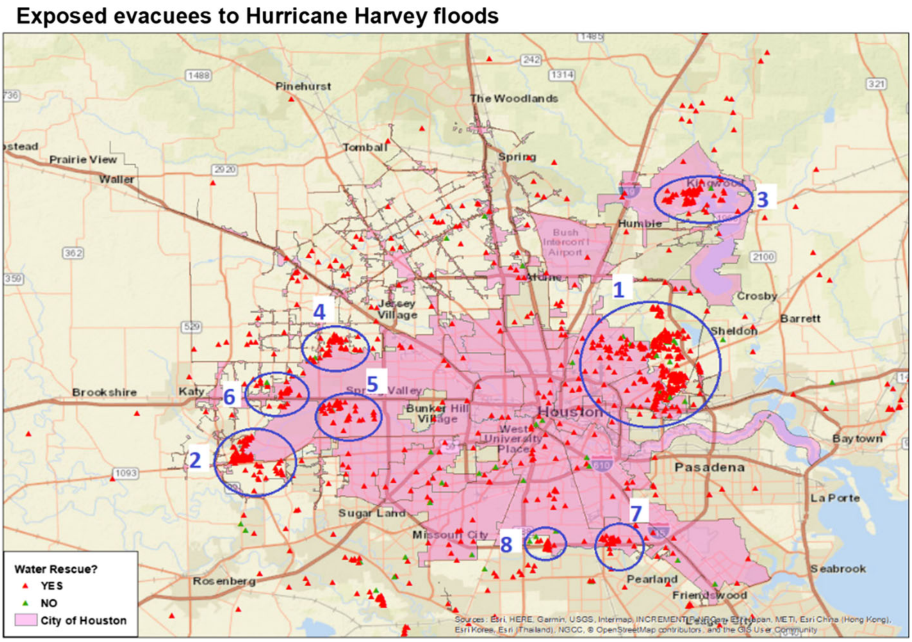

Fig. 3 The geographic distribution of exposed evacuees in Hurricane Harvey floods in Houston on Aug 29, 2017

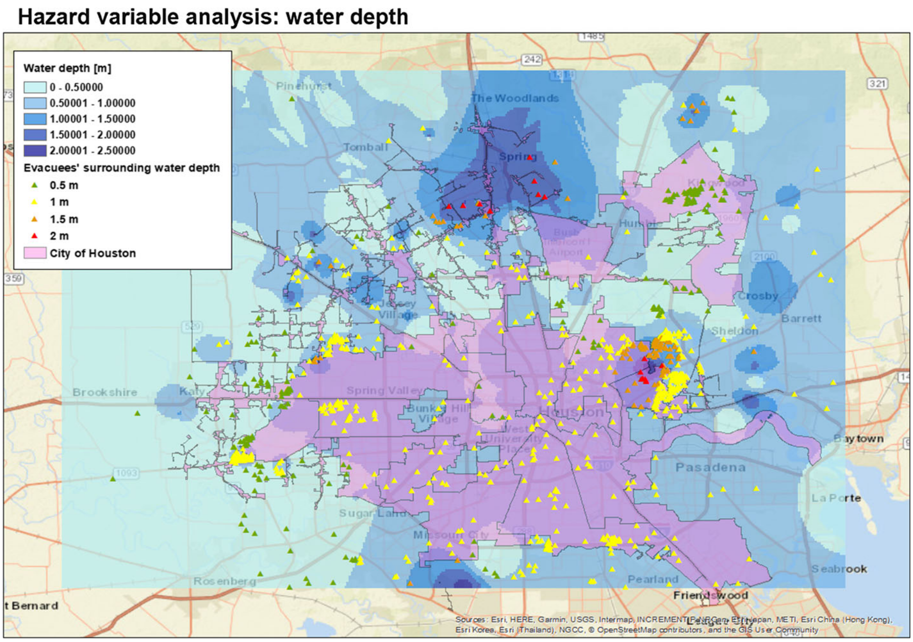

Fig. 4 The geographic visualization of hazard variable analysis in Hurricane Harvey flood at the individual scale

