https://doi.org/10.5194/nhess-2022-109
Preprint. Discussion started: 8 April 2022

Natural Hazards and Earth System Sciences
Open Access
Discussions

Exploring the utility of social media data for urban flood impact assessment in data scarce cities

Kaihua Guo[1], Mingfu Guan[1,*], Haochen Yan[1], Faith Ka Shun Chan[2]

[1]Department of Civil Engineering, the University of Hong Kong, Hong Kong, 999077, HKSAR
[2]School of Geographical Sciences, University of Nottingham Ningbo China, Ningbo 315100, China

Correspondence to: Mingfu Guan (mfguan@hku.hk)

Abstract. The growing amount of social media data is a valuable and rapidly available information source to inform flood response and recovery. In this study, a workflow framework is developed to assess urban flood impacts by extracting and analysing social media data, as well as identifying the intensive public response areas, using the case of 2020 China Chengdu rainstorm-induced flooding. A crawler-algorithm is applied to extract and filter the social media data from the commonly used social platforms, namely Weibo (static data) and Tiktok (dynamic data). Based on the spatiotemporal analysis and the identified 232 flood sites with geological locations, the study shows that, social media activities and precipitation have a significant positive correlation temporally. The temporal evolution analysis of social media topics reveals the process of flooding enabling quickly to determine the severely affected areas. Spatially, social media data can give spatial flood information and social media activities are generally associated with the demographical distribution of users. Based on a flood simulation, the framework can generate reliable data source of urban flooding from social media, which can enhance flood risk modelling with the aid of hydrodynamic model. This study demonstrates the utility of social media data for urban flood assessment.

Keywords: Social media, Spatiotemporal analysis, Information extraction, Urban flooding

1 Introduction

Cities are concentrated with population, infrastructures, as well as socio-economic activities, thus they are susceptible to various types of weather extreme events, e.g., flooding and debris flows. There is a growing need for human-centric information (Loo and Leung, 2017), which is fundamental to all phases of emergency management. It is common that people's responses towards a same adverse event can be different; therefore, it is a crucial step towards reducing damage and increasing urban resilience to develop better understanding of the spatiotemporal patterns of how general public responds to a weather extreme event (Loo and Leung, 2017; Zou et al., 2018; Fang et al., 2019; Wang et al., 2020). In recent years, the rising of social media has profoundly altered the way individuals interact and communicate with each other. Social media applications can play an important role in disaster and crisis management, acting both as an information propagator for disaster relief and a valuable data source for disaster risk analysis (Wang et al., 2016).

https://doi.org/10.5194/nhess-2022-109
Preprint. Discussion started: 8 April 2022

Natural Hazards and Earth System Sciences
Open Access
EGU
Discussions

Urban flooding has become a major threat in many cities in China because of the rapid urbanization and more frequent extreme weather (Chan et al., 2018), such as 2020 Southern China widespread flooding and 2021 Henan flooding(Chik and Xue, 2021). As a result, how to deal with the increasing frequency of urban flooding has become an urgent task for city authorities (Galloway et al., 2018). However, during and shortly after a severe flooding, limited information is available about the flood extent and consequences, as well as its impact on general publics. For example, the thematic accuracy of synthetic aperture radar satellite-based damage assessment after a flood is around 75% which can be significantly lower in urban or vegetated areas. Besides, it usually has a delay of 48-72 hours due to orbital and weather conditions(Schumann et al., 2009; Havas et al., 2017). Compared to traditional flood data sources, social media is generally believed to have advantages for information extraction and dissemination, such as the ability to be searched and shared, real-time update, self-publishing capability and widespread distribution(Guan and Chen, 2014; Havas et al., 2017).

A variety of studies have used (fused) social media data together with traditional data (mainly remote sensing imagery) to estimate the flood extent and the affected areas during a flood event (Smith et al., 2017; Schnebele and Cervone, 2013; Schnebele et al., 2014; Yu et al., 2016; Eilander et al., 2016). Their results indicate that social media data can help identify affected areas faster than traditional monitoring methods (Zou et al., 2018). However, a majority of the existing studies focus on urban flooding adopt officially initiated crowdsourcing data such as governmental or social platforms (Smith et al., 2017; Assumpção et al., 2018; Yu et al., 2016), which are usually in a unified format.

However, ordinary microblog messages are brief and informal. It is usually fragmentary, laden with typographical errors, often bereft of punctuation, and sometimes downright incoherent. Thus, a reader can understand the tweet author's intent because that reader knows the context within which that tweet is being broadcast. But current computational methods are not able to make the same inferences humans do, thus cannot achieve the same level of understanding (Imran et al., 2015). Manual flittering is the best option to minimize false classification and correctly removing irrelevant messages but time-consuming. In disaster and emergency circumstances, timeliness is critically important and non-essential information filtering should be minimized. There is no standardized process for reference with only some simple rules focusing on own study case, like conflict handling of road names and search keywords, or vague filter criteria like 'irrelevant message detection' described in studies (Wang et al., 2020; Fang et al., 2019; Zou et al., 2018). With the huge volume of information posted continuously in social media about all kind of topics, how to better filter and identify valuable urban flooding data from ordinary unstructured social media data is still a challenge.

Compared to costly, delay and time-consuming traditional observation methods, social media data can afford the sheer volume, limited cost, rapid collection times and public participants. In terms of urban flood response and emergency management, social media has been used for flood trajectory detection and social response (Wang et al., 2020; Fang et al., 2019; Smith et al., 2017), emergency information extraction and diffusion for improving situational awareness (Yin et al., 2015; Jongman et al., 2015), and flood hazard modelling and mapping (Fohringer et al., 2015; Li et al., 2018; Yu et al., 2016; Rosser et al., 2017). Specifically, some studies use geo-tagged information to visualize and map people's responses to better identify risk and assess flood damages. Another group of studies focus on the textual content and attempt to identify

https://doi.org/10.5194/nhess-2022-109
Preprint. Discussion started: 8 April 2022

information of flood response and impacts. However, these studies only focus on static information, and ignore dynamic information, videos, that can provide more detailed information.

Besides, utilization of social media data alone to draw scientific conclusions is still questionable due to the many inherent issues of social media data, such as lack of validation, and biased demographic composition of users. The integration of social media data with other traditional data (such as Radar precipitation, public facilities data, and census data) are expected to provide additional information and valid spatiotemporal analysis results. More research is necessarily addressing the types of social media data synthesized with other data sources, which can inform public response to a rainstorm induced flooding and support flood risk management for stakeholders and decision makers.

In this concern, this study develops a replicable and optimized social media data processing method that is then applied to a case of Chengdu urban flooding occurred in August 2020. In-depth spatiotemporal analysis was then conducted based on flood information identified from social media. The study specifically answers two identified research questions:

- how to better filter and identify valuable flood data from a large amount of ordinary unstructured public social information full of intensive unknown interventions?
- what can datasets identified from open-source social media inform storm-flooding public response and risk management?

2 Methodology

2.1 Study area

Chengdu, located at the western edge of the Sichuan Basin in China and sits on the Chengdu Plain, is known as the ‘Country of Heaven’ and the ‘Land of Abundance’. The Jin River flows through the city, where the elevation is high in the northwest and low in the southeast, as shown in Figure 1. It is a humid subtropical plain with a catchment area of 14335 km^2 and a total population of 20.93 million in 2020. Chengdu has direct jurisdiction over 12 districts (Jinniu(a), Qingyang(b), Wuhou(c), Jinjiang(d), Chenghua(e), Longquanyi, Qingbaijiang, Xindu, Wenjiang, Shuangliu, Pidu, Xinjin), 5 county-level cities (Dujiangyan, Pengzhou, Qionglai, Chongzhou, Jianyang) and 3 counties (Jintang, Dayi, Pujiang), as shown in Figure 1. The urbanization rate of the city is about 70% and the central urban area (a, b, c, d, e) is about 598 km^2. The urban area sits in the center of Chengdu among 12 rivers, with a relatively flat terrain and elevation of about 500 m. Rainfall occurs most frequently and is concentrated in July and August, about 76%, with very little of it in the cooler months. Rainstorms are usually intense with a short duration. The maximum hourly rainfall accounts for 35% of the total rainfall amount, and even more than 45% in some locations. Given the flat terrain, dense river networks and humid climate, Chengdu has always been a hotspot of flood risk in China.

From positive guidance of local governments to autonomous publicity of tourists and residents, Chengdu has become one of the most popular internet-famous sites among young people with tags like ‘Hotpot’, ‘Panda’, ‘Relaxing Teahouse’. People are willing to share daily life via social media. In the 2020 summer, disastrous rainstorms hit the city and its surrounding

https://doi.org/10.5194/nhess-2022-109
Preprint. Discussion started: 8 April 2022

Natural Hazards and Earth System Sciences Open Access EGU Discussions

areas. The accumulated rainfall reached over 580 mm within 8 days from August 10 to 18, causing severe urban flooding. Considering its substantial impacts and extensive social media activities involved, the case of the 2020 Chengdu rainstorm and floods was chosen in this study. The disaster emergencies evolved with time and space, presenting an important research opportunity to evaluate the role of social media in disaster process tracking and impact extraction.

