Communications Earth & Environment https://doi.org/10.1038/s43247-026-03403-4

Article in Press

A combination of social media and satellite data improves flood monitoring in China

Received: 11 August 2025

Accepted: 6 March 2026

Cite this article as: Gu, H., Xiao, J., Shen, D. *et al.* A combination of social media and satellite data improves flood monitoring in China. *Commun Earth Environ* (2026). https://doi.org/10.1038/s43247-026-03403-4

Hongji Gu, Jun Xiao, Dingtao Shen, Chunxiao Zhang, Shuting Xiao, Zhuang Niu & Fei Yu

A combination of social media and satellite data improves flood monitoring in China

Hongji Gu[1,2], Jun Xiao[1,2], Dingtao Shen*[1,2], Chunxiao Zhang[3], Shuting Xiao[1,2], Zhuang Niu[1,2], Fei Yu[1,2]

1. Key Laboratory for Geographical Process Analysis & Simulation of Hubei Province, Central China Normal University, Wuhan 430079, China;

2. College of Urban and Environmental Sciences, Central China Normal University, Wuhan 430079, China;

3. School of Information Engineering, China University of Geosciences in Beijing, Beijing 100083, China.

Corresponding author: Dingtao Shen

E-mail: dingtaoshen@outlook.com

Abstract: Global climate change has increased flood frequency worldwide, yet many urban and flash floods remain poorly captured by satellite remote sensing. Here we integrate Global Satellite Mapping of Precipitation data with 92.98 million Sina Weibo posts to trace spatiotemporal flood footprints across China from 2012 to 2024. Using a cascading threshold method, we identify 6,018 rainstorm events across 370 cities. Topic modeling of posts within rainstorm affected areas detects 1,094 flood events, far exceeding the 114 and 45 events recorded in Emergency Events Database and Dartmouth Flood Observatory. Comparison with satellite imagery on Google Earth Engine shows that approximately 50% of these events were unobservable by remote sensing. While satellites capture riverine flooding in open areas, social media fills critical gaps in densely populated urban settings. This social sensing framework reveals previously undocumented flood events and their propagation patterns, offering a complementary approach that enhances traditional flood monitoring capabilities.

Key words: flood footprint; social media; social sensing; Sina Weibo

1 Introduction

Driven by global climate change, extreme weather events are occurring with increasing frequency across the world, further exacerbating regional and local flood disaster risks[1]. According to the 2024 Global Water Resources Monitoring Report released by the Global Water Monitoring Alliance in January 2025, water-related disasters in 2024 resulted in over 8,700 deaths, displaced 40 million people, and caused economic losses exceeding 550 billion USD[2]. Notably, due to the rapid growth of population and economic assets in flood-prone areas, global flood risk is projected to triple by 2050[3]. Although flood disasters, as natural phenomena, cannot be eliminated, scientific risk assessments and refined management strategies can effectively reduce their impact on socioeconomic systems[4]. Near real-time dynamic sensing of flood occurrences, timely initiation of flood warnings, rapid collection of disaster information, and the establishment of detailed flood event archives are of great significance for government agencies to carry out flood emergency management.

To better support flood disaster prevention and control, international organizations, research institutions, and governments at all levels have actively promoted the development of global and regional flood databases. These databases not only facilitate systematic reviews and statistical analyses of historical flood events but also provide critical data support for future disaster early warning, risk assessment, and policymaking[5]. At present, several major global disaster databases record flood events, including the Global Emergency Events Database (EM-DAT), the Dartmouth Flood Observatory (DFO), the Global Disaster Alert and Coordination System

(GDACS), as well as NatCatSERVICE and Sigma, which are developed and maintained by reinsurance companies[6]. Many researchers have used these global flood databases to study the social impacts of floods[7,8]. For example, Tellman et al.[9] utilized the DFO flood database, combined with daily Moderate Resolution Imaging Spectroradiometer (MODIS) satellite imagery at 250 m resolution, to estimate the inundation extent and population exposure of 913 major flood events from 2000 to 2018. Linderson et al.[10] used the EM-DAT flood database to analyze 573 major flood events in 67 countries between 1990 and 2018, revealing a notable correlation between income inequality and flood mortality rates. Although these flood databases have greatly facilitated global-scale flood research, they still suffer from incomplete event coverage and inaccurate spatiotemporal information, which may lead to a systematic underestimation of flood risk in exposure-related studies. First, most of these databases rely heavily on major news media and government reports as primary information sources, which tend to cover only large-scale disasters, resulting in the systematic omission of small and medium-sized flood events—especially those occurring in remote areas or developing countries. Second, these databases generally apply high inclusion thresholds. For instance, EM-DAT mainly records floods that have triggered national emergency responses or requested international assistance, while GDACS adopts even stricter classification criteria—Level 1 and Level 2 flood events are defined as those causing over 1,000 deaths or 800,000 displacements, and over 100 deaths or 80,000 displacements, respectively[11]. Finally, these databases do not record the exact timing of flood outbreaks across different cities. Some databases also suffer from substantial update delays—for example, as of October 2025, the most recent flood event recorded by DFO still dates to September 2024 in Vietnam.