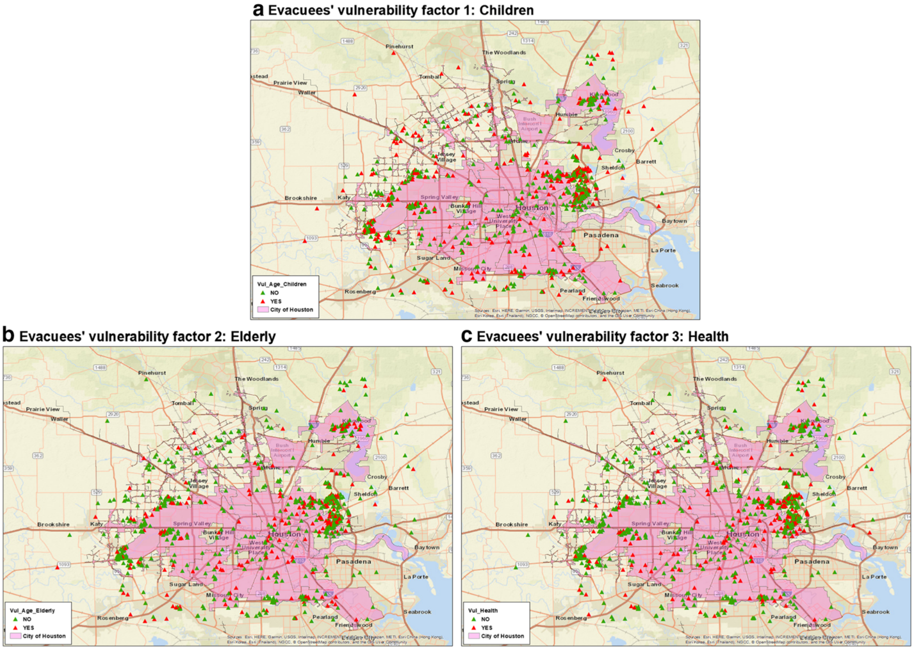

Fig. 5 The geographic visualizations of vulnerability factors including children, elderly and health issues in Hurricane Harvey flood at the individual scale

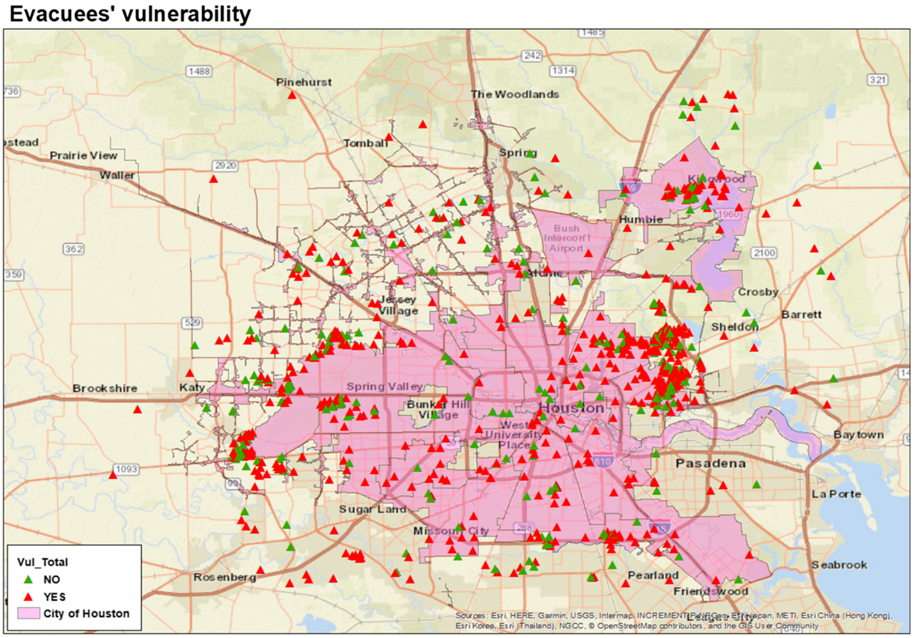

Fig. 6 The geographic visualization of whole vulnerability analysis in Hurricane Harvey flood at the individual scale

3.2 Step 2: data classification

Water depth is a common parameter used for evaluating floods hazard in previous research (e.g., Islam and Sado 2000; Pelletier et al. 2005; Ciurean et al. 2017). Hence, this research defines the water depth as the parameter to characterize the flood hazard. In terms of the exposure/vulnerability variables of the population (i.e., evacuees), this study selects age (e.g., Rahman et al. 2016; Atun and Menoni 2014) and health conditions from the dataset in *HOUSTON HARVEY RESCUE*. Furthermore, the physical damages to population mainly contain two categories, death (e.g., Menoni et al. 2016a) and affected (e.g., CRED 2018). Given the data availability, we employ the affected category to represent the physical damages to evacuees (i.e., trapped in the floods). Thereafter, this step classifies the crowdsourcing and crowdsensor data into four categories, i.e., hazard (H), exposure (E), vulnerability (V) and damage.

3.3 Step 3: data analysis

Referring to the scenario analysis approach in Menoni et al. (2016a), this step firstly analyzes the damages to exposed evacuees (i.e., population) in Hurricane Harvey flood, including their addresses and rescue needs (i.e., need water rescue or not). Secondly, this study investigates the causes of affected damages to evacuees by analysis of variables (i.e., H, E and V). This analysis starts with the investigation of causes of the damages to the individual evacuee. In addition, the geographic visualization of the variables is employed to present the analysis results. Thereafter, this research performs the aggravated-scale analysis based on the individual-scale analysis to summarize the contributions of variables (i.e., H, E and V) to the affected damages during Hurricane Harvey flood in Houston.

4 Results

4.1 Analysis of damages to the exposed evacuees

Exposure is defined as ‘the degree to which a natural or socioeconomic system or natural or socioeconomic community is exposed to potential hazards’ (Walker et al. 2011). In the practical analysis of exposure variables for the investigation of causes of damages, ‘item location’ is commonly used as an exposure factor (e.g., Menoni et al. 2016b; Garcia et al. 2016; Ogden et al. 2016). As a result, this section also conducts the exposure analysis (i.e., the geographic distribution of exposed evacuees during Hurricane Harvey flood in Houston).

