

1 **Urban anomalies in response to rainstorms based on smartphone location data: a
2 case study of eight cities in China**

3 **Jiawei Yi^{1,2}, Yunyan Du^{1,2*}, Fuyuan Liang³, Tao Pei^{1,2}, Ting Ma^{1,2}, Chenghu Zhou^{1,2}**

4 ¹State Key Laboratory of Resources and Environmental Information System, Institute of
5 Geographic Science and Natural Resources Research, Chinese Academy of Sciences, Beijing, China

6 ²University of Chinese Academy of Sciences, Beijing, China

7 ³Department of Geography, Western Illinois University, Macomb, IL, USA

8 *Corresponding author: duyy@lreis.ac.cn

9

10 **Abstract**

11 This study explored city residents' collective geo-tagged behaviors in response
12 to rainstorms using the number of location request (NLR) data generated by
13 smartphone users. We examined the rainstorms, flooding, NLR anomalies, as well as
14 the associations among them in eight selected cities across the mainland China. The
15 time series NLR clearly reflects cities' general diurnal rhythm and the total NLR is
16 moderately correlated with the total city population. Anomalies of NLR were
17 identified at both the city and grid scale using the S-H-ESD method. Analysis results
18 manifested that the NLR anomalies at the city and grid levels are well associated with
19 rainstorms, indicating city residents request more location-based services (e.g. map
20 navigation, car hailing, food delivery, etc.) when there is a rainstorm. However,
21 sensitivity of the city residents' collective geo-tagged behaviors in response to
22 rainstorms varies in different cities as shown by different peak rainfall intensity
23 thresholds. Significant high peak rainfall intensity tends to trigger city flooding, which
24 lead to increased location-based requests as shown by positive anomalies on the
25 time series NLR.

26

27 **Keywords:** urban anomaly; rainstorm disaster; human response; rainfall intensity
28 threshold; anomaly score

29

30 **1 Introduction**

31 Global climate change is making rainfall events heavier and more frequent in
32 many areas. Powerful rainstorms may flood a city once the rainfall exceeds the
33 discharge capacity of a city's drainage system. Inundation of cities' critical
34 infrastructures and populated communities tends to disrupt urban residents' social
35 and economic activities and even cause dramatic life and property losses
36 (Papagiannaki et al. 2013; Spitaliar et al. 2014; Liao et al. 2019). Floods nowadays are
37 the most common type of natural disaster, which poses a serious threat to the safety
38 of life and property in most countries (Alexander et al. 2006; Min et al. 2011; Hu et al.
39 2018). According to the released survey in the Bulletin of Flood and Drought

1 Disasters in China, more than 104 cities were struck by floods in 2017, affecting up to
2 2.18 million population and causing over 2.46 billion US dollars direct economic
3 losses (China National Climate Center 2017).

4 The impacts of a rainstorm are usually evaluated with respect to the interactions
5 among rainfall intensity, the population exposure, the urban vulnerability, and the
6 society coping capacity (Spitalar et al. 2014; Papagiannaki et al. 2017). The rainfall
7 intensity that may trigger flood disasters has been extensively investigated and
8 many studies have examined the relationship between rainfall intensities and social
9 responses (Ruin et al. 2014; Papagiannaki et al. 2015; Papagiannaki et al. 2017).
10 Nowadays the peak rainfall intensity is widely used to determine the critical rainfall
11 threshold for issuing flash flood warnings (Cannon et al. 2007; Diakakis 2012; Miao
12 et al. 2016).

13 The population exposure refers to the spatial domain of population and
14 properties that would be affected by a rainfall hazard (Ruin et al. 2008). Gradual
15 increase in the proportion of population living in urban areas due to urbanization
16 makes more people exposed and vulnerable to urban flash floods, posing great
17 challenge to flood risk reduction (Liao et al. 2019). Reduction of vulnerability
18 therefore becomes critical in urban disaster mitigation. Vulnerability is usually
19 assessed by comprehensively considering related physical, social, and
20 environmental factors (Kubal et al. 2009; Adelekan 2011; Zhou et al. 2019), and
21 their dynamic characteristics across space and time (Terti et al. 2015).

22 Coping capacity reflects the ability of a society to handle adverse disaster
23 conditions and it is one of the most important things to consider in disaster
24 mitigation (UNISDR 2015). The coping capacity is usually evaluated by examining the
25 human behaviors in response to disasters, which are mainly collected by
26 post-disaster field investigation and questionnaires (Taylor et al. 2015). Such
27 conventional approaches only provide limited samples that may not be able to fully
28 and timely reflect disaster-induced human behaviors. Recently, researchers have
29 learned the advantages of using unconventional data sets such as insurance claims
30 (Barberia et al. 2014), newspapers (Llasat et al. 2009), and emergency operations
31 and calls (Papagiannaki et al. 2015; Papagiannaki et al. 2017) to quantify the coping
32 capacity.

33 The growing use of smartphones and location-based services (LBS) in recent
34 years has generated massive geospatial data, which could be used to infer the
35 collective geo-tagged human activities. The geospatial data thus provides a new
36 perspective to study normal urban rhythm in regular days (Ratti et al. 2006; Ma et al.
37 2019) and abnormal human behaviors in response to emergencies (Goodchild &
38 Glennon 2010; Wang & Taylor 2014; Kryvasheyeu et al. 2016). Bagrow et al. (2011)

1 found the number of phone calls spiked during earthquake, blackout, and storm
2 emergencies. Dobra et al. (2015) explored the spatiotemporal variations in the
3 anomaly patterns caused by different emergencies. Gundogdu et al. (2016) reported
4 that it is possible to identify the anomalies inflicted by emergencies or
5 non-emergency events from mobile phone data using a stochastic method. In
6 addition to the afore-mentioned applications, more studies are needed to explore
7 the full potential of the mobile phone data in terms of revealing human collective
8 behaviors, particularly in response to hazards and emergencies.

9 This study explored the urban anomalies and their variations in response to
10 rainstorms using the NLR requests from smartphone users. We selected eight
11 representative cities in the mainland China to examine how urban residents response
12 to typical summer rainstorms in different regions. The anomalies of LBS requests
13 caused by rainstorms were identified using a time series decomposition method and
14 then described by multiple indices, which are used to study how rainstorms affect
15 geo-tagged human behaviors collectively. The rest of the paper is organized as
16 follows. Section 2 introduces the selected cities and the smartphone NLR dataset.
17 Section 3 presents the anomaly detection and description methods. Section 4
18 provides the analysis results including rainfall statistics, normal rhythms, and
19 rainstorm-triggered anomalies in the selected cities. Section 5 concludes the study
20 and discusses the future work.