2.2 Data collection

The data used in this study primarily involves observed hourly precipitation data, Digital Elevation Model (DEM) of Chengdu, public facilities and social media messages (Table 1). Precipitation data are used to represent the flood process and to validate the effectiveness of social media messages. Hourly rainfall data were collected from observations of five nearby meteorological stations, as shown in Figure 1, during 2020/08/10 00:00-2020/08/19 00:00. Voronoi diagram is applied to calculate the average rainfall of the five stations, used in the temporal analysis, to characterize the overall rainfall process in Chengdu. Gauge-radar-satellite combined hourly precipitation retrieval and 1×1 km/km resolution is used to analyze the dynamic process of the storm centre in Chengdu and to conduct a flood simulation. These data were obtained from the China Meteorological Data Platform (http://data.cma.cn). The terrain data of Chengdu from ALOS World 3D - 30m (AW3D30) was adopted and the terrain filtering process was applied to generate DEM of the same resolution, which are used in the hydrodynamic modelling of inundation processes across the whole city. Besides, public facilities information was obtained from National Geomatics Centre of China (2018) (http://www.ngcc.cn/ngcc).

Social media messages were collected from Sina Weibo and TikTok, the most popular information-sharing and communication platforms that provide recreational life services to the public in China. For Sina Weibo, a user may post with a 140-character limit and is allowed to use hashtags (#) and mentions (@). In June 2020, Sina Weibo reached 523 million active monthly users. Sina Weibo has a verification policy for confirming the identity of a user (celebrities, organizations, etc.) and are distinguished with different verification symbols. Celebrity accounts receive a yellow V, called yellow VIP below; government, university, organizations and companies receive a blue V, called blue VIP below. Based on yellow V, celebrities with more than 10,000 fans and monthly view reach 10 million receive a gold V, called gold VIP below. The rest are unverified accounts. Moreover, Weibo provides a mechanism for rumour reporting and complaint handling. Compared with limited coverage SMS and private WeChat, Sina Weibo can create information sharing and interaction transcending interpersonal relationships and provide more open interface for analysis. Tiktok is a short-form video-sharing focused social networking service that has a duration from fifteen seconds to three minutes. Personal users post some videos of urban flooding situations during the Chengdu storm as well, which provides important dynamic flood information.

A Python-based web crawler is applied to obtain the social media data focusing on user ID, account classification, the time when the tweet was created, spatial coordinates and posting content including text, pictures and videos. Specifically, the tweet time was used to tabulate the check-in data into different time intervals. Spatial coordinates were used to determine the location of each tweet. The content was used for identifying whether a tweet is related to urban flooding. Based on media

https://doi.org/10.5194/nhess-2022-109
Preprint. Discussion started: 8 April 2022

reports on urban flooding and trial data, the crawling keywords was set as "暴雨" (rain/rainstorm in Chinese), "淹" (flooding/inundation in Chinese) and location limited word "Chengdu" (Figure 2).

When Chengdu and any other keywords appear in the content of a tweet, it will be judged as relevant and collected focused information. Although incorporating more keywords would result in a larger collection of flood-related tweets, a shorter list of the most relevant keywords is efficient in removing irrelevant messages (Kryvasheyeu et al., 2016). During the study period (August 10-18, 2020) 50417 tweets are collected by using this method. Based on the same keywords, 20 videos of severe inundated areas with identified locations are collected manually as well.

Table 1: Data list

Data type	Resources	Amount/Resolution
Precipitation	China Meteorological Data Platform	Hourly
Terrain elevation	ALOS World 3D (AW3D)	30 m
Public facilities information	National Geomatics Center of China (2018)	
Social media	Sina Weibo	50417 tweets
Social media	TikTok	20 video tweets

2.3 Information mining and analysis (application) workflow

The workflow and approach used in this study for disaster assessment with social media are schematically depicted in Figure 2. The workflow framework mainly contains two parts: data collection and pre-processing, spatiotemporal analysis of flooding process. In order to obtain urban storm-flooding related information in time, non-essential information filtering is minimized in the follow-up.

(1) Data collection and pre-processing

The first step is to extract storm-flooding related data (45317 tweets). There are two parts of noise that need to be reduced: commercial part and celebrity worship part.

- *Commercial reduction*: Many tweets are posted automatically on social media for financial gain, exploiting the attention that a certain hashtag has received, like #Chengdu Storm# in study. These tweets totally irrelevant to storm and urban flooding, which should be deleted.
- *Celebrity worship reduction*: These tweets are usually posted by fandom to express love or habitually share life with celebrities with the hot hashtag #Chengdu Storm# to obtain more attention. These tweets usually only state the facts without effective information like it's raining today with mentions (@) to idols, which are filtered.

The second step is to extract check-in data (3463 tweets) based on the database obtained in the first step. The check-in records referred to the spatial coordinate data of locations where users sent messages on Sina Weibo which are self-choice of users with easy operation. This step includes location detection and mundane tweets reduction.

- *Location detection*: With geographical coordinate information, check-in data outside of study area, Chengdu City, is deleted.

https://doi.org/10.5194/nhess-2022-109
Preprint. Discussion started: 8 April 2022

- *Mundane reduction*: These trivial tweets without storm-flooding information were deleted, such as only express sentiment or state facts; tweets describing the impact on daily life are retained, such as power outages, delayed flights and serious inundation, etc.; moreover, repeated (reposted) messages are ignored to avoid redundancy to some extent, as in social media platforms simply repeated messages are common even encouraged.

The last step is to identify flood-points data (248 tweets), the inundation related check-in tweets. The noise in this step is ambiguous tweets.

- *Ambiguity reduction*: Tweets that do not clearly indicate the location of the flooding and inundation information is deleted.

(2) Spatiotemporal analysis of flooding process

Based on the storm-flooding related database, the temporal analysis can be carried out to look into the overall trend of the public response and sentiment to a storm-flooding event. The amount of social media activities was calculated for each hour and day. They were further compared with hourly and daily observed precipitation data. The temporal variations of social media activities were analyzed to reveal the evolution mechanism of rainstorm and the resultant urban flooding. Moreover, the relationships between social media activities and precipitation were examined using Pearson correlation coefficient (Eq1). In addition, words frequency in the text for each day was analysed to explore the evolution of storm-flooding related topics and identify hotspots affected by the urban flooding, which are manually extracted with the aid of an on-line text analysis tool (http://www.picdata.cn/indexb.php).

$$\rho_{r,n} = \frac{cov(r,n)}{\sigma_r \sigma_n} = \frac{\sum_{i=1}^{h}(r_i - \bar{r})(n_i - \bar{n})}{\sqrt{\sum_{i=1}^{h}(r_i - \bar{r})^2}\sqrt{\sum_{i=1}^{h}(n_i - \bar{n})^2}} \quad (1)$$

where r and n are the precipitation amount and number of tweets during interval, h is the number of intervals.

The spatial analysis is based on the geographic tweets database, check-in data and flood-points data. Geographic tweets data in different districts and grids are counted to analyse the spatial distribution. Dynamic information serves as effective supplementary information to provide details for spatial analysis. Choosing an appropriate spatial unit of analysis is also important. Considering positioning error and people usually post tweets only when they have time/opportunities (e.g., after they reach the next destination), a distance of 2 km is used as grid size (Wang et al., 2020). It is a core part of any analysis about urban storm-flooding that the impact of an adverse event on people's daily life and their behavioural responses. People's activities are roughly classified into home-based, non-home based, and travel-related (Loo and Leung, 2017). They typically happen at the specific public facilities like “residential area” and “transportation network and station”. Moreover, “education area” (university & research institutes), proximity to water bodies and low-lying areas and also deserve attentions (Thumerer et al., 2000; Wang et al., 2015).

https://doi.org/10.5194/nhess-2022-109
Preprint. Discussion started: 8 April 2022

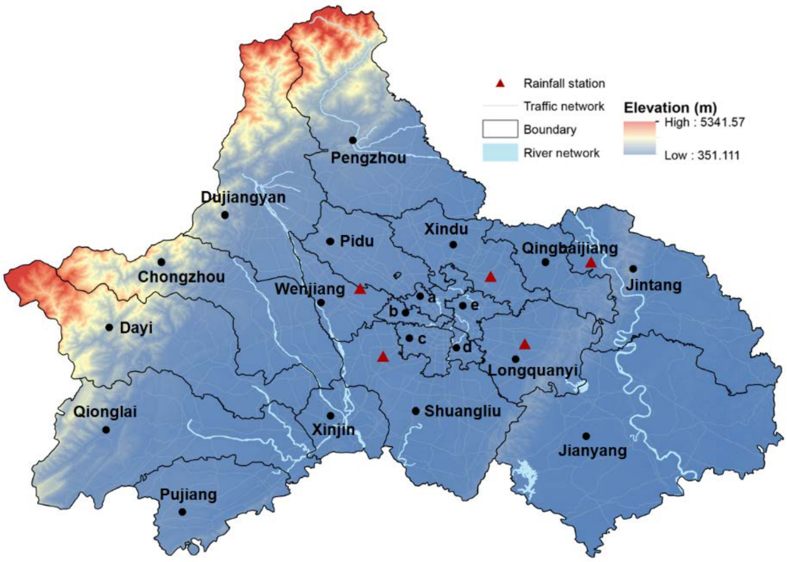


Figure 1: Basic information of Chengdu Observation site of precipitation data used in this study (a.Jinniu, b.Qingyang, c.Wuhou, d.Jinjiang, e.Chenghua).