On Aug 29, 2017, this study extracted evacuees' reports from 1524 various locations on the crowdsourcing platform *HOUSTON HARVEY RESCUE*. Two hundred and seven evacuees' reports are excluded from this dataset due to their uncompleted information input. Hence, the evacuees' data from 1317 locations are used for this analysis. Among the 1317 evacuees' reports on Aug 29, 2017, only evacuees from 56 locations reported ‘no water rescue’ (see the green triangle in Fig. 3), indicating that evacuees' from these 56 locations do not need water rescue. The evacuees from 95.75% of the 1317 locations needed water rescue (see the red triangle in Fig. 3) during Hurricane Harvey flood.

In addition, the 1317 evacuees' locations are presented in Fig. 3. To better describe the geographic distributions of the exposed evacuees in Hurricane Harvey flood, this section marks the blue oval circle for the eight zones with high-density evacuees. The oval circle 1 is near the intersection of US-90 and US-8, and has the largest number of evacuees requiring water rescue. The highways US-69 at the northeast of Houston and US-610 at the east of Houston were in 'closure' condition due to the high water on Aug 29, 2017 (Houston Public Media 2017). The road closures can block the evacuees trying to leave Houston on the way and further explain the reason for the highly dense evacuees in the oval circle 1. Moreover, seven other oval circles also present the highly dense evacuees. The summary of the locations of each oval circle appears in Table 1.

4.2 Analysis of variables

4.2.1 Hazard analysis

The crowdsensor map of USGS provides the water depth above the ground for the areas with their sensors (see Fig. 2 for the geographic distribution of USGS sensors surrounding City of Houston). The water depth values were recorded at the peak stage of Hurricane Harvey. As mentioned in step 1 of Sect. 3, this paper uses ArcMap to produce the flood map during Hurricane Harvey. Given the coverage areas of the flood map, this section filters out evacuees from 327 different locations. Then, the remaining 990 evacuees' locations are assigned with the water depth values according to the flood map. The geographic visualization of the flood hazard analysis at individual scale is shown in Fig. 4. According to Fig. 4, we can see that most evacuees' locations are marked with the orange triangles. This means that the evacuees' locations in the 1-m flood water depth accounts for the largest percentage in the 990 points. Moreover, the distributions of orange color triangle mainly fall in oval circles 1, 2, 4, 5, 7 and 8 as marked in Fig. 3. Meanwhile, evacuees' locations with the green triangles present a second significant portion in the 990 points. The distribution of evacuees' locations with 0.5-m flood water depth (see the green triangles in Fig. 4) mainly falls in oval circles 1 and 3 labeled in Fig. 3. Evacuees' locations in the 1.5-m flood water depth (the brown triangles in Fig. 4) are mainly distributed in oval circle 1. The evacuees' homes with the 2-m flood water depth (the red triangles in Fig. 4) account for the least percentage in the 990 locations.

Table 1 The summary of eight zones with highly dense evacuees

Oval circle	Location in Houston	Exact location descriptions
1	Northeast	The intersection of highways US-90 and US-8
2	West	The intersection of highways US-99 and Westpark Tollway
3	Northeast	The area at the intersection of highway US-69 and road FM-1960 in Kingwood
4	Northwest	The area surrounded by highway US-6, roads W Little York Rd, N Eldridge Pkwy and Clay Rd.
5	West	The area surrounded by highways US-6, I-10, US-8 and road FM-1093
6	West	The area surrounded by highways I-10, US-6, roads Clay Rd, N Mason Rd.
7	Southeast	The area surrounded by highways US-35, Sam Houston Tollway (Toll road), roads Monroe Blvd, Fuqua St.
8	South	The area surrounded by highway Sam Houston Tollway (Toll road), roads Cullen Blvd, Fellows Rd.

Karvonen et al. (2000) defined a threshold for adult humans' maneuverability and stability in flowing water. This section employs their threshold for judging whether the damages on the exposed evacuees can be explained by flood hazard variable. The condition for an adult to maneuver in good conditions of flow and environment (see Karvonen et al. 2000) presents in Eq. (1).

$$v * d < 0.006 * h * m + 0.3 \tag{1}$$

where v = flow velocity (m/s), d = flow depth (m), h = height of the adult (m), and m = mass of the adult (kg).

When the product of flow velocity and water depth is equal or larger than the values at the right part of Eq. (1), this research will judge that the evacuees at this location are affected by flood hazard. The average heights of US adult citizens are 5 feet 9 inches (i.e., 1.753 m) for men and 5 feet 4 inches (i.e., 1.625 m) for women (Data source: abcNEWS). This research takes the average value (i.e., 1.689 m) of men and women heights for the height parameter in Eq. (1). Meanwhile, data from The Telegraph (2012) report the average adult weight in the USA is 180.62 lb (i.e., 81.93 kg). Hence, the right part of Eq. (1) is 1.13.

Given that the velocity data are not available in Hurricane Harvey floods and the velocity range used in Karvonen et al. (2000) is 0.6–2.75 m/s, this research assumes the water velocity for the whole flood area is the average of 0.6 and 2.75. The velocity parameter in Eq. (1) is 1.675 m/s. Thereafter, this analysis compares the product of velocity and water depth with 1.13 (i.e., results of right part in Eq. 1 for U.S. adult citizens) to decide the influence of flood hazard variable (i.e., water depth and velocity) on evacuees. The hazard analysis results appear in Table 2 (i.e., aggregated-scale analysis).