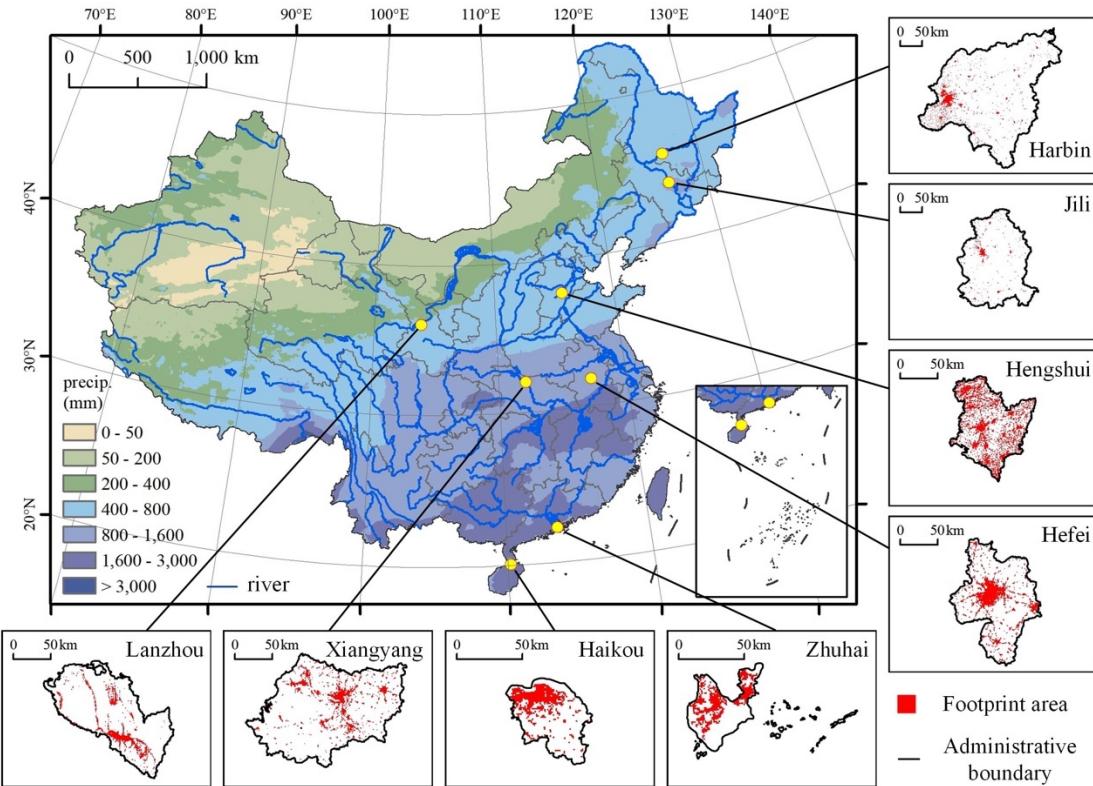
22 **2 Materials**

23 **2.1 Study area**

24 We selected eight representative cities across the mainland China for this study
25 (Fig. 1). Two cities were selected from each region except the northwestern and
26 southwestern China (Table 1). The eight cities vary significantly with respect to their
27 total population, footprint areas, and urbanization rate. In this study, the footprint of
28 a city is composed of the grids that have an hourly number of location requests (NLR)
29 no less than the median of the daily NLR time series of that grid over the whole
30 month, i.e., the grids with at least one NLR every hour in average.

31 Haikou and Zhuhai are located in southern China which has mean annual
32 precipitation between 1600 mm and 3000 mm. Among the eight cities, Zhuhai is the
33 least populated city but with the highest urbanization rate. In central China, we
34 selected Hefei and Xiangyang, which have mean annual precipitation between 800
35 mm and 1600 mm. Two cities, Lanzhou and Hengshui, were selected from a
36 semi-humid region in northern China with mean annual precipitation between 400
37 mm and 800 mm. Hengshui has the largest footprint area but the least urbanization

1 rate among the cities. Harbin and Jilin are located in the Northeastern China. The
 2 mean annual precipitation of Harbin and Jilin ranges from 400 mm to 800 mm and
 3 between 800 mm and 1600 mm, respectively. Harbin is the most populated among
 4 the eight cities.



5
 6 Figure 1 A map showing the geographic locations, annual precipitation, and
 7 footprints of the eight cities in this study.

8
 9 Table 1 Statistics of the cities

Region	City	Population (10 ⁴)	Footprint area (km ²)	Urbanization rate (%)
Southern China	Haikou	227.21	625	78.21
	Zhuhai	176.54	567	89.37
Central China	Hefei	796.50	1927	73.75
	Xiangyang	565.40	1817	59.65
Northern China	Lanzhou	372.96	1219	81.02
	Hengshui	446.04	2997	50.60
Northeastern China	Harbin	1092.90	2083	64.50
	Jilin	415.35	704	52.80

10

1 **2.2 Data collection**

2 The smartphone location data was obtained from the Tencent big data portal
3 (<https://heat.qq.com/>). The portal provides location request records of the global
4 smartphone users via the Tencent Map API. A location request record is generated
5 when a smartphone user requests any LBS, which include but are not limited to
6 navigation, car hailing, food and merchandise delivery, or social media check-ins.
7 Table 2 lists part of the LBS applications that collect user's location requests. The
8 portal releases the number of location requests per 0.01×0.01 regular grid for every
9 4-5 minutes. Comparing with other Chinese social media platforms, Tencent is the
10 most popular one with the largest social community, which reported to have nearly
11 1.1 billion monthly active users for 2018
12 (<https://www.tencent.com/en-us/company.html>). Ma (2019) compared the NLR
13 dataset with visitor numbers in a few places and confirmed that the NLR data is a
14 good proxy for investigating dynamic population changes. We collected the NLR data
15 of the grids within the administrative boundaries of the eight cities from August 1 to
16 31, 2017.

17 This study used the Version 05B GPM/IMERG 30-minute precipitation dataset
18 (Huffman et al. 2018), which has a spatial resolution of 0.1×0.1 degrees. This dataset
19 has been evaluated and widely used (Wang et al. 2017; Zhao et al. 2018; Su et al.
20 2018). The news reports about the flooding events in the eight cities were mainly
21 collected from the Chinese mainstream online media, including Xinhuanet, Ecns.cn,
22 Sohu, etc.

23 Table 2. Common smartphone applications using location-based services

Application	Type	LBS use case
WeChat	Mobile messaging app	Share location with friends
Mobile QQ	Mobile messaging app	Share location with friends
QQ Browser	Mobile web browser app	Push notifications of local news and weather
Qzone	Social network platform	Post geo-tagged microblogs

QQ Music	Mobile music app	Listen to music while running
Tencent Video	Mobile video app	Upload geo-tagged videos
Tencent News	Mobile news app	Push notifications of local news
JingDong (JOYBUY)	Mobile e-commerce platform	Location-based product recommendation
DiDi	Mobile transportation platform	Location-based car hailing
Meituan	Mobile on-demand delivery app	Location-based restaurant recommendation
Waimai		
Dianping	Mobile review and rating app	Location-based recommendation of restaurants, hotels, shops, etc.
Wangzhe Rongyao	Mobile game	Interact with nearby players

1

2

3 **3 Methods**

4 **3.1 Time series anomaly detection**

5 The smartphone location request record can be represented by a series of
 6 spatial points: $\{(x_i, y_i, T_{s_i})\}$, $i=1, 2, \dots, n$. Each point contains its geographic coordinates (x ,
 7 y) and a time (T) when the LBS is requested. The NLR was then aggregated to time
 8 series per grid or per city as illustrated below.

9 At the city level, a time series hourly NLR was established by adding up all
 10 location requests of the grids within the footprint area of that city. The magnitudes of
 11 the NLR in different cities vary significantly due to the different numbers of smart
 12 phone users. To make the NLR in different cities comparable, we normalized the NLR
 13 using the median-interquartile normalization method, which is more robust to

1 anomalies than other common approaches using sample mean and standard
2 deviation (Geller et al. 2003).

3 We employed the S-H-ESD method (Vallis et al. 2014) to detect anomalies from
4 the time series NLR, which can be represented by the following additive model

5 $Ts = T + S + R$ (1)

6 where T , S , and R denote the trend, seasonality and residual components in the time
7 series data, respectively. The S-H-ESD method assumes that the trend and the
8 seasonality would not be significantly disrupted by rapid-evolving events that last for
9 only a few hours. Two major steps are involved in the method. First, it uses the
10 piecewise median method to fit and remove the long-term trend and then the STL to
11 remove seasonality (Cleveland et al. 1990). Using the STL to remove the long-term
12 trend would introduce artificial anomalies (Vallis et al. 2014). In this study, the
13 underlying trend in the time series NLR is approached using a piecewise combination
14 of the biweekly medians, which show little changes over the whole time series.

15 In the second step, the S-H-ESD method employs the generalized Extreme
16 Studentized Deviate (GESD) statistic (Rosner 1975) to identify the significant
17 anomalies in the residuals. The GESD calculates the statistic (G) based on the mean (\bar{r})
18 and the standard deviation (s) of the observations:

19
$$G = \frac{\max |r_j - \bar{r}|}{s} \quad (2)$$

20 Given the upper bound of u suspected anomalies, the GESD performs u separate
21 tests. In each test, the GESD re-computes the statistic G after removing the

22 observation r_j that maximizes $|r_j - \bar{r}|$ and then compares G with the critical value λ
23 as defined below:

24
$$\lambda = \frac{(k-1)t_{1-\alpha/(2k), k-2}}{\sqrt{k(k-2+t_{1-\alpha/(2k), k-2}^2)}} \quad (3)$$

25 where k denotes the number of the observations in the time series after eliminating
26 a suspected anomaly in the last run, and $t_{p,d}$ represents the p^{th} percentile of a t
27 distribution with a degree of freedom d . In this study, we set the significance level α
28 as 0.05 and the number of anomalies no more than 25% of the total observations.
29 Each test identifies one anomaly in the residuals when $G > \lambda$. The identified anomaly
30 is either a positive or negative, depending upon whether the residual is greater or
31 smaller than 0, respectively.