The rapid advancement of satellite remote sensing technology has led to the widespread use of remote sensing imagery in flood monitoring and mapping, and researchers have developed a large number of remote sensing-based flood extraction methods[9,12,13]. High-resolution optical/infrared sensors can accurately delineate flood-inundated areas in the absence of cloud cover, tree canopies, or floating vegetation[14]. However, the effects of clouds and atmospheric conditions on optical/infrared imagery cannot be eliminated[15], and since most flood events occur during heavy rainfall, the application of optical/infrared remote sensing in flood monitoring is severely constrained. In contrast, microwave remote sensing can penetrate clouds and offers all-weather monitoring capabilities, providing a distinct advantage under adverse weather conditions[16,17]. Nevertheless, microwave remote sensing still faces challenges in balancing spatial and temporal resolution. Passive microwave sensors can achieve revisit times as short as 0.5 days across most of the globe[18], but their spatial resolution remains low (10–70 km). Active microwave technologies, such as Synthetic Aperture Radar (SAR), can provide high spatial resolution (1–10 m) surface observations, but their revisit cycles are typically 2–4 weeks[19]. Recently, with the development of artificial intelligence, increasing research efforts have explored its application in flood monitoring. For example, Misra et al.[20] used a lightweight MobileNet model to extract global flood events from Sentinel-1 SAR imagery covering the period from 2014 to 2024. To address the current lack of large-scale publicly available training datasets, Matosak et al.[21] developed the SenforFlood dataset, which provides abundant training samples based on Sentinel-1 and Sentinel-2 imagery, along with multiple auxiliary data sources, offering foundational support for flood-related studies. Overall, due to limitations such as cloud cover, complex terrain, and satellite revisit cycles, remote sensing remains ineffective in detecting small-scale flood events that last only one or two days and occur in urban or mountainous areas with severe occlusion—yet these types of floods are precisely the most frequent in occurrence.

With the rise of social media, researchers have increasingly recognized the tremendous advantages of web-based big data in information acquisition and dissemination. Liu et al.[22] first proposed the concept of Social

Sensing, referring to geospatial big data derived from sources such as smart card data from public transportation systems, and social media platforms. Compared with remote sensing data, socially sensed data offers greater temporal flexibility, can convey richer semantic information, and is unconstrained by terrain conditions. Its observational capability mainly depends on population distribution, giving it a distinct advantage in densely populated urban areas, where it can rapidly capture disaster information and public responses. However, in sparsely populated rural and remote regions, their monitoring capability remains relatively limited. In urban flood research, socially sensed data have gradually become an important data source[23,24], as they can effectively compensate for the extensive monitoring blind spots caused by the insufficient deployment of inundation sensors in urban flood monitoring. Numerous studies have explored the potential of social media data in disaster monitoring, impact assessment, and emergency response[24-28]. Existing research has demonstrated that social media data can effectively complement traditional monitoring approaches by enabling high-resolution analysis of flood-related features such as inundation locations, water depth, and rescue information[29-32]. Integrating social media data with conventional hydrological and hydrodynamic models also holds great promise. For instance, Scotti et al.[33] proposed a method to calibrate the Hydrologic Engineering Center's River Analysis System (HEC-RAS) model using social media data. Their study leveraged rich, localized information available from urban built-up areas, enabling the calibrated model to achieve higher simulation accuracy. In addition, researchers have developed specific algorithms for flood outbreak detection based on social media data. For example, Shoyama et al.[34] applied the E-divisive with Medians method to detect flood events by analyzing changes in the volume of tweets on social media platforms. Similarly, de Bruijn et al.[35] used a threshold-based method to detect flood outbreaks by monitoring variations in the frequency of flood-related posts. These studies strongly demonstrate the practical potential of social media for detecting and sensing global flood events, highlighting its importance as a promising pathway for future global flood monitoring and situational awareness.

China has experienced increasingly severe flood disasters in recent years, with annual direct economic losses steadily rising[36]. To comprehensively elucidate the current situation, this study designs a social sensing framework (as shown in Figure 1) that integrates Global Satellite Mapping of Precipitation (GSMaP) satellite precipitation data with Sina Weibo data to trace spatiotemporal flood footprints across China from 2012 to 2024. The main contributions of this study are as follows: (1) it enables city-scale dynamic flood sensing, refining spatiotemporal flood event characterization beyond traditional hydrological and remote sensing methods. (2) it identifies far more flood events than global databases, providing a more comprehensive depiction of China's flood distribution. (3) It offers valuable insights for the research community and governmental agencies in leveraging socially sensed data for refined global flood management and serves as a meaningful supplement to existing global flood databases.

2 Results

2.1 Spatiotemporal Distribution of Sina Weibo Posts

This study used the trained Bidirectional Encoder Representations from Transformers (BERT) model to classify Sina Weibo data from 2012 to 2024, filtering out approximately 4.87 million posts related to rainstorms and floods from a total of 92.98 million Weibo posts. Subsequently, geographic location parsing was performed on 4.87 million of these posts using the Chinese Province City Area mapper (CPCA) module, assigning the posts to 370 cities across China.

Figure 2 presents the statistics of rainstorm- and flood-related Sina Weibo posts classified by the BERT model from 2012 to 2024. As shown in Figure 2(a), from a spatial distribution perspective, rainstorm- and flood-related Weibo posts are primarily concentrated in the Yellow River Basin, the Yangtze River Basin, and the

southeastern coastal regions. These areas have higher population densities and abundant precipitation, resulting in high public attention to rainstorm and flood topics on social media platforms. As shown in Figure 2(b), in terms of temporal distribution, the total number of Weibo posts between 2012–2017 and 2018–2024 exhibits two distinctly different magnitudes, which may result from several factors. First, from a platform mechanism perspective, Sina Weibo may not have retained early user posts due to operational cost considerations, leading to missing data in the earlier period. Second, a deeper reason may lie in the platform's search strategies and application programming interface limitations, which restrict the number of early historical posts that can be crawled.

Although the total number of downloaded Weibo posts fluctuated greatly over more than a decade, the year-to-year variation in posts related to rainstorms and floods is much smaller than that of the total Weibo volume. The number of rainstorm- and flood-related posts shows a notable peak in 2021, which was caused by the extreme rainstorm event in Henan, China, in July 2021. Figure 2(c) shows a scatter plot of the number of rainstorm- and flood-related Weibo posts for cities across China. It can be observed that during 2012–2017, the annual number of related posts in all cities was below 10,000. Although over 95% of cities still had fewer than 10,000 related posts annually during 2018–2024, the disparity in the number of posts among cities has increased since 2018, as evidenced by the decreasing concentration of scatter points in the plot.