Table 2 indicates that 741 various evacuees' locations are affected by flood hazard (i.e., water depth), accounting for 74.85% of the total 990 locations. Moreover, the number of evacuees at these 741 different locations is 4505. This means that the causes of the damages to 63.32% of the 7115 affected evacuees (i.e., total evacuees on Aug 29, 2017) can be explained by flood hazard variable (i.e., water depth).

4.2.2 *Vulnerability analysis*

Data collected from the crowdsourcing platform *HOUSTON HARVEY RESCUE* in this study includes the number of children and elderly, and 'have a health issue or not' at the 990 evacuees' locations. Given that the scenario method implemented for performing the FDIs is based on data-driven method, the vulnerability variable analysis mainly considers

Table 2 The summary of flood hazard variable analysis

Water depth (m)	Number of evacuees' locations	Number of evacuees[a] (e.g., adults, children and elderly)	Velocity * depth (m^2/s)	Affected by water
0.5	249	2610	0.84	NO
1.0	608	3087	1.68	YES
1.5	117	750	2.51	YES
2.0	16	668	3.35	YES
2.5	0	0	4.19	YES
Total	990	7115	NA	NA

[a]Each evacuee's location can have more than one evacuee

the age factors (e.g., Rahman et al. 2016) including the number of children and elderly, and health conditions. To be specific, this research checks the existence of children, elderly and health issues for each evacuee's location. The evacuees' locations are assigned with 'YES' (i.e., red triangle) or 'NO' (i.e., green triangle) to represent if they have children (see Fig. 5a), elderly (see Fig. 5b) and health issues (see Fig. 5c). Moreover, 'YES' for each vulnerability factor (i.e., children, elderly and health issue) indicates that the causes of damages to evacuees at this location are due to the corresponding vulnerability factors.

By adding these three factors, we conclude the analysis results for the whole vulnerability of the evacuees at these 990 locations (see Fig. 6). Each evacuee's location with children, elderly or health issue is marked with 'YES' (i.e., red triangle) representing that this location has a vulnerability. Similarly, 'YES' means that the causes of damages to evacuees at this location can be explained by vulnerability variables. The aggregated-scale analysis results are shown in Table 3.

Table 3 indicates that children are the most critical vulnerability factor contributing to the damages to evacuees (i.e., trapped in flood). Specifically, 46.67% of the 990 evacuees' locations were affected by Hurricane Harvey flood in Houston due to their possessions of children. In terms of vulnerability factors elderly and the health issues, the percentages are 29.70 and 29.09%, respectively. Considering the effect of the whole vulnerability, this research finds that 723 evacuees' locations present vulnerability. In total, the damages to evacuees in 73.03% of the 990 locations can be illustrated by vulnerability variables (i.e., children, elderly or health issue).

Additionally, in terms of the number of children at 462 evacuees' locations, vulnerability factor 'children' still demonstrates its leading contribution to the influences on evacuees among the two age factors. This number (i.e., 1302) accounts for 18.30% of the total 7115 exposed evacuees. The number for another age factor 'elderly' (i.e., 763) makes up for only 10.72% of the 7115 exposed evacuees.

5 Discussion

Following the principle 'build back better after disasters' in SFDRR (UNISDR 2015), this paper for the first time introduces the data-driven based scenario analysis approach to perform forensic disaster investigations (FDIs). To enable a solid and complete data support for FDIs, we also for the first time integrate the crowdsourcing and crowdsensor data. The crowdsourcing data extracted from the crowdsourcing platform *HOUSTON*

Table 3 The summary of vulnerability variable analysis

Vulnerability factors	Number of evacuees	Number of evacuees' locations with 'YES'	Number of evacuees' locations with 'NO'
Children	1302	462	528
Elderly	763	294	696
Health issue	NA[a]	288	702
Whole vulnerability	NA	723	267

[a]The crowdsourcing data from *HOUSTON HARVEY RESCUE* at each evacuee's location only provide the information 'whether this location has a health issue or not.' The number of evacuees with health issues was not reported

HARVEY RESCUE provide the damage and vulnerability information of the exposed evacuees in Hurricane Harvey flood. Meanwhile, the crowdsensor data collected from the monitoring system of US Geological Survey (USGS) enable us to perform the hazard variable analysis. By the integration of crowdsourcing and crowdsensor data, this research investigates the causes of damages to exposed evacuees at individual scale by the variable analysis of hazard (H), exposure (E) and vulnerability (V). Theater, we sum up the number of evacuees whose damages can be explained by H, E and V, and conclude their contributions to the damages on exposed evacuees at aggregate scale. The aggregated-scale analysis results can be further used for developing the causal path as presented in Fig. 7. In addition, this innovative integration of crowdsourcing and crowdsensor data can also benefit the future study in the selection of data collection methods in the future disasters.

The FDIs for the case study of Hurricane Harvey validate the process map (see Fig. 1) for conducting the FDIs by scenario analysis approach. The analysis results of the FDIs in this research enable us to accept the hypothesis proposed in introduction section, i.e., using crowdsourcing data for FDIs is feasible.