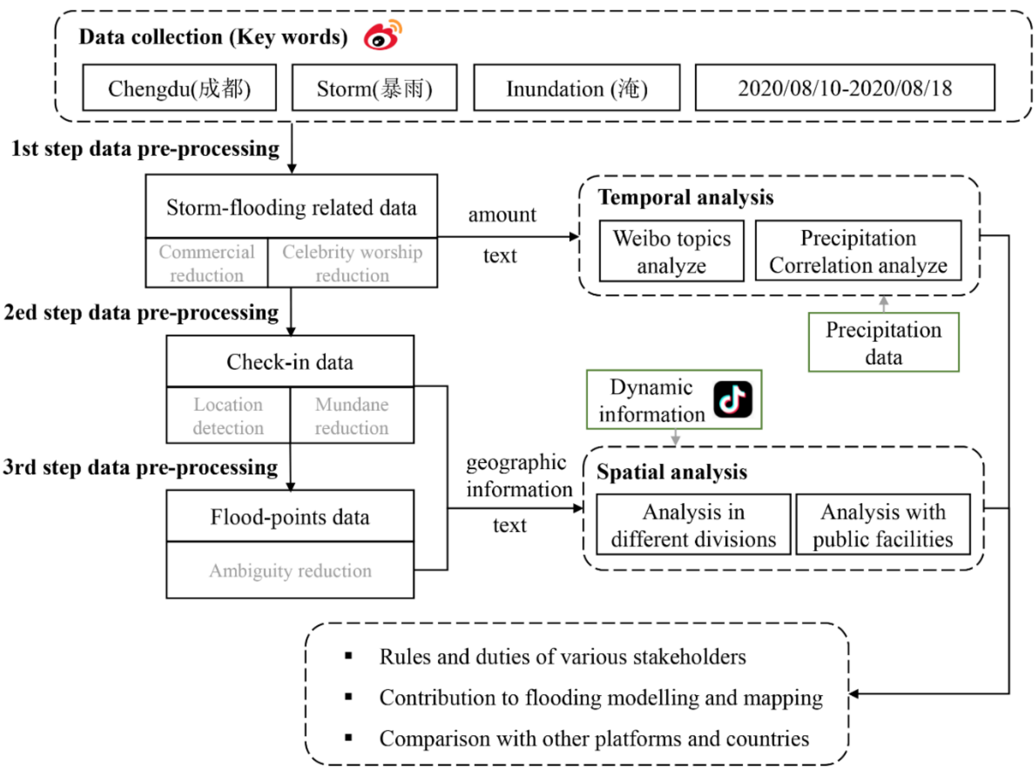


Figure 2: The workflow and procedure for collecting and analysing social media data.

https://doi.org/10.5194/nhess-2022-109
Preprint. Discussion started: 8 April 2022

Natural Hazards and Earth System Sciences
Open Access
EGU
Discussions

3 Results

3.1 Temporal analysis

3.1.1 Temporal analysis associated with precipitation

The time series of social media activities (number of tweets) and hourly precipitation from August 10 to August 18, 2020 are plotted in Figure 3. The rainstorm started from the night of August 10 and lasted till noon on August 12, causing severe impacts in Chengdu. Then the rain became small and temporally stopped on August 13. From August 14, the rain hit the city again and reached the peak on the 16th, which brought widespread flood damages. During the 36h from 12:00 on August 12 to 00:00 on August 14, the rainfall intensity was almost 0. The storm during the study period is divided into two events by 00:00 August 14, Storm 1 and Storm 2. It is seen that the variations of social media activities are generally consistent with the trend of the rainstorm process, but the response of social media had a certain lag. In other words, there were not many tweets with rainstorm topics at the beginning of the rainfall, which is understandable, especially in a humid city like Chengdu.

To further validate the similarity in the patterns of social media activities and precipitation, Pearson correlation coefficients were calculated on both original hourly data series and aggregated data on a daily basis and summarized in Table 2. For daily analysis, the correlation coefficient reaches 0.6287, showing a relatively strong correlation between rainfall and social media activities. For hourly time series, integration analysis (9 days storm duration) coefficient was small, and the highest correlation was found at the five-hour lag for the number of tweets (0.2621). Public response degree and the time to two consecutive rains were inconsistent and the two storms were analysed separately here. In Table 2, Storm 1 indicates the first storm event, and Storm 2 denotes the second storm event. For Storm 1, the highest correlation was found at the five-hour lag (0.3646), which is small as well. But it's worth noting that there was still plenty of public discussion about the storm event after the rain stopped on August 12 and August 13. Based on the content of the tweets, a video of "Dama" playing mahjong in the storm aroused the attention and discussion of people across the country.

More than 80% of tweets were discussing about this hot topic during the two days. The public were arguing about the safety issue during heavy storms. The correlation coefficient can reach 0.5889 (Storm 1 filtering) by excluding relevant discussions on this hot topic. It is seen that the hot topics or news launched by social media had a great impact on the public's attention. The elimination of hot keywords shows the public's usual degree of attention to the rainstorm event. For Storm 2, the high correlation was found at the sixteen-hour lag (0.6862). During the first peak (August 15 18:00- August 16 07:00), the public reaction was weak as continuous rainfall made people less sensitive to a rainstorm and it occurred at sleeping time. After 6am on August 16, when people typically got up and commuted from home to workplace, public attention began to increase with the arrival of another rainfall peak. These all led to a longer lag time. Moreover, Figure 4 shows the evolution of the rainstorm centre of the two storms. For Storm 1, the rainfall centre developed from a higher area, northwest, to a flatter area, southeast, while the rainfall centre of Storm 2 stayed in the flatter area, southeast. It took longer for more surface stormwater

https://doi.org/10.5194/nhess-2022-109
Preprint. Discussion started: 8 April 2022

runoff to reach the central urban area for Storm 2, which provided dominant components of social media response as shown in section 3.2.

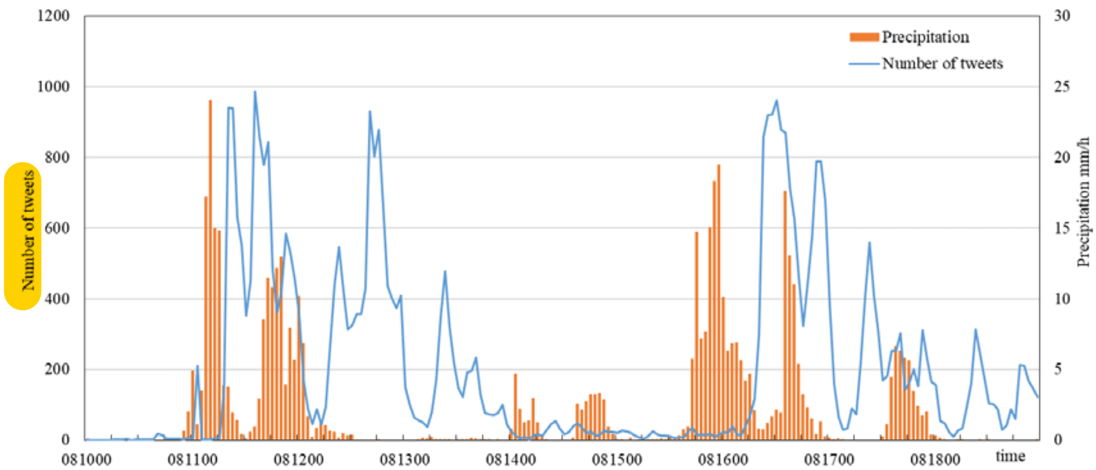
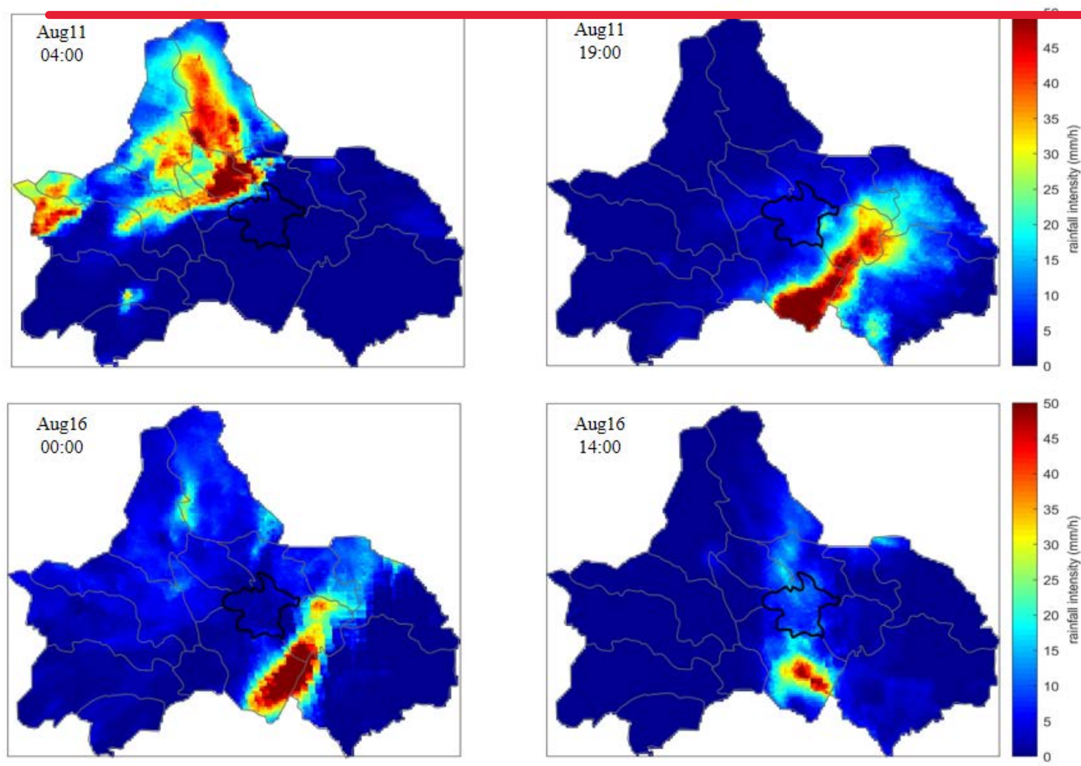


Figure 3: Temporal variations of precipitation and social media activities (number of messages) during the study period.