2.2 Spatiotemporal Distribution of Rainstorm Events

This study employed a cascade threshold method to extract rainstorm objects in China from 2012 to 2024, which were then aggregated into rainstorm events at the city scale. Figure 3 presents the statistics of rainstorm events in China during 2012–2024. Specifically, Figure 3(a) shows the spatial distribution of rainstorm objects at the grid scale, while Figure 3(b) displays the spatial distribution of rainstorm events at the city scale. Both distributions exhibit a general decreasing trend from the southeastern coastal areas toward the northwestern inland, reflecting pronounced regional differences in the spatiotemporal distribution of precipitation across China.

The statistics indicate that Sansha City experienced as many as 529 rainstorm events, far exceeding the counts in Taiwan Province (172 events) and Chongqing City (148 events). In contrast, Wuhai City, Hu Yanghe City, and Shihezi City, located in the arid northern and northwestern regions of China with relatively small urban areas, did not experience any rainstorm events during the study period. Figures 3(c) and 3(d) illustrate the temporal distribution characteristics of rainstorm events, showing clear seasonality, with most events occurring between April and October. On average, about 10 rainstorm events occur on the most concentrated dates each year, with a peak of 20 events recorded in 2020.

Figure 4 shows the annual statistics of rainstorm objects, rainstorm events, and flood events in China from 2012 to 2024. It can be observed that the period from 2016 to 2020 was a peak phase for the occurrence of both rainstorm events and flood disasters. Despite some interannual fluctuations, the overall numbers during this period were substantially higher than those in other years, except for 2024. In 2024, the number of rainstorm objects surged to 4,104, the highest among all the years analyzed and far exceeding the count in 2023. However, the number of rainstorm events and flood events in 2024 remained at a moderate level, with much smaller increases compared to the sharp rise in rainstorm objects. From an interannual perspective, 2018 recorded the highest number of rainstorm events (558), while 2020 ranked third in rainstorm event counts (525) but had the highest number of flood events (113), indicating that rainstorm events in 2020 were more likely to trigger flood disasters.

2.3 Spatiotemporal Distribution of Flood Footprints

At the city scale, this study established a spatiotemporal association between rainstorm events and Weibo data. Based on this, Weibo posts from each city covered by a rainstorm event were extracted within the time window from the event's onset to 24 hours after its end. Topic detection was then conducted using the Latent Dirichlet Allocation (LDA) model. The top 10 topic keywords output by LDA were matched against flood-related keywords to determine whether a city experienced a flood event during the rainstorm. If any city covered by the rainstorm event experienced a flood disaster, the rainstorm event was considered to have triggered a flood event. A total of 1,330 flood events automatically identified by the LDA model were manually verified in this study. The verification involved manually examining the Weibo posts from cities where floods were detected to confirm the authenticity of each event. The results indicate that the automated identification method has relatively high accuracy, ultimately confirming 1,094 real flood events, with an overall precision rate of approximately 82.25%.

Figure 5 shows the statistical results of flood events in China from 2012 to 2024. Figure 5(a) illustrates the spatial distribution of flood events at the city scale, based on manual verification. The results indicate that flood events are primarily concentrated in the middle and lower reaches of the Yangtze River Basin and the Pearl River Basin, exhibiting a clear pattern of regional clustering. Among all cities, Haikou recorded the highest number of flood events during the study period, with a total of 38 occurrences. Although Sansha City and Taiwan Province experienced a large number of rainstorm events, no flood events were detected due to the lack of Weibo data required for flood identification. Similarly, most cities in the northwestern region of China had zero recorded flood events during the study period. This outcome not only reflects the naturally arid climate and infrequency of extreme precipitation in these regions but is also closely related to the limited availability of Weibo data in these areas.

Furthermore, this study analyzed the spatiotemporal footprint of floods in China. First, flood events were classified into localized floods, simultaneous multi-city floods, and cascading wave floods. Among the 1,094 flood events, there were 742 localized floods, 45 simultaneous multi-city floods, and 307 cascading wave floods. As shown in Figure 5(b), the number of events follows the pattern: localized floods > cascading wave floods > simultaneous multi-city floods, with localized floods—those occurring in only one city—being the most common. The box plot in Figure 5(c), which shows flood event duration and the number of affected cities, reveals an increasing dispersion in event duration over the years, indicating growing complexity and variability in flood disasters. From 2012 to 2022, the interannual variation in flood event impact scope (i.e., number of affected cities) remained relatively stable, with most fluctuations occurring in the outliers. In contrast, the years 2023 and 2024 exhibited a clear concentration trend, with approximately 76% of flood events affecting only a single city, indicating a narrowing of spatial impact. Figure 5(d) presents the statistics on flood event duration and the number of affected cities. The results show that most flood events in China are short in duration and limited in spatial extent—typically affecting fewer than five cities and lasting less than five days. This suggests that social media can capture many small-scale flood disasters that have long been overlooked by mainstream news outlets and traditional natural disaster databases.