1

2 **3.2 Anomaly measures and scores**

3 In this study, an individual anomaly is represented with a vector,

4
$$v=(x, y, t, obs, res) \quad (4)$$

5 where x and y denote the coordinates of the grid centroid, t denotes the observation
6 time, obs and res denote the observation and the residual (R in equation 1) in the
7 time series. This study uses an anomaly's absolute residual to describe its unusual
8 deviation from its expectation.9 A rainstorm disaster, once significantly impacts the cities, usually can trigger an
10 outbreak of NLR anomalies in multiple places across the city. To collectively
11 characterize the abnormal human responses, this study defines three indices: the
12 total number (N_t), the total residual (R_t), and the mean density (D_t) of the positive or
13 negative anomalies. The mean density is defined as follows:

14
$$D_t = \frac{\sum_{i=1}^{N_t} B_i}{N_t} \quad (5)$$

15 where B_i denotes the number of neighborhood anomalies within a Manhattan
16 distance of a 5-grid (~ 5 km) radius of the i^{th} anomaly. The radius is large enough to
17 cover most urban facilities nearby the anomaly.18 An anomaly score is then defined based on the afore-mentioned indices to
19 evaluate the city residents' responses to a rainstorm event. First, we surveyed the
20 hourly changes of the indices and calculated the quartiles (Q_1, Q_2, Q_3) and
21 interquartile range (IQR) of each index for every hour every day. The score of an
22 index is defined by:

23
$$S_{V,t} = \begin{cases} \frac{V_t - Q_1}{IQR} & , \text{if } V_t \leq Q_1 \\ 0 & , \text{if } Q_1 < V_t < Q_3 \\ \frac{V_t - Q_3}{IQR} & , \text{if } V_t \geq Q_3 \end{cases} \quad (6)$$

24 where V_t represents one of the three indices at time t . According to Tukey's fences
25 (Tukey 1977), the score is considered an outlier if its absolute value is greater than
26 1.5 or an extreme if it is greater than 3. The final anomaly score is the mean of the
27 three index scores.

28

29 **3.3 Characterization of a rainfall event**30 In this study, we examined the city residents' responses to the rainfall events in
31 August 2017. The national average precipitation of this month is 124.6 mm, which is
32 the highest in 2017 and 21.3% more than the August average precipitation in
33 previous years.

34 We defined a rainfall event as a precipitation process that lasts for at least two

1 hours and with no rain preceding it for at least one hour. The severity of a rainfall
2 event is described by its duration, accumulated precipitation, and peak rainfall
3 intensity. The duration refers to how long a rainfall event lasts, and the accumulated
4 precipitation is the total precipitation received during a rainfall event. The peak
5 rainfall intensity (I_d) is widely used to estimate the possible rainfall intensity
6 threshold that triggers city (Cannon et al. 2007; Diakakis 2012) and is defined as
7 below:

8

$$I_d = \frac{\max\{\sum_{i=j}^{j+d-1} P_i\}}{d}, \quad j = 1, 2, \dots, N - d + 1 \quad (7)$$

9 where P_i denotes the precipitation during the i^{th} time interval, N denotes the total
10 number of the time intervals in a rainfall time series, and d denotes the width of the
11 moving time window that was used to search for the maximum accumulated
12 precipitation in a rainfall event. Based on the peak rainfall intensity, the August
13 rainfall events in the eight cities can be categorized into moderate rainstorm (0.5
14 mm/h $< I_1 \leq 4$ mm/h), heavy rainstorm (4 mm/h $< I_1 \leq 8$ mm/h), and violent
15 rainstorm ($I_1 \geq 8$ mm/h).

16 For calculation purpose, we downscaled the precipitation data to the same
17 spatial resolution as that of the NLR using the nearest-neighbor interpolation method.
18 At the city level, the rainfall of a city is defined as the total of the half-hour TRMM
19 precipitation within the human footprint. At the grid level, the rainfall of each grid
20 refers to the total precipitation received by that grid within a certain time period.
21

22 **4 Results**

23 **4.1 Rainfall characteristics and peak rainfall intensity thresholds**

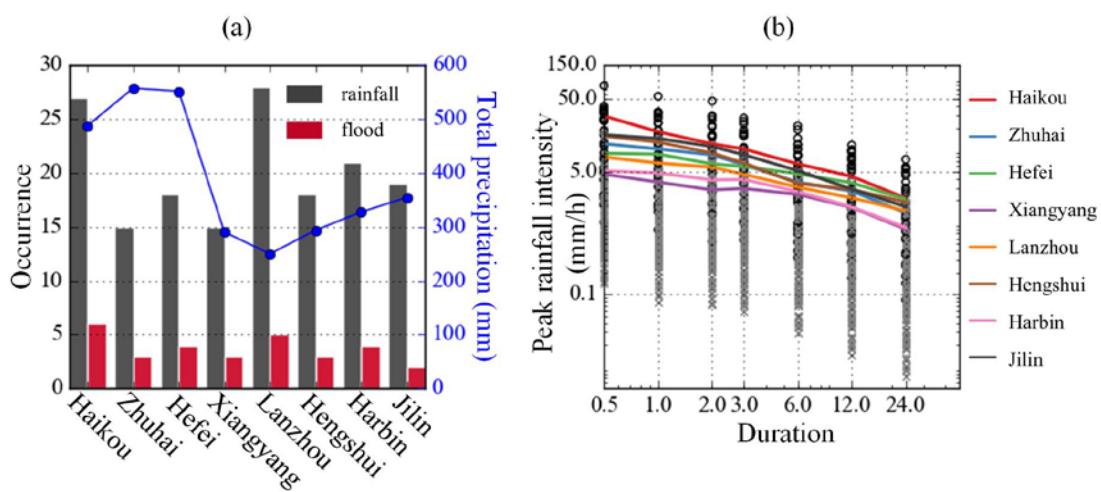
24 The eight cities could be categorized into two groups in terms of the total
25 precipitation amount in August 2017 (Fig. 2a). The first group includes Haikou, Zhuhai,
26 and Hefei, with total precipitation more than 400 mm. The summer monsoon brings
27 plenty of water to the two coastal cities (i.e. Haikou and Zhuhai). The Typhoon Hato,
28 when it made landfall on August 23, further dumped 68- and 108-mm water to
29 Haikou and Zhuhai, respectively. By contrast, the inland city Hefei, received 47.6%
30 more precipitation in 2017 than the average mainly due to a few unusual rainstorms
31 in August 2017 (Hydrology and Water Resource Bureau of Hefei 2018)

32 The second group includes all the other cities, which have less than 400 mm
33 precipitation in August 2017. The city Lanzhou is located in the dry northwestern
34 China and has the least precipitation of 250 mm. The two inland cities, Xiangyang and
35 Hengshui both have slightly higher precipitation of 300 mm. The precipitation of the
36 two northeastern cities, Harbin and Jilin, ranges between 320 and 350 mm and is

1 mainly brought in by the northwestern vortexes.

2 There are at least 15 rainstorms and two flooding events in each city. The city
3 Haikou, Lanzhou, and Harbin witnessed more than 20 rainstorms and about 1/4 out
4 of them caused serious flooding problems. The number of rainstorms in the other
5 cities ranges from 15 to 20 and about two to four out of them caused flooding
6 problems in the cities.

7 We identified the peak rainfall intensity threshold value that likely triggers city
8 flooding using the method developed by Cannon et al. (2008) and Diakakis (2012).
9 The method plots peak rainfall intensity of different time windows against the
10 corresponding rainfall duration. The flood-triggering threshold is defined as the
11 upper limit of the peak rainfall intensity that tends to lead to urban flooding but
12 actually not. As shown in Fig. 2b, for the rainfall thresholds calculated based on 0.5-,
13 1-, 2-, and 3-hour time window, the city ranking shows no change with an order of
14 Haikou, Jilin, Hengshui, Zhuhai, Hefei, Lanzhou, Harbin, and Xiangyang. The ranking
15 shows some fluctuations when the flooding-triggering rainfall threshold values were
16 calculated with a more than 3-hour time window. However, Haikou and Harbin are
17 always the top two cities whereas Xiangyang is the last one on the ranking list. It is
18 worthy to note that a rainstorm with a peak rainfall intensity over the threshold 5
19 mm/h would definitely trigger floods in Xiangyang. However, in Haikou, such a
20 threshold value is 30 mm/h. In other words, city flooding would occur in Haikou
21 when it is hit by a rainstorm with peak rainfall intensity over 30mm/h. In general, the
22 difference between the threshold values among these cities reduces with a longer
23 time window, indicating that the rainfall in a shorter time window is more critical to
24 evaluate whether a city is prone to flooding.



25
26 Figure 2. Total August precipitation and frequency of rainfall and city flooding events
27 (a). Variations in peak rainfall intensity (circles) and the flooding-triggering
28 precipitation threshold values (lines) that are derived from time windows ranging
29 from 0.5 to 24 hours (b).