225 **Figure 4: The rainfall intensity distribution at different peak time of the two storm events**

Table 2: Correlation coefficients between precipitation and the number of messages for hourly time series with different time lags and for daily aggregated data.

Time lag (h)	0	1	2	3	4	5	6	7	8
Integration	0.0835	0.0606	0.0784	0.1249	0.2041	**0.2621**	0.2553	0.2457	0.2373
Storm 1	0.0444	0.0401	0.0994	0.1899	0.3213	**0.3646**	0.2753	0.2036	0.2472
Storm 1 filtering	0.2493	0.2408	0.3028	0.4001	0.5472	**0.5889**	0.4636	0.3493	0.3589
Time lag (h)	0	2	4	6	8	10	12	14	16
Storm 2	0.1454	0.0861	0.1280	0.2680	0.3169	0.3601	0.4713	0.5653	**0.6862**
Daily					**0.6287**				

3.1.2 Temporal evolution of flood-related Weibo topics

The evolution of flooding-related topics was explored through word frequency analysis of Weibo messages for each day from August 10 to August 18, as shown in Table 3. “Chengdu”, “Rainstorm”, “Weather” and “Sichuan” were the words with the highest frequency which had no analytical value and were removed. During the initial phase of each heavy rain event (August 10 to August 13), forecasting-related topics were frequently mentioned, such as “Warning/forecast”, “Issue”, “Signal”, “Thunder”, “Orange” (colour used by China Meteorological Administration to issue second-worst rainstorm alert). When the rainstorm hit the city, Weibo messages were dominated by instant impact-related topics, such as the high-frequency words of “Traffic”, “Inundation”, “Avenue” on August 12-13 and August 15-16. While when the rain weakened or temporarily stopped, words about flood damages and response activities like “Flood defence”, “Relief/rescue”, “Affected”, “Disaster” increased their frequency, as on August 16 and August 17. Hotspots affected by the flooding could also be detected. Only place-related words included in the study area, Chengdu, were analysed.

For example, Jintang, Shuangliu and Longquanyi had higher frequency as some inundation events had occurred in these areas with extensive attention on social media; and Tianfu Avenue where heavy rain paralyzed these roads. It can help rescue units to initially determine the severely affected areas. But place-related words also need to be further analyzed based on tweets content, such as Pengzhou, which attracted public attentions because the maximum rainfall was observed at the observatory in Pengzhou on August 11. In addition, social media messages can provide abundant information about disaster impacts from various aspects. Impacts on people's emotions and psychological activities can be inferred through social media messages: the words “praying, “scary”, “lose sleep” and “worried” frequently appeared in tweets.

Table 3: Top 10 topics with the highest frequency in Weibo tweets related from August 10 to 18.

Aug10 (1.3mm)	Aug11 (235.5mm)	Aug12 (39.9mm)	Aug13 (0.2mm)	Aug14 (39.9mm)	Aug15 (86.4mm)	Aug16 (122.7mm)	Aug17 (55.7mm)	Aug18 (1.4mm)
Warning 88.95%	Warning 86.76%	Dama 95.48%	Dama 88.78%	Dama 92.26%	Rainfall amount 84.77%	Inundation 91.65%	Heavy rainfall 94.21%	Snow mountain 99.01%

https://doi.org/10.5194/nhess-2022-109
Preprint. Discussion started: 8 April 2022

Observatory 87.02%	Dark as night 82.21%	Mahjong 94.06%	Peak discharge 88.64%	Calm 82.03%	Yellow warning 83.50%	Heavy rainfall 90.45%	Inundation 91.02%	Sunny 89.42%
Rainfall amount 86.74%	Thunder and lightning 81.29%	Inundation 92.78%	Inundation 85.34%	Hot discussion 81.74%	Observatory 82.64%	Tianfu avenue 82.43%	Jintang 89.30%	Inundation 75.95%
Blue warning 85.18%	Observatory 79.41%	Heavy rainfall 85.56%	Yaan 84.16%	Save 81.22%	Inundation 82.63%	Longquanyi 82.41%	Fireman 88.58%	Chengduness 76.35%
Heavy rainfall 84.71%	Orange warning 79.27%	Dark as night 81.68%	Sunny 82.93%	Flooding 80.60%	Deyang 82.54%	Warning 82.23%	Rescue 86.50%	Hilarious 73.26%
Meishan 84.59%	Yaan 78.20%	Orange warning 80.89%	Leshan 82.19%	License plate 79.31%	Guangyuan 82.20%	Shuangliu 80.26%	Pupies 84.56%	Wading 73.75%
Thunderstorm 83.83%	Heavy Rainfall 77.92%	Hot discussion 79.14%	Meishan 81.88%	Meishan 78.18%	Mianyang 82.15%	Jintang 79.79%	Traffic police 83.48%	Fireman 73.11%
Deyang 83.80%	Safty 76.34%	Observatory 78.40%	Leshan Giant Buddha 80.65%	Traffic police 76.89%	Longquanyi 78.72%	Rescue 78.75%	Water level 83.22%	Flood prevention 71.76%
Yaan 82.29%	Pengzhou 75.63%	Torrent 78.40%	Upstream 78.43%	Rainfall amount 74.56%	Yaan 77.75%	Police 81.01%	Tunnel 82.14%	Enthusiastic 70.40%
Guangyuan 82.25%	Inundation 74.44%	Flood prevention 77.35%	Water level 78.30%	Sunny 74.44%	Cloudy 77.28%	Flood prevention 78.57%	Shuangliu 80.64%	Kitty in police station 69.55%

3.2 Spatial analysis

3.2.1 Distribution analysis of spatial data in different districts

Figure 5 shows the spatial patterns of public responses towards urban flooding on Sina Weibo during the rainstorm periods. It was observed that check-in data concentrated on the central urban area and surrounding districts (Longquanyi, Xindu, Pidu, Wenjiang, Shuangliu), which accounts for 93.82% and only the central urban area alone accounts for 51.17%. As analyzed in section 3.1.2, Jintang, Shuangliu and Longquanyi suffered severe flood influence. Correspondingly, check-in data of Shuangliu and Longquanyi is accounting for 18.94% and 8.35%, respectively; while only 1.07% for Jintang. In reality, Sanhuangmiao gauge station in Tuo revier, Jintang observed the peak water level of 446.29m and the peak discharge of 7910 m3/s at 9pm 16th August. The flood return period was 50 years, which led to most area of Qingjiang Town in Jintang being flooded and seriously damaged.

https://doi.org/10.5194/nhess-2022-109
Preprint. Discussion started: 8 April 2022

Based on the word frequency analysis, there were plenty of tweets discussing the severe inundation in Jintang, but most of these tweets were published by people in other districts who concerned about the safety of Jintang people. For local people in Jintang, the proportion of tweets published was not large, and most of them were just expressing negative emotions without the ~~inclusion of~~ useful geographic information, including text and geographic coordinates. But these small amounts of data can also show the enrichment in Qingyuan Town in Jintang (red box in Figure 5a and Figure 5a), giving the correct spatial information. It is demonstrated that more severely affected areas do not mean having more check-in data but the density indicates the degree of flood severity. Besides, the districts with more check-in data include Pidu (219), Xindu (161) and Wenjiang (152), accounting for 6.32%, 4.65% and 4.32%, respectively. The closer to the central city, the more active people are to share information about urban flooding on social media platforms. This is clearly demonstrated in Figure 6a.

This study identified a total of 232 flood-points with detailed geographical locations, as shown in Figure 5b. Similar to the check-in data, flood-points concentrated on the central urban area and surrounding districts (Longquanyi, Xindu, Pidu, Wenjiang, Shuangliu), which account for 83.47% and only the central urban area alone accounts for 45.16%. As discussed in section 3.1.2, Jintang, Shuangliu and Longquanyi suffered severe inundation accounting for 10.48% (26), 16.13% (40) and 5.65% (14), respectively. Especially, Tianfu Avenue, the north-south main road through Shuangliu, was seriously flooded. Although there was not much check-in data in Jintang, most of them gave effective flooding information as 27 out of 37 check-in data points are identified as flood-points data. Flood-points data showed the enrichment in Qingyuan Town in Jintang as well (red box in Figure 5b and Figure 6b).