As shown in Figure 6, this study takes Flood Event No. 967, which occurred from June 4 to 24, 2022, as an example to illustrate the spatiotemporal evolution of rainstorm objects, rainstorm events, and flood events. This flood event affected a wide region, covering 27 cities across five provinces: Guangdong, Guangxi, Fujian, Hunan, and Zhejiang. For clarity, only the five most severely affected cities are used as case examples. Figure 6(a) shows the evolution of three rainstorm objects at 12:00 noon from June 18 to June 20. All three rainstorm objects

covered Shaoguan City and were thus aggregated into a single rainstorm event, whose spatial extent includes all cities affected by any of the three objects. By further extracting Weibo posts from these cities during the rainstorm period and applying LDA topic modeling combined with manual verification, five cities were identified as having experienced flood disasters: Shaoguan, Qingyuan, Zhaoqing, Ganzhou, and Yongzhou. Figure 6(b) presents a selection of flood-related images posted on social media from Shaoguan, Qingyuan, and Yongzhou during the flood event. Figure 6(c) shows the temporal trends in the number of rainstorm- and flood-related Weibo posts from the five cities during the event. Notably, Qingyuan experienced a sharp rise in Weibo activity on June 21–22, far exceeding the other cities. Qingyuan was also the most severely affected area in this flood event, demonstrating the aggregation effect of social media in disseminating disaster information.

2.4 Comparative Analysis with Global Flood Databases

This study conducted a comparative analysis between the 1,094 validated flood events and the EM-DAT and DFO flood databases (noting that the DFO database is only updated through 2021), with results shown in Table 1. During the comparative analysis, we considered a flood event from this study to correspond to an event in the EM-DAT or DFO databases if there was an overlap in the affected cities and the duration of the event. Furthermore, due to the spatial scale differences between the provincial and city levels, it is possible that multiple flood events extracted in this study correspond to a single flood event recorded in the EM-DAT and DFO databases.

For the EM-DAT flood database, this study detected 77 flood events out of the 114 events recorded in EM-DAT, resulting in a hit rate of 67%. Conversely, EM-DAT recorded 284 flood events out of the 1,094 flood events extracted in this study, yielding a hit rate of 25%. For the DFO flood database, this study detected 29 flood events out of 45 recorded in DFO, corresponding to a hit rate of 64%. Meanwhile, DFO recorded only 80 flood events out of the 914 flood events extracted by this study from 2012 to 2021, with a hit rate of 8%. The notably lower hit rate of the DFO database compared to EM-DAT is primarily due to the smaller number of flood events recorded in DFO, as well as its use of vector files to mark more precise flood impact areas, whereas EM-DAT only records affected administrative regions in textual form. Additionally, because the number of flood events extracted by this study far exceeds those recorded in both flood databases, there is a large interannual variability in hit rates. For example, the hit rate of this study's extracted flood events against the DFO database ranges from a minimum of 33% to a maximum of 100%. Compared with the spatial and temporal coverage of flood events recorded in the EM-DAT and DFO databases, the flood events extracted in this study exhibit clear advantages in both dimensions.

As illustrated in Supplementary Figure 1, taking Flood Event No. 835 extracted in this study as an example, the corresponding event IDs in the EM-DAT and DFO databases are 2021-0380-CHN and 5089, respectively. This study first extracted the rainstorm event associated with this flood event, which affected four provinces: Jiangxi, Anhui, Zhejiang, and Fujian (Supplementary Figure 1(a)). Based on the movement trajectory of the rainstorm event across these provinces, floods were sequentially identified in Shangrao City and Jiujiang City in Jiangxi Province, and Tongling City in Anhui Province. Moreover, the occurrence and end times of flood disasters in each city can be precisely determined at the hourly scale. In the EM-DAT flood database, the affected area of each flood event is marked by combinations of administrative divisions at the provincial, city, and county levels, with most records only precise to the provincial level—for example, in this case, EM-DAT locates the flood event only at the Jiangxi Province level (Supplementary Figure 1(b)). The DFO database uses a different approach by providing vector files representing the global flood event impact areas, with polygons indicating affected regions. As shown in the Supplementary Figure 1(c), there remain some differences between the flood

impact areas marked by DFO and those extracted in this study.

2.5 Comparison with Satellite Remote Sensing Observations

This study conducted a comparative validation of the extracted flood events against satellite remote sensing observations based on the Google Earth Engine (GEE) platform. During the comparison, two types of remote sensing data were used: microwave data (Sentinel-1) and optical data (Sentinel-2, Landsat 4/5/7/8/9), involving a total of seven remote sensing data sources. For different remote sensing images, water extraction was performed using the Otsu's thresholding method on Sentinel-1 images and the Normalized Difference Water Index (NDWI) on optical images. In the validation process, for each flood event and its affected cities, remote sensing images from one month prior to the event were extracted to construct a reference water body extent. This reference was visually compared with water extraction results during the flood event period to determine whether the flood event was observable by satellite. Due to the inherent temporal resolution limitations (revisit cycles) of the satellite platforms, a pragmatic image selection window was adopted. This window spans from the flood event start date to 10 days after the event end date, as a compromise to maximize the likelihood of obtaining at least one usable observation per event. It is recognized that flood dynamics may change within this window, potentially affecting direct comparability. Additionally, a 50% cloud cover threshold was applied to optical images, with images exceeding this threshold regarded as unusable.

As shown in Table 2, approximately 50% of the 1,094 flood events identified in this study were observable by satellite remote sensing. Among these, the number of flood events detected by microwave remote sensing was about twice that detected by optical remote sensing. Since Sentinel-1A was launched in 2014, microwave observations were not available for 2012 and 2013. Before Sentinel-1B began operation in 2016, the system operated in single-satellite mode with a revisit cycle of 12 days, limiting observation capability during 2014–2015. After the dual-satellite system was established in 2016, the revisit cycle shortened to 6 days, increasing the number of observed flood events, exceeding 50 annually between 2016 and 2021, and peaking at 97 events in 2020. In contrast, optical remote sensing was overall less effective due to interference from clouds and fog.