In the data-driven based scenario analysis, the uses of crowdsourcing and crowdsensor data can reduce the bias resulted from expert interview/consultation (e.g., Dorussen et al. 2005). For instance, the three case studies in the IDEA project evaluated the causes and drivers by sectors (e.g., people and properties) with scores according to their specific case studies (e.g., Umbria flood) and then classified these causes and drivers by factors of H, E and V for each damaged sector. However, the selection of the committee members for the evaluation on the causes and drivers was not discussed and the evaluation results can vary significantly by the committee with different expert members. The consequence is the data bias brought by the evaluation committee members. The crowdsourcing and crowdsensor data used in this paper are taken from evacuees' direct reports and crowdsensor system of USGS, which can avoid this problem in our FDIs.

In addition, this paper provides a new solution for the 'lack of a specific data category' issue in previous research (e.g., Zurich insurance 2016). On the one side, the crowdsensor data from USGS monitoring system provide the water depth above the ground (i.e., flood hazard parameter), a critical parameter influencing the damages to both people (e.g., Jonkman and Penning-Rowsell 2008) and buildings (e.g., Pistrika et al. 2014). This

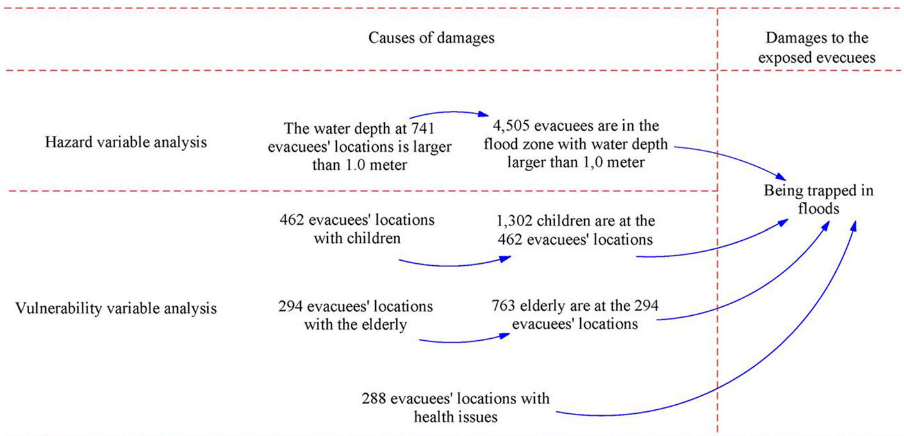

Fig. 7 The causal path for the FDIs at aggregate scale

resolves the lack of flood depth data (i.e., flood hazard factor) in the analysis by variables for the investigation of causes of building damages in Menoni et al. (2016a). On the other side, the crowdsourcing data enable this research to analyze the damages and vulnerability of individual exposed evacuees, which is not possible in most of the previous FDIs-related research (e.g., German Committee for Disaster Reduction 2012) and case studies (e.g., IDEA and IRDR projects, CEDIM FDA and Zurich insurance case studies). As a result, this research uses the scenario analysis approach for FDIs to summarize the contributions of the three variables (i.e., H, E and V) to the damages on exposed evacuees from the individual-scale data analysis. The bottom-up approach also provides a new idea for the further research in the FDIs on the selection of data scales (i.e., individual and aggregated scales).

One limitation in this research falls in the use of water depth at peak stage of Hurricane Harvey for the flood hazard analysis. To cover maximum areas in Houston with the sensors of USGS, we select data of 1169 out of 1258 USGS sensors. However, the water depth collected from the 1169 sensors (see Fig. 2 for the distribution of USGS sensors) was recorded from Aug 25, 2017, to Sep 2, 2017, which are further applied for making the flood hazard map in Houston. The damage data of exposed evacuees were mainly collected on Aug 29, 2017. Consequently, the water depth recorded after Aug 29, 2017, may not explain the damages to exposed evacuees on this day. Moreover, the use IDW interpolation tool in ArcMap can introduce the estimation errors of the water depth in the areas without sensors. Another limitation is the coverage of vulnerability parameters of people in the use of crowdsourcing data. Previous research indicates the other vulnerability parameters of people such as preparedness to disasters (e.g., Atun and Menoni 2014), annual household income (e.g., Rahman et al. 2016), level of education (e.g., Schneiderbauer 2007; Velasquez and Tanhueco 2005; Haki et al. 2004), etc. As the scenario analysis in this paper is based on data-driven method, we mainly analyze vulnerability variables on age and health conditions of evacuees (i.e., the only available data). The age parameter is represented by the numbers of children and elderly as these two kinds of people cannot deal with their evacuations during the flood by themselves (e.g., Karvonen et al. 2000). Similarly, the health condition is also considered as a vulnerability parameter in this paper with the assumption that evacuees reporting with health issues cannot evacuate independently. Future research on the use of crowdsourcing method for collecting damages and vulnerability data of people during disasters can focus on the establishment of a complete database with the attributes of all the critical exposure/vulnerability parameters (e.g., preparedness to disasters).