Figure 5a shows that 50.74% of VIP Users' check-in tweets were identified as flood-points, while only 3.86% of Ordinary Users' check-in tweets were identified as flood-points (different levels of Weibo users can be referred to in Section 2.2). Obviously with 8.59% of VIP Users' check-in tweets providing 55.24% of flood-points data, VIP Users' information is more effective. As manual filtering is very time-consuming, only paying attention to VIP Users can also get most flood-points data and effectively improve efficiency.

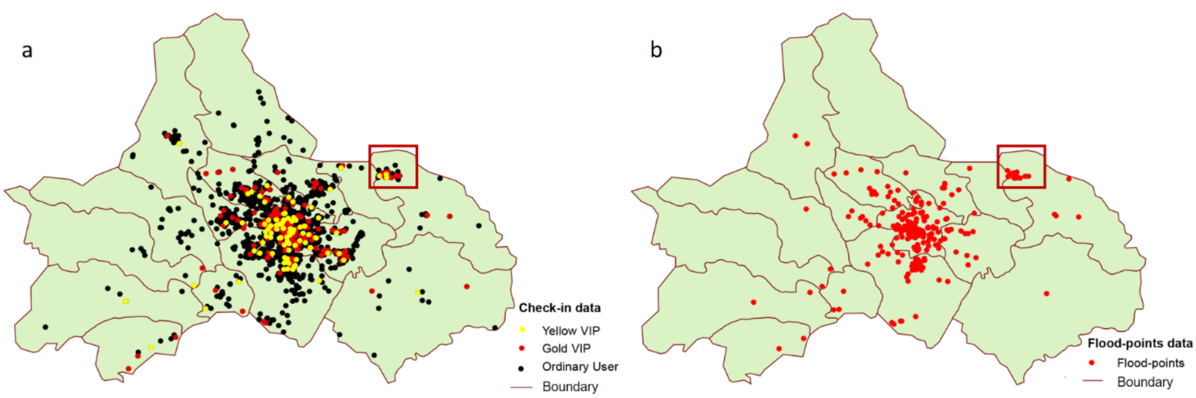


Figure 5: The geography of check-in data and flood-points data towards urban flooding in Chengdu during the rainstorm periods.

https://doi.org/10.5194/nhess-2022-109
Preprint. Discussion started: 8 April 2022

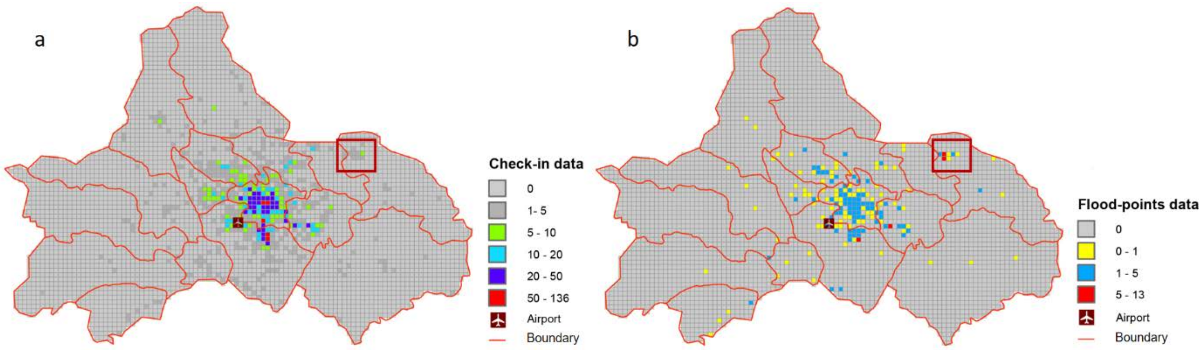


Figure 6: The enrichment distribution of check-in data and flood-points data towards urban flooding in Chengdu during the rainstorm periods.

Figure 7: Inundation condition of Tianfu Avenue in Shuangliu district which is published in Weibo (figure source: left: https://wx2.sinaimg.cn/mw2000/006ZVfWjgy1ghtq6ofykfj30u0140qb1.jpg; right: https://video.weibo.com/show?fid=1034:4538552097374239)

Table 4: The distribution of Check-in data and flood-points data in each district towards urban flooding in Chengdu during the rainstorm periods.

Check-in data	Area	Xinjin	Shuangliu	Longquanyi	Central urban	Wenjiang	Chongzhou	Qingbaijiang	Pidu
	No.	24	656	289	1772	152	22	32	219
	%	0.69	18.94	8.35	51.17	4.39	0.64	0.92	6.32
	Area	Xindu	Dujiangyan	Pujiang	Qionglai	Jianyang	Dayi	Jintang	Pengzhou
	No.	161	32	10	5	16	7	37	29
	%	4.65	0.92	0.29	0.14	0.46	0.20	1.07	0.84
Flood-points data	Area	Xinjin	Shuangliu	Longquanyi	Central urban	Wenjiang	Chongzhou	Qingbaijiang	Pidu
	No.	3	33	13	112	6	1	1	16
	%	1.29	14.22	5.60	48.28	2.59	0.43	0.43	6.90
	Area	Xindu	Dujiangyan	Pujiang	Qionglai	Jianyang	Dayi	Jintang	Pengzhou
	No.	13	2	2	2	1	2	25	0
	%	5.60	0.86	0.86	0.86	0.43	0.86	10.78	0.00

https://doi.org/10.5194/nhess-2022-109
Preprint. Discussion started: 8 April 2022

3.2.1 Distribution analysis of spatial data with public facilities

Figure 8 shows the spatial patterns of public facilities, including train stations, bus stations, residential area and traffic network, and spatial data extracted from Weibo tweets. It is clearly shown that the areas with relatively lower elevation had higher concentrations of check-in data and flood-points data. This outcome is reasonable since people living in areas with a lower altitude tend to show a higher level of urban flooding concern. As Chengdu has the mountainous and hilly bed terrain on both west and east sides, the surface stormwater runoff accumulates in the middle flat area where the central urban area is located. This explains the cause of more check-in data in the areas with lower elevation. Besides, densely-populated areas tend to have more check-in data and flood-points data. The residential area point data provide the coordinate information of the entrance of the community and building area is usually 0.3-3.0 km^2 (Zhou et al. 2016). The inundation location may be only within the visual range and observers do not need to be at the exact points.

Therefore, if check-in data and flood-points data is within 3 km^2 of the residential area data, it is deemed as check-in data and flood-points data coincide with the residential area. The coincidence rate of check-in data and flood-points data were quantified to be 79.34% and 87.09%, respectively. In other words, areas without residential communities have almost no reported storm-flooding related information. As expected, population density is a significant factor affecting public social media response, which coincides with the high number of tweets in the densely populated central city.

In terms of the traffic aspect, there is no concentration of spatial data around bus stations and train stations. Based on the text information, there are 62 tweets (tiny proportion) with check-in information near train stations, and they are all located in nearby hotels. Based on the tweets postponed travelling to Chengdu influenced by rainstorm, it is speculated that for land transportation, most people would choose to cancel the schedule in adverse weather without causing passengers stranded at these stations. Figure 9 shows the spatial data of the central urban area with more details. The social media data in subways, the special transportation in central urban area, is also scarce. Only 21 check-in tweets and 20 flood-points data relate to subway were identified.

That is interesting to note that amount of check-in data around the airport were reported in social media because of the cancellation and delay of flights. In addition, there is not much check-in data and flood-points data in proximity to water bodies, except for the Tuo River where overtopping embankment occurred. In other words, residents near the water body in Chengdu did not show greater concern about upcoming flooding than residents in other places. Last but not least, we do not find much social media reports in such education areas as university and research institutes. This probably because the rainstorm occurred in the middle of August, the summer vacation period, thus almost no students stayed on campus.

https://doi.org/10.5194/nhess-2022-109
Preprint. Discussion started: 8 April 2022

Natural Hazards and Earth System Sciences
Open Access
Discussions

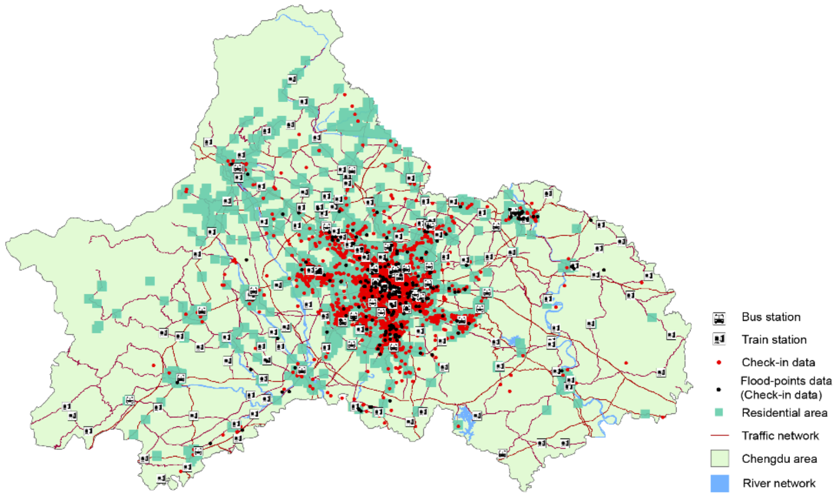


Figure 8: The distribution of spatial data towards urban flooding during the rainstorm periods and public facilities in Chengdu.