This study further conducted a detailed analysis of the satellite remote sensing capabilities for flood event observation, with the results shown in Figure 7(a). Between 2015 and 2020, the proportion of flood events observable only by microwave remote sensing was relatively high. From 2021 to 2024, about half of the flood events detected by microwave remote sensing were also captured by optical imagery. However, the number of events observable solely by optical remote sensing remained consistently low, highlighting the irreplaceable advantage of microwave remote sensing in flood monitoring. Additionally, this study performed an attribution analysis on flood events not detected by satellite remote sensing, categorizing the causes into three types: 1) cloud and fog obscuration or insufficient image coverage; 2) lack of available imagery data; and 3) absence of identifiable flood water features in the imagery. It should be noted that this statistic only includes events unobservable by both microwave and optical sensors; events detected by either sensor were excluded. As shown in Figure 7(b), before 2014, insufficient availability of remote sensing imagery was the primary reason for poor satellite observation performance. After 2014, the dominant factor shifted to the lack of distinct flood features in the imagery, while the impact of cloud and fog obscuration or insufficient coverage gradually declined with the development of microwave remote sensing.

Taking Flood Event No. 898 as an example (Figure 8), this study conducted remote sensing flood mapping in Fuzhou City, Fujian Province, China. The event occurred from late July to early August 2023. Figure 8(a) shows the temporal distribution of available multi-source satellite imagery during the flood, indicating that most images were captured within a few days after the disaster. Despite this temporal lag, Figures 8(b–g) demonstrate

that satellites were still able to effectively identify the flood inundation extent. Among them, optical imagery extracted a larger water area than microwave imagery. This is mainly because Fuzhou is located in an estuarine basin surrounded by dense mountains and pronounced terrain variation, whereas flat basins are concentrated in the urban built-up areas, and the complex terrain substantially limits the flood detection capability of microwave imagery. In addition, Figures 8(d) and 8(g) indicate that the inundated areas identified by remote sensing were primarily outside the urban built-up area, whereas flood points extracted from social sensing data were more concentrated within the urban area. This difference is mainly due to data characteristics and observation stages. Social sensing is more sensitive to disaster responses in densely populated areas but under-monitors suburban areas, while remote sensing is limited by urban structures and high-rise buildings, making it difficult to capture flooding within urban areas. In summary, social sensing and remote sensing exhibit clear spatiotemporal complementarity in flood detection: spatially, social sensing compensates for remote sensing deficiencies in urban and complex terrain areas, while remote sensing enhances monitoring in remote areas; temporally, social sensing excels at real-time detection during rainstorms, whereas remote sensing is more suitable for post-disaster inundation detection.

3 Discussion

This study integrates satellite precipitation with social media data to trace spatiotemporal flood footprints, addressing the limitations of both global databases in capturing small-to-medium floods and remote sensing in monitoring urban pluvial and flash floods. Compared to previous studies such as de Bruijn et al.[35] and Shen et al.[37], which detect flood events solely based on social media data, our approach offers clear advantages. Those studies typically rely on abrupt changes in the volume or frequency of social media posts to detect flood outbreaks. However, due to the large spatial and temporal variation in social media activity across different cities and years, the design of activation thresholds for flood detection requires careful tuning and still struggles to accommodate such data heterogeneity. For instance, Shen et al.[37] reported poor detection performance in earlier years when Weibo data was sparse. In contrast, as illustrated in Figures 2 and 4 of this study, the number of detected flood events in earlier years with limited Weibo data is not substantially lower than in more recent years with much higher data volume. Moreover, prior studies often conduct flood detection at the scale of a single city or province, making it difficult to trace the progression of flood events with fine granularity. To overcome this limitation, our study introduces the concept of flood footprint tracing, enabling each detected flood event to be associated with precise spatiotemporal attributes, including its propagation across multiple cities. Furthermore, the integrated framework proposed in this study is highly suitable for incorporation into various flood monitoring information platforms. By integrating satellite precipitation data and social media data, it can perform near-real-time social media data cleaning and heavy rainfall signal detection, enabling city-scale, hourly flood outbreak detection and visualization of flood propagation. This facilitates early flood warning and provides valuable time for flood emergency response and rescue operations.

However, this study also has certain limitations. First, flood events can be triggered by a variety of complex factors, whereas the proposed approach primarily focuses on rainstorm-induced floods and is largely unable to capture events caused by dam failures, snowmelt, or other non-precipitation-driven mechanisms. Second, satellite-based precipitation products inherently contain estimation errors and cannot be fully equated with ground-truth rainfall measurements. Previous research (e.g., Wang[38]) has shown that GSMaP tends to underestimate extreme precipitation, which may lead to undetected rainstorm events and, consequently, missed flood events in certain cities. Third, social media data exhibit systematic urban biases in their spatial distribution. This bias arises from uneven population density, communication infrastructure, and social media usage, which

can result in a lack of real-time records and social feedback for flood disasters in remote areas.

Moreover, there are still several aspects of the methodology that require further refinement and optimization in future work. First, the CPCA address parsing tool used in this study is not yet sufficiently intelligent. Future research could consider incorporating large language models with stronger semantic understanding capabilities (such as ChatGPT or DeepSeek) to improve both the quantity and accuracy of address parsing. Second, the end time of a flood was approximated by the end time of the rainstorm. This aligns with the intuitive perception that once the rain stops, the skies clear and floodwaters recede, yet it underrepresents the complete dissipation of flood-induced waterlogging. Finally, despite the regional limitations introduced by using Sina Weibo as the primary data source, the proposed flood footprint spatiotemporal tracing framework still has strong portability and potential for global application. This scalability is mainly reflected in the modular design of the methodological framework: in the natural language processing part, except for the CPCA, which needs to be replaced with place-name address parsing tools specifically designed for the English context, such as Stanford NER, Gazetteer, and Geoparser, the BERT and LDA models were originally developed for English-language scenarios. By replacing or adjusting the geographic parsing module to adapt to different language structures and corpus characteristics, and integrating global mainstream social media platforms (such as Twitter, Facebook, Instagram, etc.), this research framework can be replicated and promoted in multilingual, multi-platform environments.