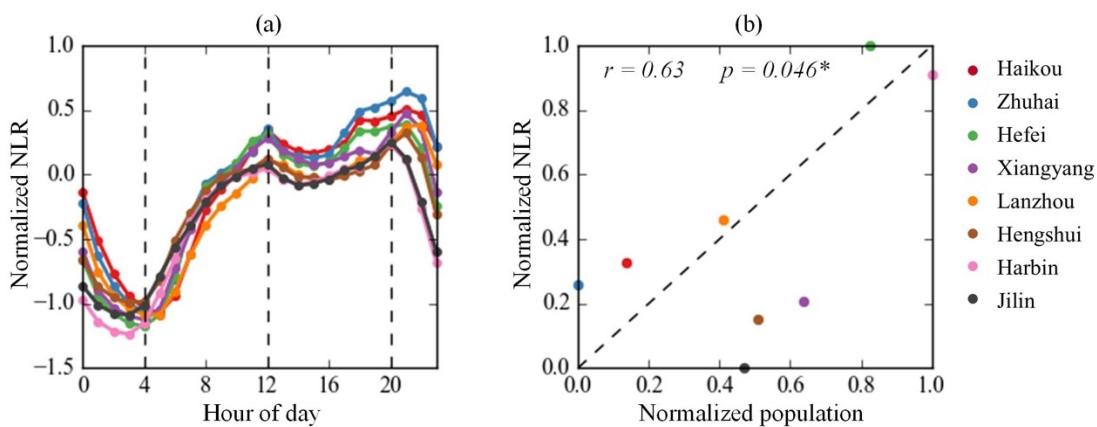
1

2 **4.2 Normal rhythm of city**

3 The NLR records can serve as a proxy of the city residents' normal daily routine.
 4 The normalized NLR show the eight cities have a similar diurnal rhythm (Fig. 3a). The
 5 normalized NLR median climbs from a minimum at around 4:00 and to a peak right at
 6 12:00. It starts to drop slightly and then peaks again at around 20:00. This general
 7 pattern reflects the smartphone usage patterns of the city residents. Phone usage
 8 starts to drop after the midnight when most residents start to rest. It reaches its first
 9 peak during the lunch time as residents may request more LBS to find a place to eat.
 10 After lunch time, phone usage remains at a high plateau, probably due to more LBS
 11 requests for business purposes. Phone usage reaches the highest peak of the whole
 12 day right after the normal work hours, indicating a significant increased need for the
 13 LBS such as hailing nearby taxis to socialize with friends, go back home, or sending
 14 geo-tagged posts for socializing.

15 The general diurnal pattern was superposed with subtle short-term NLR
 16 variations. The NLR in the southern cities peaks and hits the bottom later at night and
 17 before dawn, respectively, than that of the northern cities. This is very likely due to
 18 the different lifestyles between the northern and southern residents in response to
 19 the economic activities and day length. It is well-known that the southern China is
 20 more active in economic and social activities and the southerners enjoy the night
 21 activities more (Ma et al. 2019). By contrast, the northerners tend to end their
 22 nightlife earlier and also become active earlier as the day breaks earlier in the north.

23 The total NLR is moderately correlated with the population of these cities (Fig.
 24 3b). The 0.63 Pearson correlation coefficient (with a p value of 0.046) indicates a
 25 statistically significant positive relationship between the normalized NLR and the
 26 population. As a result, we believe the NLR data could reflect the collective
 27 geo-tagged behaviors of the city residents as a whole and consequently it could serve
 28 as a proxy of the human responses to different environmental and social events.



29

30 Figure 3. The diurnal variation patterns of the NLR in the eight cities (a) and a positive

1 correlation between the NLR and the total number of residents (b).

2

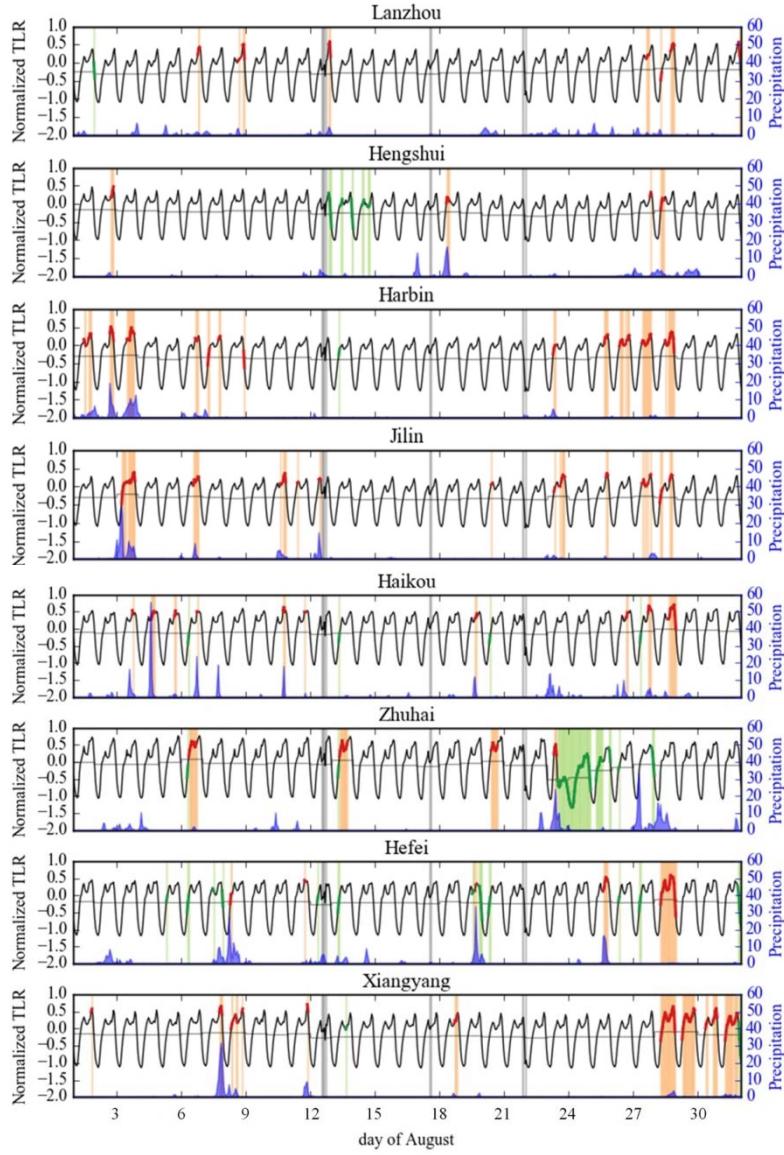
3 **4.3 Urban anomalies during rainstorms**

4 **4.3.1 City-scale analysis**

5 There are more positive than negative anomalies in the August time series
6 hourly NLR and most positive anomalies were found in pair with precipitation spikes
7 (Fig. 4). For example, two significant precipitation spikes in Harbin in the afternoon of
8 August 2nd and 3rd were closely associated with positive NLR anomalies. Few NLR
9 negative anomalies were identified in the eight cities except Zhuhai. This city was
10 significantly affected by Typhoon Hato, which brings huge amount of precipitation
11 and leads to a negative anomaly since the Afternoon of August 23rd in Zhuhai. Such a
12 significant negative anomaly could be attributed to serious communication
13 interruption or damages caused by the typhoon.

14 It is worthy to note that both positive and negative anomalies were also
15 identified when there is no rain in the cities. For example, two positive anomalies
16 were identified around August 28th in Harbin when there is no rain at all. The no-rain
17 anomalies must be triggered by other major events in the cities. However, at this
18 moment it is not easy to trace what local events may trigger such anomalies.

19 It is very interesting to notice that a couple of no-rain positive anomalies were
20 identified in the last week of August for almost all eight selected cities except Zhuhai.
21 These positive anomalies were obviously not associated with any special rainstorm
22 events. Instead, they are more likely to be associated with sort of national-wide
23 events, such as the college students' back to school and move-in events, which are
24 mainly scheduled in the last week of August every year in China. Such positive
25 anomalies were not found in Zhuhai, of which the 2017 back to school and move-in
26 events was postponed to the first week of October due to the significant damages
27 caused by Typhoon Hato. However, further studies, such as of the NLR of other cities
28 in China, are needed to consolidate this argument.



1

2 Figure 4. The time series NLR and rain events during August 2017. Positive and
 3 negative anomalies were shown in orange and green colors, respectively. The light
 4 gray columns show the periods when NLR data is missing.