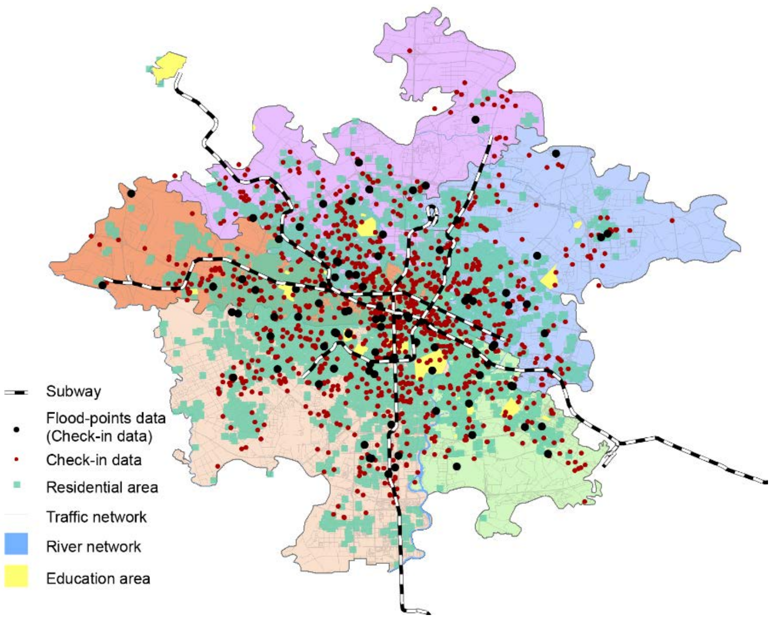


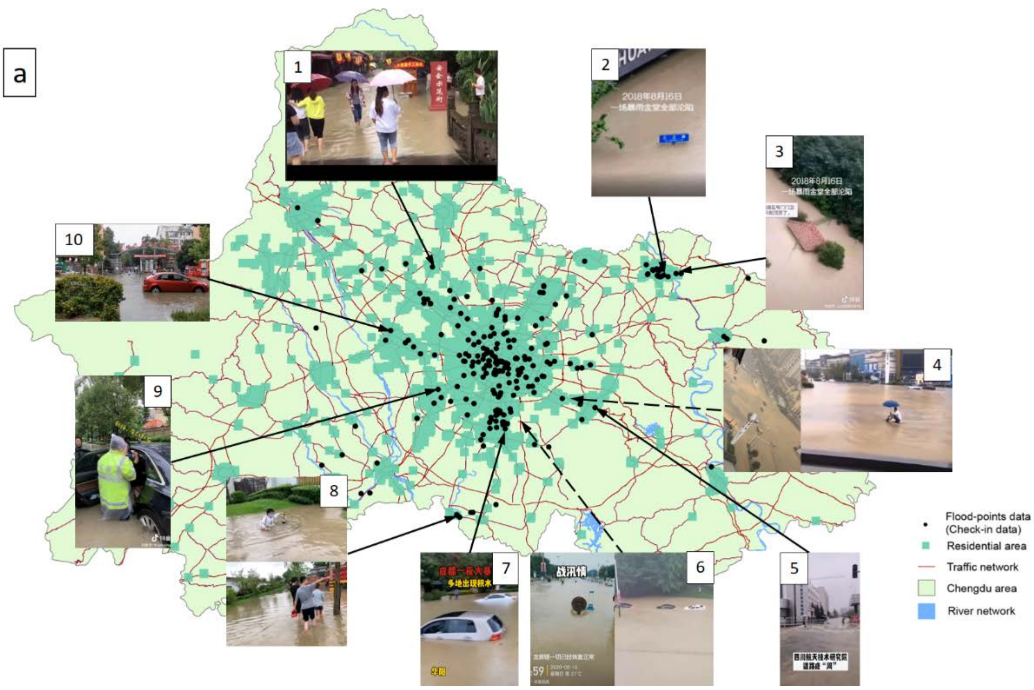
Figure 9: The distribution of spatial data towards urban flooding during the rainstorm periods and public facilities in central area.

3.3 Dynamic information

Compared to the static information from Sina Weibo platform, dynamic information collected from Tiktok is more repetitive and covers less locations. But it can provide extra details with flood videos. Based on the 20 videos with specific geographic information, 10 flood points are identified and shown in Figure 10(a) and detailed annotation of available information are shown in Figure 10(b), which are taken from video screenshots. And the video's details can be accessed in the supplement. Geography of these videos are all implicit references: text in video or landmark like blue box in Figure 10(b). 8 of them

https://doi.org/10.5194/nhess-2022-109
Preprint. Discussion started: 8 April 2022

coincide with the flood points extracted from Weibo (solid arrows) and the remaining 2 provide new inundation location (dashed arrows). Water level, flood range, flow direction, and local condition of an inundated area can be observed based on the dynamic information shared. Water level can be estimated based on a reference point, which can be vehicles, human bodies or public facilities. For instance, the water level in photo 5 of Figure 10(a) can be estimated by a human body, which is approximately below the knee position of an adult man and about 30-40 cm marked in Figure 10(b). The water level can also be estimated by vehicles as shown in photo 4 of Figure 10(A), where the seat of the e-bike almost inundated, so the water depth is about 60-70 cm marked in Figure 10(b). The water depth in photo 6, 7 ,9 and 10 also can be estimated by submerged depth of a car. The inundation range can be identified as well, as shown in Figure 10(b) and the inundated area is surrounded by the red lines. The water level can be identified by public facilities as well, such as street sign in photo 2 and guard pavilion in photo 3. Although obtaining detailed information from dynamic information is uncertain, as it is subjective and estimated based on reference points, this is a rare source of data and should not be ignored. The flow direction can be determined as in the video of Figure 11b. The local condition of an inundated area can be observed. For instance, the sewer manhole covers on Yidu West Road were all opened to shorten the duration of inundation showing in left part of photo 6. Traditional data hardly provide such details, but they have a great impact on local hydrodynamics. The synthesis of static information and dynamic information could be a strong support to scientific researchers and software developers on flooding modelling and mapping.



https://doi.org/10.5194/nhess-2022-109
Preprint. Discussion started: 8 April 2022

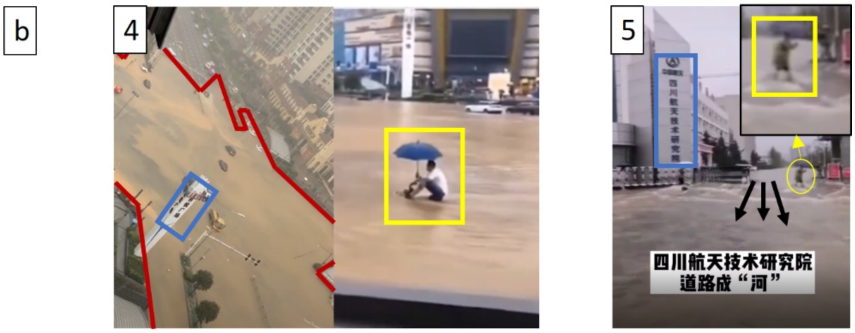


Figure 10: (a) Details condition of flood points collected from TikTok (video source: 1: https://v.douyin.com/NpDgNct/; 2: https://v.douyin.com/NpD6gMP/; 3: https://v.douyin.com/NpD6gMP/; 4(left): https://v.douyin.com/NpDqcve/; 4(right): https://v.douyin.com/NpAm3A2/; 5: https://v.douyin.com/NpDjb9w/; 6(left): https://v.douyin.com/NpAKoPL/; 6(right): https://v.douyin.com/NpAcyU4/; 7: https://v.douyin.com/NpDuQhn/; 8: https://v.douyin.com/NpDgNct/; 9: https://v.douyin.com/NpDahF7/; 10: https://v.douyin.com/NpDqcve/) and (b) annotation of available information of photo 4 and 5: geography (blue box), water level (yellow box), flood range (surrounded by red lines) and flow direction (black arrows).

4 Results

4.1 Rules and duties of various stakeholders from the lens of social media data

The spatiotemporal analysis above reveals that social media platforms are a valid and useful source of information throughout a flood lifecycle. In view of the inevitable gaps between those who publish the information and those who put it into use, speedy and concerted efforts from all stakeholders are indispensable. For people affected social media can provide more opportunities to communicate with the outside to obtain a greater sense of security. People can publish first-hand flood information in near real-time to those outside the impact zone and rescue organizations, which is of great value to help timely warning and rescue. For formal response agencies, flood-point data filtered from the check-in data have great guiding significance during a flooding as focusing on VIP accounts can greatly improve collection and pre-processing efficiency and achieve real-time feedback. Besides serve rescue operations can be supported by the in-time data source from social media, such as coordinate information, the depth and area of inundation area and the status of the trapped people by pictures or videos. Successful rescue videos and timely responses published either by the local public or the response agencies are of great encouragement to the greater public. That would be valuable for decision makers to incorporate flood information posted on social media platforms into flood risk management processes. The general public reactions can be spotted quickly by temporal evolution analysis of related topics, which can help decision-makers to detect severe inundation areas and better coordinate/ deploy the rescue resources with public's needs in a timely manner. The emerging topics during and after the rainstorm can also serve as a reference for the direction of essential safety education by the government.