4 Method

Supplementary Figure 2 presents the workflow for tracking spatiotemporal flood footprints by integrating satellite precipitation and social media data. The entire technical process consists of three main steps:

Step 1: Data Preprocessing — This includes masking the GSMaP satellite precipitation grids to extract hourly precipitation data covering the Chinese region. Meanwhile, 44 predefined keywords are used to download rainstorm- and flood-related posts from the Sina Weibo platform (Supplementary Figure 2(a)).

Step 2: Rainstorm Object Extraction, Social Media Filtering, and Geoparsing — For the GSMaP precipitation data, a cascading threshold method is applied to identify rainstorm objects, which are then aggregated into rainstorm events. Spatiotemporal associations are established between these events and cities (Supplementary Figure 2(b)). For the social media data, a BERT model is used to classify each post, retaining only those genuinely related to floods. The CPCA geoparsing tool is then employed to extract city-level location information from each post. Flood-related Weibo posts are assigned to corresponding cities, and spatiotemporal associations between posts and cities are established (Supplementary Figure 2(c)).

Step 3: Flood Event Detection — For each city affected by a rainstorm event, Weibo posts posted during the rainstorm period are extracted (Supplementary Figure 2(d)). An LDA topic model is applied to detect flood-related topics. Based on the temporal sequence of flood occurrences across cities, a spatiotemporal flood footprint dataset is automatically constructed.

4.1 Data

4.1.1 GSMaP Satellite Precipitation Data

This study selects satellite precipitation products (SPPs) as the source of precipitation data. Compared to traditional ground-based rain gauge observations, SPPs offer advantages such as long temporal coverage, wide spatial extent, high temporal resolution, and spatial continuity[39]. Their grid-based data storage format is particularly valuable for the extraction of rainstorm objects. The study by Gu et al.[40] indicates that among various SPPs, GSMaP performs best in rainfall estimation across most river basins in China, particularly in humid

climate regions. Furthermore, the spatiotemporal distribution of floods in China closely aligns with that of extreme precipitation events[41], making GSMaP particularly suitable for flood disaster research in the Chinese context.

The GSMaP is a satellite-based precipitation product developed by the Japan Aerospace Exploration Agency (JAXA) in 2002. GSMaP primarily estimates rainfall using multi-band passive microwave and infrared radiometers onboard GPM satellites, providing global hourly precipitation data at a spatial resolution of 0.1° × 0.1°. The GSMaP product suite includes GSMaP_NRT, GSMaP_MVK, and GSMaP_Gauge. Among them, GSMaP_Gauge is a high-accuracy satellite precipitation product derived by calibrating GSMaP_MVK with data from over 3,000 Climate Prediction Center rain gauge stations worldwide[42]. This dataset is freely available from the JAXA website (https://sharaku.eorc.jaxa.jp/GSMaP/).

This study ultimately selected the GSMaP_Gauge_v6 hourly precipitation product. After spatial and temporal clipping, a total of 113,976 scenes of gridded data covering China from 2012 to 2024 were obtained.

4.1.2 Sina Weibo Data

Currently, most flood-related studies using social media data focus on international platforms such as Twitter, Facebook, and Flickr[26,43-45]. However, the coverage of these platforms in China is very limited, as the most widely used social media platform in the country is Sina Weibo—a large-scale Chinese-language platform similar to Twitter. By the end of 2024, Sina Weibo had 590 million monthly active users and 260 million daily active users, generating hundreds of millions of Chinese-language posts each day. Therefore, this study selected Sina Weibo as the source of social media data.

This study employed a keyword-based retrieval method to collect flood-related social media data from the Sina Weibo platform. This approach offers high filtering efficiency and effectively reduces irrelevant noise. The key to downloading Weibo data lies in the careful selection of keywords—poorly chosen keywords may result in ineffective filtering or the omission of a large number of posts related to specific flood events. To enhance both the coverage and accuracy of data retrieval, the selected keywords included not only flood-related terms but also vocabulary associated with rainstorms, given the close relationship between flood disasters and heavy rainfall. Social media users often mention severe weather conditions when posting about flood events. As shown in Supplementary Table 1, a total of 44 Chinese keywords highly relevant to rainstorms and floods were ultimately used in this study. Based on these keywords, a web crawler was developed in Python to retrieve a total of 92.98 million Weibo posts from 2012 to 2024. The data were collected from the Sina Weibo platform (https://weibo.com/).

4.2 City-Level Detection of Rainstorm Events

Flood outbreaks are typically triggered by extreme precipitation. Based on this fundamental assumption, accurately detecting flood events from social media data first requires identifying where and when heavy rainfall occurred. For satellite precipitation grid data, the simplest strategy is to directly detect rainstorm events at the pixel level. However, since location information extracted from Weibo posts cannot be refined to the pixel scale, the most feasible approach is to match rainstorm events with social media data at the city level. To achieve this spatiotemporal matching, this study adopts a three-step strategy. First, rainstorms are detected at the pixel level. Second, adjacent rainstorm-affected pixels are aggregated in space and time to construct rainstorm objects with three-dimensional spatiotemporal characteristics. Finally, these rainstorm objects are further aggregated to extract city-scale rainstorm events, enabling the capture of the movement trajectories of rainstorms across cities.