5

6 We further quantitatively examined the association between rainfall events and
 7 the NLR anomalies. Table 3 lists the R_{pos} and R_{neg} , which are the ratios of the positive
 8 and negative anomalies corresponding to the four scenarios (no rains, moderate,
 9 heavy and violent rainstorm events) to the total number of anomalies identified over
 10 the whole time series, respectively. As shown in Table 3, in total we identified 27, 19,
 11 78, and 166 violent, heavy, moderate, and no rainstorm events in the eight cities,
 12 respectively. Under different scenarios, the R_{pos} is always higher than R_{neg} except the
 13 no rain scenario, in which there is no significant difference between these two ratios.
 14 The rainstorm-related R_{pos} increases from 0.22 to 0.70 as the rainstorms level up
 15 from moderate to violent as compared to a no-rain R_{pos} of 0.12. The rain-related or

1 no-rain R_{neg} is no more than 0.22. The R_{pos} is much higher than R_{neg} when the cities
 2 are affected by stronger rainfall events. For example, when the cities are affected by
 3 violent storms, the R_{pos} and R_{neg} are 0.70 and 0.22 respectively. By contrast, the R_{pos}
 4 and R_{neg} are 0.22 and 0.06, respectively when the cities are affected by moderate
 5 rainstorms. It is very likely that, when there are severe rainstorms, people may send
 6 out more LBS requests in order to, for instance, search a route free of inundation
 7 spots and less congested roads, order delivery food, or post geo-tagged photos of the
 8 terrible moments.

9 A lower R_{pos} of the heavy and moderate rainstorms may also be partly
 10 attributed to the effect of data aggregation at the city scale. It is very common that a
 11 rainstorm may influence only a part of the city and only lead to certain local positive
 12 anomalies. In such a case, increase of the NLR in a small number of grids may not
 13 result in significant changes of the NLR of the entire city and consequently no
 14 anomalies at the city level. Analysis at the grid level, as reported in the next section,
 15 would show how residents respond to the local rainstorm events.

16 The difference between the R_{pos} and R_{neg} also varies for different cities. For
 17 example, the two violent rainstorms both trigger a positive anomaly in Xiangyang
 18 and Harbin. By contrast, the five violent rainstorms in Zhuhai lead to the same
 19 percent positive and negative anomalies. City Hefei is special. The same percent of
 20 positive and negative anomalies are triggered by the five violent storms. However,
 21 when Hefei is affected by the moderate and heavy rainstorms or even no rainfalls,
 22 there are slightly more negative than positive anomalies.

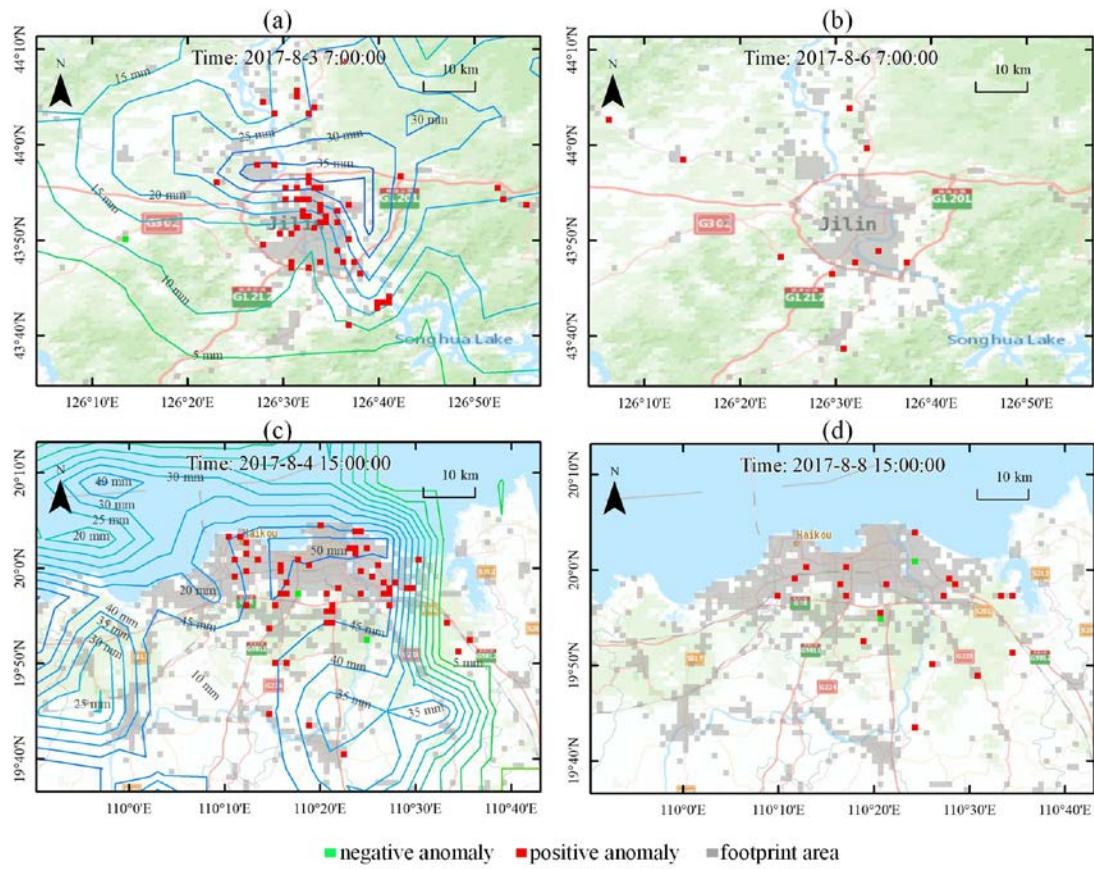
23
 24 Table 3. Numbers of different categories of rainstorms and the corresponding R_{pos}
 25 and R_{neg} .
 26

Cities	No rainfall			Rainstorms					
	N	Moderate		N	Heavy		N	Violent	
		R_{pos}	R_{neg}		R_{pos}	R_{neg}		R_{pos}	R_{neg}
Haikou	27	0.04	0.22	14	0.21	0.00	3	0.33	0.00
Zhuhai	16	0.19	0.25	5	0.20	0.20	3	0.00	0.00
Hefei	19	0.05	0.32	7	0.00	0.14	2	0.50	1.00
Xiangyang	15	0.33	0.33	7	0.29	0.00	0	-	-
Lanzhou	29	0.07	0.10	17	0.24	0.06	5	0.20	0.20
Hengshui	19	0.00	0.21	11	0.18	0.09	2	0.00	0.00
Harbin	21	0.24	0.10	7	0.14	0.14	3	1.00	0.00
Jilin	20	0.15	0.15	10	0.40	0.00	1	1.00	0.00
Overall	166	0.12	0.20	78	0.22	0.06	19	0.37	0.11
							27	0.70	0.22

27
 28 **4.3.2 Grid-scale analysis: anomaly indices**

29 The S-H-ESD method was also used to detect the NLR anomalies at the grid level.
 30 There are always more grids showing anomaly when the city was affected by a

1 rainstorm. Figure 5 provides an example to illustrate the grids with anomaly detected
 2 during a rainstorm and the same time period in another day without rainfall in Jilin
 3 and Haikou, respectively. Anomalies were identified in 56 grids in Jilin when it was hit
 4 by a rainstorm at 7am on August 3, 2017. By contrast, anomalies are observed in only
 5 10 grids during the same time period on August 6, 2017 when there is no rain at all.
 6 In Haikou, anomalies are found in 52 grids during a rainstorm and only 19 grids when
 7 there is no rain.



8
 9 Figure 5. Grid with negative and positive anomalies within the footprint areas of
 10 Haikou and Jilin. The contour lines show the precipitation.

11
 12 The total number, total residual, and mean density of these anomalies were
 13 then calculated (Fig. 6) for the cities when they were affected by flooding caused by a
 14 typical rainstorm event (Table 4). The three anomaly indices show similar diurnal
 15 variations as of the NLR diurnal rhythm and they all spiked to the level of an outlier
 16 or even to an extreme value when the city was significantly affected by flooding

1 issues.

2 After the spikes, the anomaly indices usually bounce back to the same level
3 before for almost all the cities except Zhuhai, indicating most cities return to their
4 normal rhythms after the rainstorm interruption. However, Zhuhai was hit by the
5 category-3 Typhoon Hato at around 12:50 on August 23. The typhoon brought
6 intense rain, strong winds, and caused significant flooding issues and damages to the
7 city infrastructures, causing a sharp decline and persistent negative anomalies after
8 the landfall of Hato. It took more than 72 hours for the anomaly indices to bounce
9 back to the same level before Hato (not shown in Fig. 6).