A common issue involved with crowdsourcing data is the data quality (e.g., Goodchild and Glennon 2010; Alexander 2014). This research pays many efforts to improve the reliability of crowdsourcing data from *HOUSTON HARVEY RESCUE*. To be specific, the original crowdsourcing data include 17 attributes: ID, date, address, name, phone, imminent threat/death or not, age, power (have or not), food (have or not), water (have or not), health issues (have or not), destination (has a place to go after rescue or not), adults (number), children (number), elderly (number), rescue status and water rescue (need or not). Hence, the exact locations of the evacuees can be tracked, which can benefit this research for the data validation. For instance, when a report from an evacuee's location says they have 100 children, we can check the building type at their exact address from Google satellite map. If the building is a common residential house, we can judge that the capacity of their house is not able to contain 100 children and the report from this evacuee's location is not reliable. Besides, this research also collects the phone number and name of the reporters at each evacuee's location during Hurricane Harvey flood in

Houston. To improve the data reliability, we can also contact the reporters for the data validation. To protect evacuees' privacy, this research will not discuss this part in detail. Though we have taken practical measures to improve the crowdsourcing data quality, the development of more advanced algorithms to achieve the automatic crowdsourcing validation is suggested for the future study.

6 Conclusion

This paper conducts the FDIs by the scenario analysis approach for Hurricane Harvey with a solid and complete data support. The process map for the implementation of the FDIs is established and further validated by the Hurricane Harvey case study. The FDIs for the first time integrate the crowdsourcing and crowdsensor data to support the analysis of damages to exposed evacuees and the investigation of the causes of the damages. Furthermore, this research summarizes the causes of damages to exposed evacuees by analysis of variables (i.e., hazard, exposure and vulnerability) from individual scale to aggregated scale. At individual-scale level, the causes of damages to individual evacuee are explained by hazard (i.e., water depth), exposure (i.e., location in the flood zone) and vulnerability (i.e., age and health condition). The individual-scale analysis results are geographically visualized in Figs. 4, 5 and 6. Thereafter, we aggregate the number of evacuees by the causes of their damages explained by hazard, exposure and vulnerability, respectively. Our results enable us to accept the hypothesis in this paper, i.e., using crowdsourcing data for performing the FDIs is feasible. Future research on the development of advanced algorithms is suggested to achieve automatic validation of crowdsourcing data.

References

abcNEWS Why Have Americans Stopped Growing Taller? https://abcnews.go.com/Technology/apples-headquarters-facilities-now-powered-100-percent-renewable/story?id=54362901. Access 16 Apr 2018

Alexander DE (2014) Social media in disaster risk reduction and crisis management. Sci Eng Ethics 20(3):717–733

Atun F, Menoni S (2014) Vulnerability to earthquake in Istanbul: An application of the ENSURE methodology. ITU J Fac Arch 11(1):99–116

Barrington L, Ghosh S, Greene M, Har-Noy S, Berger J, Gill S, Lin YM, Huyck C (2011) Crowdsourcing earthquake damage assessment using remote sensing imagery. Ann Geophys 54(6):680–687

Burton I (2010) Forensic disaster investigations in depth: a new case study model. Environ Mag 52(5):36–41

Campoy A (2017) A three-day-old crowdsourcing website is helping volunteers save lives in hurricane-hit Houston. https://qz.com/1065089/volunteers-are-coordinating-harvey-rescue-efforts-through-crowdsourcing-and-zello/. Access 16 Apr 2018

Center for Research on Epidemiology of Disasters. EM-DAT: The International Disaster Database. http://www.emdat.be/. Access 16 Apr 2018

Cervone G, Sava E, Huang Q, Schnebele E, Harrison J, Waters N (2016) Using Twitter for tasking remote-sensing data collection and damage assessment: 2013 Boulder flood case study. Int J Remote Sens 37(1):100–124

Chatfield AT, Reddick CG (2015) Understanding risk communication gaps through E-Government website and Twitter Hashtag content analyses: the case of Indonesia's Mt. Sinabung eruption. Homel Secur Emerg Manag 12(2):351–385

Ciurean RL, Hussin H, Van Westen CJ et al (2017) Multi-scale debris flow vulnerability assessment and direct loss estimation of buildings in the Eastern Italian Alps. Nat Hazards 85:929–957

Cool CT, Claravall MC, Hall JL, Taketani K, Zepeda JP, Gehner M, Lawe-Davies O (2015) Social media as a risk communication tool following Typhoon Haiyan. West Pac Surv Response J 6(Suppl 1):86–90

De Groeve T, Poljansek K, and Ehrlich D (2013) Recording disasters losses: recommendations for a European approach. JRC Sci Policy Rep. http://publications.jrc.ec.europa.eu/repository/bitstream/JRC83743/lbna26111enn.pdf. Access 16 Apr 2018

Deng Q, Liu Y, Zhang H, Deng X, Ma Y (2016) A new crowdsourcing model to assess disaster using microblog data in typhoon Haiyan. Nat Hazards 84:1241–1256

Dorussen H, Lenz H, Blavoukos S (2005) Assessing the reliability and validity of expert interviews. Eur Union Polit 6(3):315–337

Eskenazi M, Levow GA, Meng H et al (2013) Crowdsourcing for speech processing: applications to data collection, transcription and assessment. Wiley, London

Faustino-Eslava DV (2013) Predictive forensics for averting possible disasters: a FORIN template for tackling issues related to the valley fault system and the Angat Dam in Luzon, Philippines. FORIN Report. University of the Philippines, Laguna, Philippines, Los Banos