To extract rainstorm events from satellite precipitation data, it is first necessary to define rainstorms at the

pixel level. Most studies adopt a threshold-based method to define heavy rainfall[46,47]. However, since rainfall intensity during a rainstorm often fluctuates and does not always exceed a fixed threshold, this method tends to overlook periods of relatively low intensity within a rainstorm event. In contrast, rainstorm definitions that consider rainfall duration can more comprehensively capture the full rainstorm process and better reflect real-world precipitation patterns. As shown in Supplementary Table 2, this study refers to the rainstorm classification standards issued by the China Meteorological Administration and divides rainstorms into three levels. When the continuous rainfall duration is less than 24 hours, events with accumulated precipitation exceeding 48 mm, 96 mm, and 240 mm are classified as Level 1, 2, and 3 rainstorms, respectively. When the continuous rainfall duration exceeds 24 hours, events with average rainfall intensities above 2 mm/h, 4 mm/h, and 10 mm/h are categorized as Level 1, 2, and 3 rainstorms, respectively.

This study adopts the cascading threshold method proposed by Skok et al.[48] to progressively aggregate rainstorm-affected pixels into rainstorm objects and ultimately construct a complete set of rainstorm events. The object-based approach to rainstorm detection was first introduced by Davis et al.[49] through the Method for Object-Based Diagnostic Evaluation (MODE) framework. The core idea is to identify spatially contiguous regions in the precipitation field that exceed a predefined threshold as rainstorm objects in the spatial dimension. In the temporal dimension, two-dimensional rainstorm objects are merged by applying a distance threshold between the centers of rainstorm objects in adjacent time steps, thereby forming spatiotemporally continuous rainstorm objects. Skok et al.[50] refined the MODE framework by introducing the concept that rainstorm objects should maintain closure in both space and time. Building on this, Skok et al.[48] further optimized the methodology by incorporating several key improvements, including object growth, cascading thresholds, and forward object identification, which substantially enhanced the accuracy of rainstorm object detection.

The core idea of the cascading threshold-based rainstorm object extraction method is to use the highest-level rainstorm area as the core of the rainstorm object and progressively absorb surrounding lower-level rainstorm areas. This study integrates the cascading threshold method, object expansion algorithm, and forward object identification technique to construct the rainstorm object detection workflow illustrated in Supplementary Figure 3. As shown in Supplementary Figure 3(a), hourly GSMaP precipitation grid data are first processed to extract time series rainfall values for each grid cell, treating continuous non-zero values as an individual rainfall process. Each rainfall event is then classified into Levels 1 to 3 based on the rainstorm definitions in Supplementary Table 2. Events that do not meet the threshold for rainstorm classification are marked as 0. The spatial structure of each rainfall event is restored to form a corresponding rainstorm level raster. Supplementary Figure 3(b) illustrates the expansion process of rainstorm objects at a given time step t. This process begins with the highest-level rainstorm grids at the current time step as the core and gradually expands to adjacent grids with the next lower rainstorm level, until all grids assigned a rainstorm level are absorbed into complete two-dimensional rainstorm objects. If multiple non-contiguous high-level rainstorm regions exist at the same time step, they are treated as separate rainstorm objects and expanded independently. Supplementary Figure 3(c) shows how two-dimensional rainstorm objects evolve into three-dimensional objects over time. Two-dimensional rainstorm objects identified in the initial time step are assigned unique object IDs. Then, each two-dimensional rainstorm object at a given time step is spatially overlapped with those from the previous time step. Each object inherits the ID of the overlapping object with the largest intersection area, thus maintaining temporal continuity. If a two-dimensional rainstorm object has no overlap with any object from the previous time step, it is considered a new rainstorm object.

Furthermore, this study takes cities as the basic spatial unit and aggregates rainstorm objects into rainstorm

events based on their spatiotemporal distribution within each city. A common approach is to group rainstorm objects into the same event if they overlap in space and time. However, this method may lead to misclassification. For instance, when two rainstorm objects affect the same city in close succession, they may be separate processes from a data perspective, but the public often perceives and discusses them on social media as a single, continuous rainstorm event. To address this issue, a time tolerance is introduced in the aggregation process. Specifically, if multiple rainstorm objects affect the same city within a 24-hour period, they are considered part of the same rainstorm event. Otherwise, they are treated as separate, independent events.

4.3 City-Level Flood Event Perception

4.3.1 Data Cleaning and Classification of Sina Weibo Posts

During the process of crawling Weibo data from the Sina Weibo platform, although keyword-based searches can filter out most irrelevant information, the complexity of Chinese text means that not all posts containing keywords related to rainstorms or floods are actually associated with real flood events. As a result, the downloaded Weibo data still contains a large amount of noise. To further eliminate irrelevant noisy data and retain only posts related to actual flood events, this study employs the BERT deep learning model to classify the Weibo posts and determine whether their content is directly related to rainstorms or floods. BERT, proposed by Devlin et al.[51], is a neural network model for natural language processing (NLP) that enhances contextual understanding of unannotated text. It has been widely applied to various NLP tasks. Due to its powerful context-aware word embedding capabilities, BERT has become one of the core technologies across nearly all subfields of NLP[52]. At present, the model has been successfully applied in multiple domains and has demonstrated strong performance in text classification tasks[Error! Reference source not found.-55].

The BERT model program used in this study was obtained from GitHub: https://github.com/illiterate/BertClassifier. This program is a library specifically designed for Chinese language processing. To train the BERT model, this study collected Weibo posts using the keyword "Zhengzhou rainstorm," targeting the severe rainstorm event that occurred in Zhengzhou, Henan Province, in July 2021. A total of 28,000 rainstorm- or flood-related Weibo posts were manually selected as positive samples. Additionally, 28,000 Weibo posts unrelated to floods were randomly selected from data downloaded during a non-flood period in April 2023, serving as negative samples. Together, these formed a dataset of 56,000 posts. These samples were divided into training and validation sets at a ratio of 8:2. The constructed validation set contains 11,327 Weibo posts, including 5,791 posts related to rainstorm and flooding, and 5,536 posts unrelated to such events. To evaluate the classification performance of the model, three commonly used metrics—Precision, Recall, and F1-score—were adopted. Their calculation formulas are shown below (Equations 1–3).