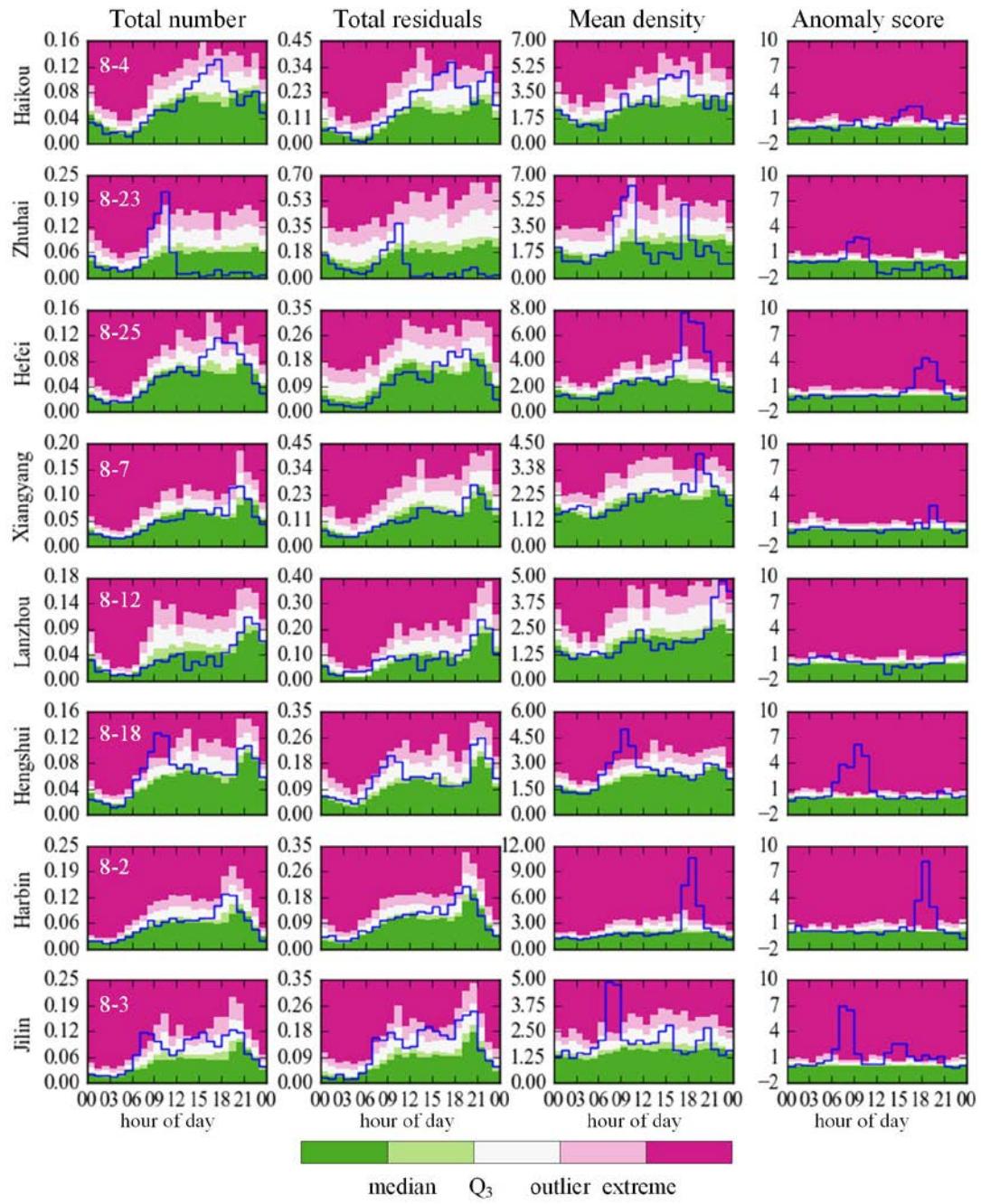
10

11 Table 4. An exemplary flooding event in each of the cities.

12

City	Urban flood event	Rainfall duration (h)	Accumulated precipitation (mm)	Half-hour peak rainfall intensity (mm/h)
Haikou	8-4 15:00	10	117	77
Zhuhai	8-23 12:50	23	108	32
Hefei	8-25 17:00	13	72	25
Xiangyang	8-7 18:00	30.5	140	34
Lanzhou	8-12 21:00	9.5	14	5
Hengshui	8-18 08:00	15	67	18
Harbin	8-2 17:00	12.5	61	26
Jilin	8-3 07:00	38.5	185	31

13



1

2 Figure 6. Intra-day variations in NLR, total residuals, mean density, and anomaly score
 3 within 24 hours of a typical flooding event in each of the cities.

4

5 4.3.3 Grid-scale analysis: anomaly score and rainfall intensity

6

7 Given the anomaly score is indicative of the unusual responses of residents to
 8 rainstorms, we further examined the relation between the anomaly score and the
 9 rainfalls in these cities during the August 2017.

1 The grid-level R_{pos} is much higher than the city-level counterpart with respect to
2 all types of events (Fig. 7a). Such a difference is mainly due to the different analysis
3 levels. We can easily identify the local anomalies per grid, which are more likely to be
4 obliterated at the city level due to the data aggregation. At the grid level, the R_{pos}
5 and R_{neg} also vary in response to the different levels of rainstorm events. All cities
6 show a higher R_{pos} when they are affected by violent rainstorms (85%) than heavy
7 rainstorms (68%), in comparison with the R_{pos} (56%) when the cities are not affected
8 by any rainfall events. However, the R_{pos} of moderate rainstorms (45%) is less than
9 the no-rain R_{pos} , likely suggesting that low-intensity rainfall events may not
10 necessarily trigger NLR anomalies and other factors may contribute to the NLR
11 anomalies at the grid level.

12 How easily the rhythm of a city would be disrupted by a rainstorm is strongly
13 related to the anomaly-triggering peak rainfall intensity threshold (Fig. 7b), which
14 was calculated using the same the ideas in the methods developed by Cannon et al.
15 (2008) and Diakakis (2012). We plotted the peak rainfall intensity with respect to
16 whether there are anomalies or not for each city. The anomaly-triggering peak
17 rainfall intensity is defined as the upper limit of the rainfall intensity that tends to
18 lead to an NLR anomaly but actually not.

19 Every rainstorm with its peak intensity higher than the threshold would
20 definitely trigger an NLR anomaly. As a result, the cities with a lower threshold tend
21 to be more easily disrupted by a moderate or heavy rainstorm. For example,
22 Xiangyang has a very low threshold value of 1.4 mm/h. In August 2017, there are six
23 rainstorm events with peak rainfall intensity exceeding this threshold and they all
24 caused anomalies in this city.

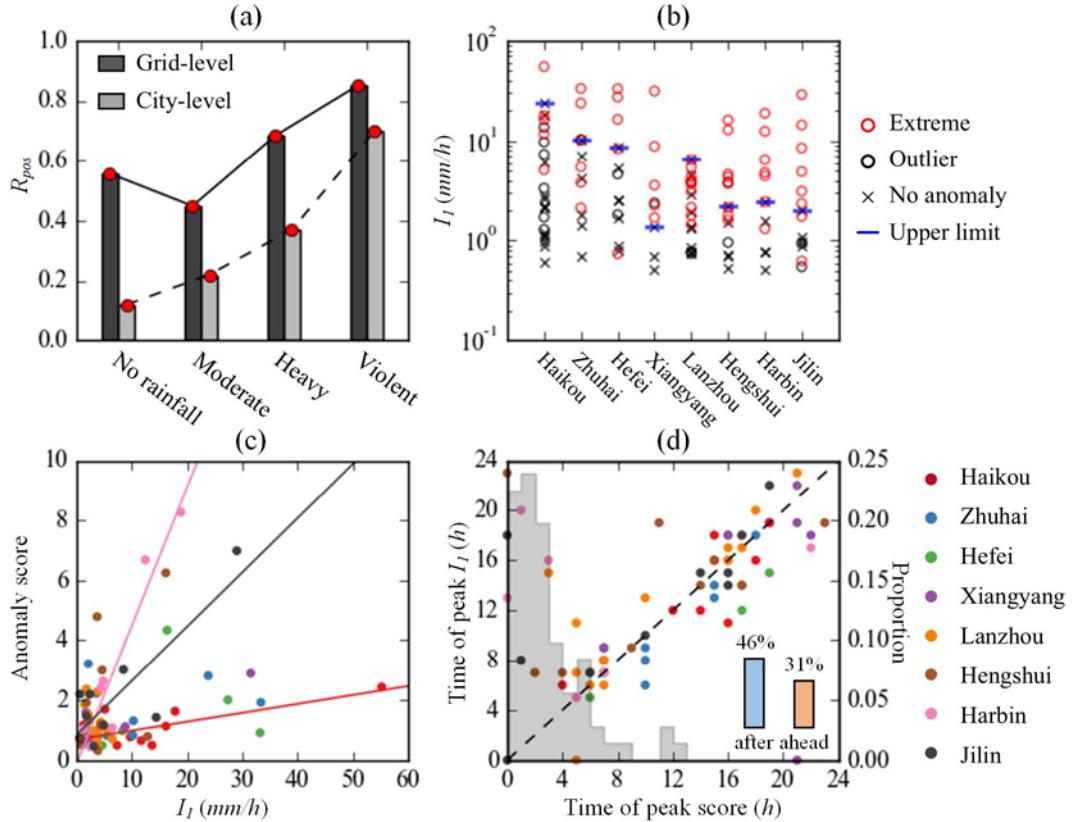
25 However, even a rainstorm with its peak rainfall intensity below the threshold
26 may also trigger an NLR anomaly. For example, quite a few NLR anomalies were
27 found in Lanzhou, of which most rainstorms have its peak rainfall intensity below the
28 threshold (6.6 mm/h). This is because a heavy rainstorm at around 24:00 failed to
29 trigger an NLR anomaly as most people were sheltered at home and hence were not
30 affected. However, this rainstorm is included in the process to calculate the peak

1 rainfall intensity and increase the threshold. As a result, rainstorms with their peak
2 rainfall intensity below the threshold may also trigger anomalies, particularly in the
3 cities with more heavy and violent rainstorms after late night and before dawn.