Garcia M, Predes R, Menoni S, Mendoza M, Jimanez M, Garcia-Fernandez M, Cedazo C, Mata R, Prades R (2016) Deliverable B.3: forensic investigation in the case studies in Spain (Lorca and Vall D'Aran). Technical report, IDEA project

German Committee for Disaster Reduction (Ed.) (2012) Detecting disaster root causes: a framework and an analytic tool for practitioners. DKKV Publication Series 48, Bonn. Technical report, IRDR project

Ghahremanlou L, Sherchan W, Thom JA (2014) Geotagging Twitter messages in crisis management. Comput J 58(9):1937–1954

Goodchild MF, Glennon JA (2010) Crowdsourcing geographic information for disaster response: a research frontier. Int J Digit Earth 3(3):231–241

Gotangco CK, Josol J, Padilla M, Dalupang JP, See J, Elumba R (2013) Harmonizing FORIN for climate change adaptation & disaster risk management to develop multi-sectoral narratives for Metro Manila. Technical report, IRDR project

Guan X, Chen C (2014) Using social media data to understand and assess disasters. Nat Hazards 74:837–850

Guha-Sapri D, Santos I (2012) The economic impacts of natural disasters. Oxford University Press, Oxford

Haki Z, Akyuerek Z, Duezguen S (2004) Assessment of social vulnerability using geographic information systems: Pendik, Istanbul case study. In: 7th AGILE conference on geographic information science (Heraklion, Greece, 2004), Middle East Technical University of Ankara, Turkey

Hirth M, Hoβfeld T, Tran-Gia P (2011) Anatomy of a crowdsourcing platform: using the example of microworkers.com. In: Proceedings of the 2011 fifth international conference on innovative mobile and internet services in ubiquitous computing, Seoul, Korea

HOUSTON HARVEY RESCUE (2017) Houston Harvey rescue. http://houstonharveyrescue.com/rescuesindex.php. Accessed 3 Sept 2017

Houston Public Media (2017, Aug 29) Road closures due to high water in Houston, Tuesday. https://www.houstonpublicmedia.org/articles/news/2017/08/29/233730/road-closures-due-to-high-water-in-houston-tuesday/. Access 13 Apr 2018

Howe J (2006) The rise of crowdsourcing. WIRED website. http://www.wired.com/wired/archive/14.06/crowds.html. Access 16 Apr 2018

Huang T, Hsiang-Chieh L, Hui-Hsuan Y, Chung-Sheng L (2013) Towards a generic framework for synthesising the societal disturbance from Typhoon Morakot. National Science and Technology Center for Disaster Reduction, Taipei City. Technical report, IRDR project

Hughes AL, Palen L (2009) Twitter adoption and use in mass convergence and emergency events. Int J Emerg Manage 6:248–260

IRDR (2013) Affiliated projects. http://www.irdrinternational.org/projects/affiliated-projects-3/. Access 16 Apr 2018

IRDR: Integrated Research on Disaster Risk (2011) Forensic investigations of disasters: the FORIN project (IRDR FORIN Publication No. 1), Beijing, China

Islam M, Sado K (2000) Flood hazard assessment in Bangladesh using NOAA AVHRR data with geographical information system. Hydrol Process 14(2000):605–620

Jonkman SN, Penning-Rowsell E (2008) Human instability in flood flows. JAWRA 44(5):1208–1218

Karvonen RA, Hepojoki HK, Huhta HK, & Louhio A (2000) The use of physical models in dam-break flood analysis, Development of Rescue Actions Based on Dam-Break Flood Analysis (RESCDAM). Final report of Helsinki University of Technology. Finnish Environment Institute

Kron W (2005) Flood Risk = Hazard · Values · Vulnerability. Water Int 30(1):58–68

Kryvasheyeu Y, Chen H, Obradovich N, Moro E, Van Hentenryck P, Fowler J, Cebrian M (2016) Rapid assessment of disaster damage using social media activity. Sci Adv 2(e1500779):1–11

Li J, Rao HR (2010) Twitter as a rapid response news service: an exploration in the context of the 2008 China earthquake. Electron J Inf Sys Dev Ctries 42(4):1–22

Liang WT, Lee JC, Chen KH, Hsiao NC (2017) Citizen earthquake science in Taiwan: from science to hazard mitigation. J Disaster Res 12(6):1174–1181

Menoni S, Molinari D, Ballio F, Minucci G, Mejri O, Atun F, Berni N, Pandolfo C (2016a) Flood damage: a model for consistent, complete and multipurpose scenarios. Nat Hazards Earth Syst Sci 16:2783–2797

Menoni S, Molinari D, Ballio F, Minucci G, Costantini S, Berni N, Pandolfo C (2016b) Deliverable B.2: forensic investigation in the Umbria case study. Technical report, IDEA project

Mühr B, Kunz M, Kunz-Plapp T, Daniell J, Khazai B, Vannieuwenhuyse M, Comes T, Elmer F, Schröter K, Leyser A, Lucas C, Fohringer J, Münzberg T, Trieselmann W, Zschau J (2012) CEDIM FDA-report on Hurricane Sandy 22–30 October 2012. Technical report, Center for Disaster Management and Risk Reduction Technology