For example, the hot discussion of the news in Chengdu, such as Dama playing mahjong in the storm and swimming on the flooded avenue, all demonstrate the lack of public crisis awareness facing urban flooding. In the meantime, the effectiveness of the social media data largely hinges on the clarity of the location information. Geographical coordinates are thus highly

https://doi.org/10.5194/nhess-2022-109
Preprint. Discussion started: 8 April 2022

Natural Hazards and Earth System Sciences
Open Access
EGU
Discussions

encouraged to attached as it is useful for a number of tasks in disaster response, the visualization of information on a map and possibly making it more actionable for emergency responders, which is easy to operate and more effective than implicit references to names of places (Imran et al., 2015).

4.2 Social media data contribution to flood modelling and mapping

Flood-point data provide valuable in-situ information, contributing to a reliable resource for flood modelling and mapping. Most models require large amounts of data, both for model establishment and model usage in order to have an adequate representation of urban flooding. In view of its capability of collecting and processing massive and unstructured datasets, the present workflow could provide a scalable solution for the lack of flooding-related data during simulation. Specifically, the data types serve to simulate different stages: model setup, calibration and validation.

Based on the dynamic information shared, the local condition of an inundated area, such as blocked sewer manhole cover, backflow from pipe or sediment accumulation, can be observed to update the model setup. This has a great impact on local hydrodynamics. Flood model calibration and validation can also be supported by the more detailed information extracted from social media, such as the geotagging information of flood-points data, water level, flood extent and speed. The timestamp is considered to be vital in urban flood simulation as well (Kutija et al., 2014).

This study demonstrates the value of the filtered flood-points data in supporting large-scale flood modelling. A two-dimensional full shallow water-based inundation model with OpenMPI-accelerated approach is applied and it has been tested and verified with a range of boundary conditions in a number of water environments (Guan et al.,2013, 2017). The terrain data with 30m resolution and gauge-radar-satellite combined precipitation mentioned is used. The infiltration and drainage network and associated sewers are not explicitly considered within the model because of unavailability and uncertainty of suitable data regarding the grates and gullies. More crucially because its effects and quality of operation during an extreme rainfall are likely to be minimal especially suffering a heavy rain just before the study period.

A uniform Manning coefficient across the domain is assumed to be 0.035 $s/m^{1/3}$ to partially compensate for street furniture which is neglected, as well as the mixture of surfaces that include long grass and woodland, paving stones, and asphalt. The spatial distribution of inundated water depth (>0.05m) at the peak time (7pm on 16th August) is illustrated in Figure 11 (greater Chengdu area) and Figure 12 (Chengdu city). Notedly, overtopping embankments in Jintang is reproduced by the model, and flood-points data captured from Weibo is indeed concentrated in the flooding area, as shown in the black box in Figure 11.

By considering positioning deviation, buffered area with a radius of 200m is created to extract relevant information of corresponding flood-points data from the output raster files. The results demonstrate that 71% of the flood-points data with buffered area average simulation water depth greater than 0.05m and 95% of the flood-points data with buffered area maximum simulation water depth greater than 0.03m. The simulation result is acceptable and also verifies the quality of social media data from the side. In addition, as areas without residential communities have almost no reported storm-flooding related information, inundation to be inferred elsewhere in the city with increased detail and accuracy is allowed with the framework coupled with hydrodynamic modelling.

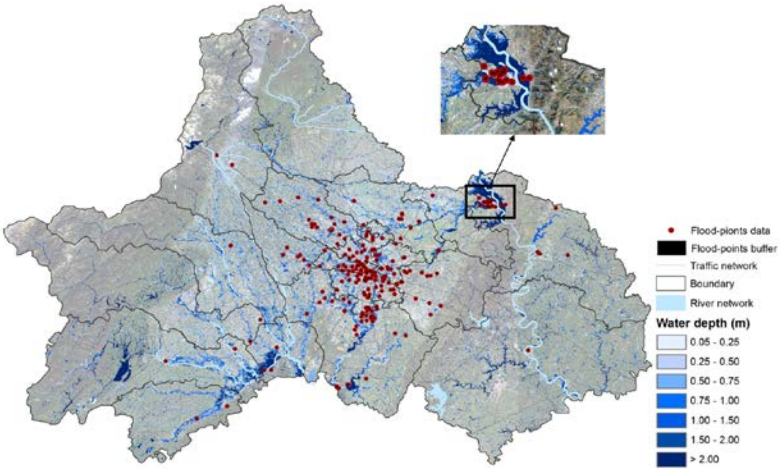


Figure 11: The distribution of inundated water depth at the peak time (7pm on 16th August) of the flooding area (>0.05m) in Chengdu.

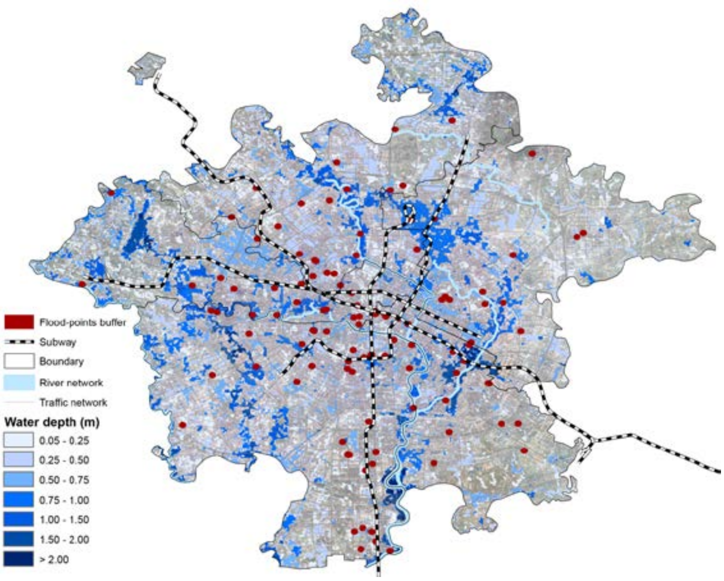


Figure 12: The distribution of inundated water depth at the peak time (7pm on 16th August) of the flooding area (>0.05m) in central urban area of Chengdu.

4.3 Comparison with other platforms and countries

Effective information management and easily accessible communication practice play a key role to identify urban flooding risk and to capture public responses (Unisdr, 2009). Many open source platforms have been designed to collect and proliferate storm-flooding information with a prescriptive syntax and generate urban flooding maps based on social media such as Twitter and Flickr, the platform “Tweak the Tweet”, “Twitter Alerts”, “GeoFeedia” and many others (Thumerer et al., 2000; Eilander et al., 2016; Jongman et al., 2015; Sun et al., 2016). Weibo did not launch a similar system to integrate storm-inundated related information during the Chengdu storm event.

However, some actions were taken during the Henan storm event in the late July, 2021, such as the launch of the hashtag #Henan Storm Mutual Aid# to uniformly publish trapped-rescue information. Such kind of initiatives are valuable and needs to be promoted. Further, it is worthy to promote a more standardized and detailed platform, actively collecting and categorizing information with time, location, account, the coordinate system, and language, =to increase the accuracy of the social media data spatially and temporally.

PetaJakarta, an app that automatically collects geotagged photos of flooding observations from Twitter users in Jakarta, Indonesia, has been widely cited as a model of community engagement in urban flooding response (Wang et al., 2020). Similarly, it would be highly valuable to establish a well-organized social media based urban flooding database based on Weibo to extract and share more quantitative information of each event. Collaboration with industries can offer more flexible and reliable service. The Swiss government collaborated with national building insurance association and cantonal insurers to launch a free, nationwide, interactive online map detailing areas with potential urban flooding threat. Nowadays, urban command centers have been set up to collect and analyze different kinds of big data in many big cities in China. It is possible for these smart city control centers to include many potential data sources from industries and general public (Zhong et al., 2010).

Given that the most negative effects of urban flooding tends to fall on traffic-related activities, social media can cooperate with map services companies and meteorological departments to integrate real-time rainfall, transport information (e.g., road systems and traffic flows), urban flood information and other relevant data into real-time digital maps (Loo and Wang, 2017). Previous study suggested that public and the communities enabled the flood risk perception, understanding and awareness by social media effectively particularly on the flood zones, in prior improving the flood resilience (Wang et al., 2020). That is the evidence to reduce urban flood risk for the communities in Chinese cities. We recommended the municipal governments should take a leading role in the management and operation of cities by establishing clear guidelines for the industry and public in making appropriate responses.

5 Conclusion

This study investigated the utility of social media in urban flood assessment by using the case of 2020 China Chengdu rainstorm-induced flooding. We presented an efficient workflow to collect, process and identify unstructured flood related data in near real-time during a storm event. 232 flood sites with geological locations are identified in Chengdu 2020 storm. Based on identified social media database, the spatiotemporal analysis of the urban flooding process and general public response were conducted to support risk management. This study shows that social media data can provide valuable spatial and timely information on public responses towards urban flooding for use by local governments in emergency management.