4 The anomaly score is correlated with rainfall intensity for some cities (Fig. 7c).
5 Specifically, there are three cities, i.e. Harbin, Jilin, and Haikou, showing a positive
6 linear relationship between the anomaly score and rainfall intensity. The p-values in
7 the linear regression analysis are less than 0.05, suggesting the correlations are
8 statistically significant. As the rainfall intensity increases, the anomaly scores of the
9 three cities increase linearly. Furthermore, the slope coefficients of the correlations
10 indicate how sensitive the rainfall intensity may trigger anomalies. The city Harbin
11 has the steepest slope thus slight increase in rainfall intensity would trigger
12 anomalies more easily. By contrast, the gentlest slope indicates Haikou is a city where
13 the residents, in terms of their LBS request, are not very sensitive to the increase of
14 the rainfall intensity. Such diverse sensitivity may be essentially due to different
15 climatic conditions, infrastructure levels or other potential factors in these cities. The
16 city Haikou is situated in humid climate zone with average rainfall exceeding 1600
17 mm per year, greater than the other two cities. Meanwhile, considering the
18 drainpipe density, which is defined as the ratio of the total length of drainpipes to the
19 built-up area of a city, we found Haikou also has higher drainpipe density (11.74 km^{-1})
20 than the rest two cities (5.73 km^{-1} for Jilin and 7.44 km^{-1} for Haikou). So, although
21 rainstorms frequently hit Haikou, their impacts on local residents could be much
22 lessened by efficient drainage system.

23 Around 31%, 23%, and 46% of the maximum anomaly scores were detected
24 before, at the same time, and after the rainfall intensity reaches its peak (Fig. 7d).
25 Specifically, 23%, 24%, and 20% of the anomaly score peaks simultaneously, within 1
26 hour, and within 2 hours of the rainfall intensity peaks, respectively. About 46% of
27 the anomaly score peaks after the rainfall intensity peaks, which is 50% more than
28 the number of the cases that anomaly score peaks ahead of the rainfall intensity
29 peak. As a result, we usually see the maximum positive anomalies (i.e. significant
30 disturbance in city rhythm) after the rainfall intensity reached a maximum value. It is

1 also possible for the anomaly to reach its peak before the peak of the rainfall
 2 intensity if, for example, the cumulative rainfall is high enough to significantly impact
 3 the city.



5 Figure 7. Correlation between peak rainfall intensity and anomaly score.

7 5. Conclusions

8 This study shows the potentials of the NLR data in reflecting city residents'
 9 collective geo-tagged behaviors. First of all, the NLR was moderately correlated with
 10 the population of the cities. Secondly, the time series NLR data well corresponds to
 11 the regular diurnal rhythm in all eight cities, which is characterized by limited
 12 activities from the midnight to early morning and very active LBS requests from noon
 13 to the evening. Thirdly, the time series NLR also reflects the different lifestyles in the
 14 northern and southern China, showing southerners enjoy late night life more
 15 whereas the northerners start their days earlier in the morning.

16 The anomalies of the NLR data are well with that the rainstorms, especially the
 17 violent ones, were very likely to trigger positive NLR anomalies at city level. At the

1 grid level, the anomalies in response to rainstorms show a significant increase in the
2 anomaly indices in terms of the total number, total residual, and mean density. The
3 time series composite score derived from these three anomaly indices clearly shows
4 how city residents respond to rainstorms in terms of their LBS requests.

5 Rainstorms of the same magnitude may not trigger NLR anomalies in the same
6 way in every city. Essentially, the peak rainfall intensity of the rainstorms seems to be
7 the key and such a threshold is significantly different among different cities. As a
8 result, high peak rainfall intensity tends to trigger flooding and subsequently
9 anomalies in the NLR data. Furthermore, the peak rainfall intensity is well associated
10 with the peak anomaly score, further indicating it is the key factor that can trigger
11 rainstorm-induced NLR anomalies.

12 It is worthy to note that other events may also contribute to NLR anomalies.
13 There are a couple of positive anomalies in the last week of August for almost all
14 cities except Zhuhai. The last week of August is the school registration time for
15 college students in China. It is reasonable to expect such a nation-wide event may
16 trigger NLR anomalies as shown in this study. However, some college cities may
17 postpone the registration time and Zhuhai is one of them due to the significant
18 damages caused by Typhoon Hato right before the registration week.

19 We are also aware of limitation of the Tencent location request dataset.
20 Although the dataset is generated by more than one billion monthly active users, it
21 still cannot fully represent the entire population of a city. The collective response
22 analysis based on the location requests consequently may underestimate rainstorm
23 impacts upon infrequent users, particular the older and children. Future studies will
24 combine multi-source geospatial datasets to address the limitation and further
25 explore human responses to various weather events.

26

27 **Acknowledgements**

28 This research was funded by the National Key Research and Development Program of
29 China (Grant Nos. 2017YFB0503605 and 2017YFC1503003), and the Strategic Priority

1 Research Program of the Chinese Academy of Sciences (Grant No. XDA19040501).
2 The IMERG data were provided by the NASA/Goddard Space Flight Center's PMM
3 and PPS through <http://pmm.nasa.gov/data-access/downloads/gpm> (accessed 14
4 April 2019), and archived at the NASA GES DISC.

5

6 **References**

7 Adelekan, I. O.: Vulnerability assessment of an urban flood in Nigeria: Abeokuta
8 flood 2007, *Nat. Hazards*, 56(1), 215–231, doi:10.1007/s11069-010-9564-z,
9 2011.

10 Alexander, L. V., Zhang, X., Peterson, T. C., Caesar, J., Gleason, B., Klein Tank, A. M. G.,
11 Haylock, M., Collins, D., Trewin, B., Rahimzadeh, F., Tagipour, A., Rupa Kumar, K.,
12 Revadekar, J., Griffiths, G., Vincent, L., Stephenson, D. B., Burn, J., Aguilar, E.,
13 Brunet, M., Taylor, M., New, M., Zhai, P., Rusticucci, M. and Vazquez-Aguirre, J.
14 L.: Global observed changes in daily climate extremes of temperature and
15 precipitation, *J. Geophys. Res. Atmos.*, 111(5), 1–22,
16 doi:10.1029/2005JD006290, 2006.

17 Bagrow, J. P., Wang, D. and Barabási, A.-L.: Collective Response of Human
18 Populations to Large- Scale Emergencies, *PLoS One*, 6(3), 1–8, doi:10.1371/
19 journal.pone.0017680, 2011.

20 Barberia, L., Amaro, J., Aran, M. and Llasat, M. C.: The role of different factors
21 related to social impact of heavy rain events : considerations about the intensity
22 thresholds in densely populated areas, *Nat. Hazards Earth Syst. Sci.*, 14,
23 1843–1852, doi:10.5194/nhess-14-1843-2014, 2014.

24 Cannon, S. H., Gartner, J. E., Wilson, R. C., Bowers, J. C. and Laber, J. L.: Storm rainfall
25 conditions for floods and debris flows from recently burned areas in
26 southwestern Colorado and southern California, *Geomorphology*, 96(3–4),
27 250–269, doi:10.1016/j.geomorph.2007.03.019, 2007.

28 China National Climate Center: Bulletin of Flood and Drought Disaster in China 2017,
29 2017.

30 Cleveland, R. B., Cleveland, W. S., MaRae, J. E. and Terpenning, I.: STL: A
31 Seasonal-Trend Decomposition Procedure Based on Loess, *J. Off. Stat.*, 6(1),
32 3–73, 1990.