Mühr B, Daniell J, Wisotzky C, Wandel J, Becker F, Buchholz M, Baumstark S, Schäfer A, Dittrich A (2016) CEDIM Forensic Disaster Analysis Group (FDA)-Hurricane Matthew 24 October 2016. Technical report, Center for Disaster Management and Risk Reduction Technology

Mühr B, Daniell J, Kron A, Jahanbazi M, Bartsch M, Raskob W, Wisotzky C, Barta T, Kunz M, Wandel J, Becker F, Latt C, Mohr S (2017a) CEDIM Forensic Disaster Analysis Group (FDA)-Hurricane/tropical storm Harvey information as of 29 August 2017. Technical report, Center for Disaster Management and Risk Reduction Technology

Mühr B, Ottenburger S, Kunz M, Wandel J, Becker F, Latt C, Mohr S (2017b) CEDIM Forensic Disaster Analysis Group (FDA)-Hurricane Irma information as of 09 October 2017. Technical report, Center for Disaster Management and Risk Reduction Technology

Müller A, Reiter J, Weiland U (2011) Assessment of urban vulnerability towards floods using an indicator-based approach: a case study for Santiago de Chile. Nat Hazards Earth Syst Sci 11:2107–2123

Naruchaikusol S, Beckman M, Mocjizuki J (2013) Disaster response and adaptive capacity of upland communities in the face of increasing climate risk. A discussion of changing livelihoods, land use, and natural-resource management in Northern Thailand. Technical report, IRDR project

Ogden R, Walliman N, Dolan M, Amouzad S (2016) Deliverable B.4: document describing forensic assessment of damage to business and the utilities sector in the case study area in UK. Technical report, IDEA project

Pelletier J, Pearthree P, House P et al (2005) An integrated approach to flood hazard assessment on alluvial fans using numerical modeling, field mapping, and remote sensing. GSA Bull 117(9/10):1167–1180

Pistrika A, Tsakiris G, Nalbantis I (2014) Flood depth-damage functions for built environment. Environ Process 1(4):553–572

Rahman MH, Aldosary AS, Nahiduzzaman KM, Reza I (2016) Vulnerability of flash flooding in Riyadh, Saudi Arabia. Nat Hazards 84:1807–1830

Riccardi MT (2016) The power of crowdsourcing in disaster response operations. Int J Disaster Risk Reduct 20:123–128

Schneiderbauer S (2007) Risk and vulnerability to natural disasters-from broad view to focused perspective. Dissertation, Freie Universiẗat Berlin

The Balance (2017) Hurricane Harvey facts, damage and costs. https://www.thebalance.com/hurricane-harvey-facts-damage-costs-4150087. Access 16 Apr 2018

The Telegraph (2012) The world’s fattest countries: How do you compare? https://www.telegraph.co.uk/news/earth/earthnews/9345086/The-worlds-fattest-countries-how-do-you-compare.html. Access 16 Apr 2018

UNISDR (United Nations Office for Disaster Risk Reduction) (2015) Sendai Framework for Disaster Risk Reduction 2015–2030. https://www.unisdr.org/files/43291_sendaiframeworkfordrren.pdf. Access 16 Apr 2018

Velasquez G, Tanhueco R (2005) Know risk. In: Proceedings of the United Nations ‘world conference on disaster reduction’, Hyogo, Japan

Walker G, Deeming H, Margottini C, Menoni S (2011) Introduction to sustainable risk mitigation for a more resilient Europe. In: Inside risk: a strategy for sustainable risk mitigation. Springer, Milano, pp. 1–22

Wang Z, Ye X, Tsou MH (2016) Spatial, temporal, and content analysis of Twitter for wildfire hazards. Nat Hazards 83(1):523–540

Wenzel F, Zschau J, Kunz M, Daniell JE, Khazai B, Kunz-Plapp T (2013) Near real-time forensic disaster analysis. In: Proceeding of the 10th ISCRAM, Baden-Baden, Germany

Yang D, Zhang D, Frank K, Robertson P, Jennings E, Roddy M, Lichtenstern M (2014) Providing real-time assistance in disaster relief by leveraging crowdsourcing power. Pers Ubiquit Comput 18(8):2025–2034

Yuan F, Liu R (2018a) Feasibility study of using crowdsourcing to identify critical affected areas for rapid damage assessment: Hurricane Matthew case study. Int J Disaster Risk Reduct 28(2018):758–767

Yuan F, Liu R (2018b) Integration of social media and unmanned aerial vehicles (UAVs) for rapid damage assessment in Hurricane Matthew. In: Proceedings of the construction research congress 2018, New Orleans, LA, USA

Yuan F, Liu R, Mejri O (2017) An information system for real-time critical infrastructure damage assessment based on crowdsourcing method: a case study in Fort McMurray. In: Proceedings of the international conference on sustainable infrastructure 2017, New York, USA

Zurich Insurance (2013) The PERC manual. https://www.zurich.com/en/corporate-responsibility/flood-resilience/learning-from-post-flood-events. Access 16 Apr 2018

Zurich Insurance (2016) Risk nexus-flash floods: the underestimated natural hazard. Technical report, Zurich Insurance Group