The study reveals that temporally social media activities have a significant positive correlation with precipitation. And the severely affected areas can be effectively determined through the temporal evolution analysis of storm-flooding related topics. Spatially, social media data can provide spatial flood information and valuable in-situ information. More severely affected areas do not mean having more social media data, which are generally associated with the demographical distribution of users, but the density can indicate the degree of flood severity.

In addition, the flexibility of train and bus traffic leads to less social media data in relevant stations than airports where social media events are more proactive. Numerical simulation of the flooding indicates that the proposed framework is able to obtain excellent extra valuable data of urban flooding from social media to support flood dynamic modelling. The quality of social media data can be further improved by initiating campaigns in social media platforms such as a dedicated campaign during a specific flood event.

In the future, with the establishment of a more standardized collection-categorization platform and concerted efforts from all stakeholders, social media data can be more effectively incorporated into the flood risk management process.

Authorship contribution statement

KG contributed to the social media data processing and simulation and prepared the original manuscript. MG acquired the research funding, conceptualized and developed the study and investigation, and edited the manuscript. HY contributed to the terrain data curation and validation. FC also reviewed the manuscript.

Declaration of competing interest

The authors declare that they have no conflict of interest.

Acknowledgement

The study is financially supported by Early Career Scheme from the Hong Kong Research Grants Council (grant number: 27202419).

References

Assumpção, T. H., Popescu, I., Jonoski, A., and Solomatine, D. P.: Citizen observations contributing to flood modelling: Opportunities and challenges, Hydrology and Earth System Sciences, 22, 1473-1489, doi:10.5194/hess-22-1473-2018, 2018.

Chan, F. K. S., Griffiths, J. A., Higgitt, D., Xu, S., Zhu, F., Tang, Y.-T., Xu, Y., and Thorne, C. R.: "Sponge City" in China—a breakthrough of planning and flood risk management in the urban context, Land use policy, 76, 772-778, https://doi.org/10.1016/j.landusepol.2018.03.005, 2018.

China floods: how Zhengzhou's 'once in a thousand years' rainfall compares: https://www.scmp.com/news/china/science/article/3142157/china-floods-how-zhengzhous-once-thousand-years-rainfall, last access: 22 July 2021.

Eilander, D., Trambauer, P., Wagemaker, J., and Van Loenen, A.: Harvesting social media for generation of near real-time flood maps, Procedia Engineering, 154, 176-183, https://doi.org/10.1016/j.proeng.2016.07.441, 2016.

Fang, J., Hu, J., Shi, X., and Zhao, L.: Assessing disaster impacts and response using social media data in China: A case study of 2016 Wuhan rainstorm, International journal of disaster risk reduction, 34, 275-282, https://doi.org/10.1016/j.ijdrr.2018.11.027, 2019.

Fohringer, J., Dransch, D., Kreibich, H., and Schröter, K.: Social media as an information source for rapid flood inundation mapping, Natural Hazards and Earth System Sciences, 15, 2725-2738, doi:10.5194/nhess-15-2725-2015, 2015.

Galloway, G. E., Reilly, A., Ryoo, S., Riley, A., Haslam, M., Brody, S., Highfield, W., Gunn, J., Rainey, J., and Parker, S.: The growing threat of urban flooding: A national challenge, College Park and Galveston: University of Maryland and Texas A&M University, 2018.

Guan, X. and Chen, C.: Using social media data to understand and assess disasters, Natural hazards, 74, 837-850, doi:10.1007/s11069-014-1217-1, 2014.

Havas, C., Resch, B., Francalanci, C., Pernici, B., Scalia, G., Fernandez-Marquez, J. L., Van Achte, T., Zeug, G., Mondardini, M. R. R., and Grandoni, D.: E2mc: Improving emergency management service practice through social media and crowdsourcing analysis in near real time, Sensors, 17, 2766, doi:10.3390/s17122766, 2017.

Imran, M., Castillo, C., Diaz, F., and Vieweg, S.: Processing social media messages in mass emergency: A survey, ACM Computing Surveys (CSUR), 47, 1-38, https://doi-org.eproxy.lib.hku.hk/10.1145/2771588, 2015.

Jongman, B., Wagemaker, J., Romero, B. R., and De Perez, E. C.: Early flood detection for rapid humanitarian response: harnessing near real-time satellite and Twitter signals, ISPRS International Journal of Geo-Information, 4, 2246-2266, doi:10.3390/ijgi4042246, 2015.

Kryvasheyeu, Y., Chen, H., Obradovich, N., Moro, E., Van Hentenryck, P., Fowler, J., and Cebrian, M.: Rapid assessment of disaster damage using social media activity, Science advances, 2, e1500779, doi: 10.1126/sciadv.1500779, 2016.

Kutija, V., Bertsch, R., Glenis, V., Alderson, D., Parkin, G., Walsh, C., Robinson, J., and Kilsby, C.: Model validation using crowd-sourced data from a large pluvial flood, 11th International Conference on Hydroinformatics, New York, USA2014.

Li, Z., Wang, C., Emrich, C. T., and Guo, D.: 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, 45, 97-110, https://doi-org.eproxy.lib.hku.hk/10.1080/15230406.2016.1271356, 2018.

Loo, B. P. and Leung, K. Y.: Transport resilience: the occupy central movement in Hong Kong from another perspective, Transportation Research Part A: Policy and Practice, 106, 100-115, https://doi.org/10.1016/j.tra.2017.09.003, 2017.

Loo, B. P. and Wang, B.: Progress of e-development in China since 1998, Telecommunications Policy, 41, 731-742, 2017.

Rosser, J. F., Leibovici, D., and Jackson, M.: Rapid flood inundation mapping using social media, remote sensing and topographic data, Natural Hazards, 87, 103-120, doi:10.1007/s11069-017-2755-0, 2017.

Schnebele, E. and Cervone, G.: Improving remote sensing flood assessment using volunteered geographical data, Natural Hazards and Earth System Sciences, 13, 669-677, doi:10.5194/nhess-13-669-2013, 2013.

Schnebele, E., Cervone, G., and Waters, N.: Road assessment after flood events using non-authoritative data, Natural Hazards and Earth System Sciences, 14, 1007-1015, doi:10.5194/nhess-14-1007-2014, 2014.

Schumann, G., Di Baldassarre, G., and Bates, P. D.: The utility of spaceborne radar to render flood inundation maps based on multialgorithm ensembles, IEEE Transactions on Geoscience and Remote Sensing, 47, 2801-2807, doi: 10.1109/TGRS.2009.2017937, 2009.

Smith, L., Liang, Q., James, P., and Lin, W.: Assessing the utility of social media as a data source for flood risk management using a real-time modelling framework, Journal of Flood Risk Management, 10, 370-380, https://doi.org/10.1111/jfr3.12154, 2017.

Sun, D., Li, S., Zheng, W., Croitoru, A., Stefanidis, A., and Goldberg, M.: Mapping floods due to Hurricane Sandy using NPP VIIRS and ATMS data and geotagged Flickr imagery, International Journal of Digital Earth, 9, 427-441, https://doi.org/10.1080/17538947.2015.1040474, 2016.

Thumerer, T., Jones, A., and Brown, D.: A GIS based coastal management system for climate change associated flood risk assessment on the east coast of England, International Journal of Geographical Information Science, 14, 265-281, https://doi-org.eproxy.lib.hku.hk/10.1080/136588100240840, 2000.

UNISDR, U. N. I. S. f. D. R.: 2009 Global Assessment Report on Disaster Risk Reduction: Risk and Poverty in a Changing Climate Secretariat, 2009.

Wang, B., Loo, B. P., Zhen, F., and Xi, G.: Urban resilience from the lens of social media data: Responses to urban flooding in Nanjing, China, Cities, 106, 102884, https://doi.org/10.1016/j.cities.2020.102884, 2020.

Wang, B., Zhen, F., Wei, Z., Guo, S., and Chen, T.: A theoretical framework and methodology for urban activity spatial structure in e-society: Empirical evidence for Nanjing City, China, Chinese Geographical Science, 25, 672-683, doi:10.1007/s11769-015-0751-4, 2015.

Wang, Z., Ye, X., and Tsou, M.-H.: Spatial, temporal, and content analysis of Twitter for wildfire hazards, Natural Hazards, 83, 523-540, doi:10.1007/s11069-016-2329-6, 2016.

Yin, J., Karimi, S., Lampert, A., Cameron, M., Robinson, B., and Power, R.: Using social media to enhance emergency situation awareness, Twenty-fourth international joint conference on artificial intelligence, doi: 10.1109/MIS.2012.6,

Yu, D., Yin, J., and Liu, M.: Validating city-scale surface water flood modelling using crowd-sourced data, Environmental Research Letters, 11, 124011, doi:10.1088/1748-9326/11/12/124011, 2016.

Zhong, M., Shi, C., Fu, T., He, L., and Shi, J.: Study in performance analysis of China Urban Emergency Response System based on Petri net, Safety science, 48, 755-762, https://doi.org/10.1016/j.ssci.2010.02.017, 2010.

Zou, L., Lam, N. S., Cai, H., and Qiang, Y.: Mining Twitter data for improved understanding of disaster resilience, Annals of the American Association of Geographers, 108, 1422-1441, https://doi-org.eproxy.lib.hku.hk/10.1080/24694452.2017.1421897, 2018.