33 Diakakis, M.: Rainfall thresholds for flood triggering. The case of Marathonas in
34 Greece, *Nat. Hazards*, 60(3), 789–800, doi:10.1007/s11069-011-9904-7, 2012.

35 Dobra, A., Williams, N. E. and Eagle, N.: Spatiotemporal detection of unusual human
36 population behavior using mobile phone data, *PLoS One*, 10(3), e0120449,
37 doi:10.1371/journal.pone.0120449, 2015.

38 Geller, S. C., Gregg, J. P., Hagerman, P. and Rocke, D. M.: Transformation and
39 normalization of oligonucleotide microarray data, *Bioinformatics*,
40 doi:10.1093/bioinformatics/btg245, 2003.

41 Goodchild, M. F. and Glennon, J. A.: Crowdsourcing geographic information for

1 disaster response : a research frontier, *Int. J. Digit. Earth*, 3(3), 231–241,
2 doi:10.1080/17538941003759255, 2010.

3 Gundogdu, D., Incel, O. D., Salah, A. A. and Lepri, B.: Countrywide arrhythmia:
4 emergency event detection using mobile phone data, *EPJ Data Sci.*, 5(1),
5 doi:10.1140/epjds/s13688-016-0086-0, 2016.

6 Hu, P., Zhang, Q., Shi, P., Chen, B. and Fang, J.: Flood-induced mortality across the
7 globe: Spatiotemporal pattern and influencing factors, *Sci. Total Environ.*, 643,
8 171–182, doi:10.1016/j.scitotenv.2018.06.197, 2018.

9 Hydrology and Water Resource Bureau of Hefei: *Bulletin of Flood and Drought
Disaster in Hefei*, 2018.

10 Kryvasheyeu, Y., Chen, H., Obradovich, N., Moro, E., Van Hentenryck, P., Fowler, J.
11 and Cebrian, M.: Rapid assessment of disaster damage using social media
12 activity, *Sci. Adv.*, 2(3), e1500779, doi:10.1126/sciadv.1500779, 2016.

13 Kubal, C., Haase, D., Meyer, V. and Scheuer, S.: Integrated urban flood risk
14 assessment – adapting a multicriteria approach to a city, *Nat. Hazards Earth Syst.
Sci.*, 9, 1881–1895, doi:10.5194/nhess-9-1881-2009, 2009.

15 Liao, X., Xu, W., Zhang, J., Li, Y. and Tian, Y.: Global exposure to rainstorms and the
16 contribution rates of climate change and population change, *Sci. Total Environ.*,
17 663, 644–653, doi:10.1016/j.scitotenv.2019.01.290, 2019.

18 Llasat, M. C., Llasat-Botija, M., Barnolas, M., López, L. and Altava-Ortiz, V.: An
19 analysis of the evolution of hydrometeorological extremes in newspapers: The
20 case of Catalonia, 1982–2006, *Nat. Hazards Earth Syst. Sci.*, 9(4), 1201–1212,
21 doi:10.5194/nhess-9-1201-2009, 2009.

22 Ma, T.: Quantitative responses of satellite-derived night-time light signals to urban
23 depopulation during Chinese New Year, *Remote Sens. Lett.*, 10(2), 139–148,
24 doi:10.1080/2150704X.2018.1530484, 2019.

25 Ma, T., Pei, T., Song, C., Liu, Y., Du, Y. and Liao, X.: Understanding geographical
26 patterns of a city's diurnal rhythm from aggregate data of location-aware
27 services, *Trans. GIS*, doi:10.1111/tgis.12508, 2019.

28 Miao, Q., Yang, D., Yang, H. and Li, Z.: Establishing a rainfall threshold for flash flood
29 warnings in China's mountainous areas based on a distributed hydrological
30 model, *J. Hydrol.*, 541, 371–386, doi:10.1016/j.jhydrol.2016.04.054, 2016.

31 Min, S. K., Zhang, X., Zwiers, F. W. and Hegerl, G. C.: Human contribution to
32 more-intense precipitation extremes, *Nature*, 470(7334), 378–381,
33 doi:10.1038/nature09763, 2011.

34 Papagiannaki, K., Lagouvardos, K. and Kotroni, V.: A database of high-impact weather
35 events in Greece: A descriptive impact analysis for the period 2001–2011, *Nat.
36 Hazards Earth Syst. Sci.*, 13(3), 727–736, doi:10.5194/nhess-13-727-2013, 2013.

37 Papagiannaki, K., Lagouvardos, K., Kotroni, V. and Bezes, A.: Flash flood occurrence
38 and relation to the rainfall hazard in a highly urbanized area, *Nat. Hazards Earth
39 Syst. Sci.*, 15(8), 1859–1871, doi:10.5194/nhess-15-1859-2015, 2015.

40 Papagiannaki, K., Kotroni, V., Lagouvardos, K., Ruin, I. and Bezes, A.: Urban Area
41 Response to Flash Flood–Triggering Rainfall, Featuring Human Behavioral
42 Factors: The Case of 22 October 2015 in Attica, Greece, *Weather. Clim. Soc.*,

1 9(3), 621–638, doi:10.1175/wcas-d-16-0068.1, 2017.

2 Ratti, C., Frenchman, D., Pulselli, R. M. and Williams, S.: Mobile landscapes: Using
3 location data from cell phones for urban analysis, *Environ. Plan. B Plan. Des.*,
4 33(5), 727–748, doi:10.1068/b32047, 2006.

5 Rosner, B.: On the detection of many outliers, *Technometrics*, 17(2), 221–227, 1975.

6 Ruin, I., Creutin, J. D., Anquetin, S. and Lutoff, C.: Human exposure to flash floods -
7 Relation between flood parameters and human vulnerability during a storm of
8 September 2002 in Southern France, *J. Hydrol.*, 361(1–2), 199–213,
9 doi:10.1016/j.jhydrol.2008.07.044, 2008.

10 Ruin, I., Lutoff, C., Boudevillain, B., Creutin, J.-D., Anquetin, S., Rojo, M. B., Boissier, L.,
11 Bonnifait, L., Borga, M., Colbeau-Justin, L., Creton-Cazanave, L., Delrieu, G.,
12 Douvinet, J., Gaume, E., Gruntfest, E., Naulin, J.-P., Payrastre, O. and Vannier, O.:
13 Social and Hydrological Responses to Extreme Precipitations : An
14 Interdisciplinary Strategy for Postflood Investigation, *Weather. Clim. Soc.*, 6,
15 135–153, doi:10.1175/WCAS-D-13-00009.1, 2014.

16 Spitalar, M., Gourley, J. J., Lutoff, C., KIRSTETTER, P. E., Brilly, M. and Carr, N.: Analysis
17 of flash flood parameters and human impacts in the US from 2006 to 2012, *J.*
18 *Hydrol.*, 519(PA), 863–870, doi:10.1016/j.jhydrol.2014.07.004, 2014.

19 Taylor, H. L., Webber, D., Becker, J. S., Gruntfest, E., Wright, K. C. and Doody, B. J.: A
20 Review of People's Behavior in and around Floodwater, *Weather. Clim. Soc.*,
21 7(4), 321–332, doi:10.1175/wcas-d-14-00030.1, 2015.

22 Terti, G., Ruin, I., Anquetin, S. and Gourley, J. J.: Dynamic vulnerability factors for
23 impact-based flash flood prediction, *Nat. Hazards*, 79(3), 1481–1497,
24 doi:10.1007/s11069-015-1910-8, 2015.

25 Tukey, J. W.: *Exploratory Data Analysis*, Addison-Wesley Publishing Company., 1977.

26 UNISDR: Sendai Framework for Disaster Risk Reduction 2015-2030., 2015.

27 Vallis, O., Hochenbaum, J. and Kejariwal, A.: A Novel Technique for Long-term
28 Anomaly Detection in the Cloud, *Proceedings 6th USENIX Work. Hot Top. Cloud*
29 *Comput.* (USENIX '14), 1–6, doi:10.1016/j.pupt.2014.03.006, 2014.

30 Wang, Q. and Taylor, J. E.: Quantifying human mobility perturbation and resilience in
31 hurricane sandy, *PLoS One*, 9(11), 1–5, doi:10.1371/journal.pone.0112608,
32 2014.

33 Zhou, Y., Shen, D., Huang, N., Guo, Y., Zhang, T. and Zhang, Y.: Urban flood risk
34 assessment using storm characteristic parameters sensitive to
35 catchment-specific drainage system, *Sci. Total Environ.*, 659, 1362–1369,
36 doi:10.1016/j.scitotenv.2019.01.004, 2019.

37