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$F1 - score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (3)$$

In the equations, TP (true positive) refers to the number of Weibo posts that were classified as rainstorm- or flood-related by both the model and manual annotation. FP (false positive) refers to the number of posts classified as rainstorm- or flood-related by the model but not by manual annotation. FN (false negative) refers to the number of posts that were manually labeled as rainstorm- or flood-related but were classified otherwise by the model. During the BERT model training process, a total of 10 epochs were conducted, and the accuracy

evaluation results are presented in Supplementary Table 3. The results show that the Precision, Recall, and F1-score of the BERT model all exceeded 0.95, indicating that the trained BERT model achieved high accuracy and strong performance in text classification. Using the trained BERT model, this study classified all Weibo data from 2012 to 2024, ultimately identifying approximately 4.87 million posts related to rainstorms and floods.

4.3.2 Location Parsing of Rainstorm- or Flood-Related Weibo Posts

After using the BERT model to filter all Weibo posts related to rainstorms and floods, it is necessary to further identify the geographic location information contained in the posts so that they can be assigned to corresponding cities, thereby enabling spatiotemporal association with rainstorm events. Based on an analysis of the characteristics of Weibo data, we classify the texts into four categories: (1) posts that contain only city (district/county) names; (2) posts that contain both city (district/county) names and address (place name) information; (3) posts that do not contain city (district/county) names but include address (place name) information; and (4) posts that contain neither city (district/county) names nor address (place name) information.

In this study, we employed the Chinese Province City Area mapper (CPCA) tool to extract location information from the Weibo text. CPCA is a Python module specifically designed for parsing administrative divisions from Chinese text. It supports mapping, validation, and basic visualization of geographic entities (https://github.com/DQinYuan/chinese_province_city_area_mapper). With its high precision and low computational cost, CPCA is particularly well-suited for extracting geographic information in China. When text is input into the CPCA program, it returns strings in the format of {province, city, district (county), address (place name)}. In addition, CPCA features an automatic completion function: when only a district (county) name appears in the text, it can automatically supplement the corresponding province and city names.

An analysis of the above four categories of Weibo posts shows that the first category is the most abundant and is the primary focus of this study. A typical example is “Heavy rainstorm in Zhengzhou today, streets turned into rivers”, for which CPCA can correctly parse the result as {Henan Province, Zhengzhou City, NULL, NULL}. All such posts are retained and assigned to their corresponding cities. For the second category, CPCA can extract complete address (place name) information only when the city (district/county) name and the address (place name) appear consecutively in the text. For example, in “Wan’antang West Industrial Zone, Luojiang District, Quanzhou City”, CPCA parses the result as {Fujian Province, Quanzhou City, Luojiang District, Wan’antang West Industrial Zone}. However, CPCA cannot automatically concatenate city (district/county) names with address (place name) information when they are separated in the text. For instance, in “The heavy rainstorm in Wuhan lasted from last night until noon today, and Central China Normal University was flooded”, CPCA parses the result as {Hubei Province, Wuhan City, NULL, NULL}. Although the address (place name) information is missing, such posts are still retained and included in subsequent LDA topic analysis. For the third category, these posts contain only place names (e.g., road names, restaurant names, or other POIs) without explicitly mentioning the corresponding province, city, or district/county. CPCA is unable to parse such texts, and therefore these posts are discarded. Compared with the first and second categories, the number of such posts is relatively small, and their exclusion has a limited impact on subsequent LDA topic analysis. For the fourth category, because neither city (district/county) names nor address (place name) information are present, CPCA returns results such as {NULL, NULL, NULL, NULL} or only a province name, e.g., {province name, NULL, NULL, NULL}. These posts are automatically discarded.

The above analysis indicates that CPCA performs well in parsing administrative division information, but its ability to extract Point of Interest (POI) information remains limited. It should be noted that this study operates

at the city scale and aims to sense flood events by mining flood-related topics posted by users that are associated with specific cities, rather than by identifying flooded POIs. Therefore, the CPCA tool is sufficient for the purposes of this study. We constructed a sample consisting of 400 Weibo posts, from which the geographic locations of flood occurrences were manually extracted and compared with the locations parsed by CPCA. The results show that only one post was incorrectly parsed: CPCA misidentified Foshan City as Shanwei City. Overall, the geographic location parsing accuracy was close to 100%. However, on the other hand, Weibo texts often contain multiple city names, and CPCA cannot, as a human would, determine which city actually experienced flooding. As a result, CPCA may extract multiple geographic locations. Among the 400 tested posts, 53 posts contained geographic locations that were not flood occurrence locations, accounting for 13.25%. In this study, such posts containing multiple city names are assigned to each identified city (i.e., the same post is retained for all corresponding cities). This operation introduces a certain degree of uncertainty. To mitigate flood sensing errors arising from this uncertainty, satellite precipitation data are incorporated to identify rainstorm events in each city. Since most incorrectly assigned posts occur during non-rainstorm periods, and LDA topic analysis is conducted only during city-level rainstorm periods, the majority of these incorrectly assigned posts are excluded from the LDA analysis and therefore have a limited impact on the results.

In this study, the CPCA program is used to perform geographic location parsing on Weibo posts related to rainstorms and flooding after BERT-based classification. Based on the parsing results, each post is automatically assigned to its corresponding city, so that each city has its own Sina Weibo dataset. For each city-level Weibo dataset, posts are sorted by their posting time, thereby enabling the automatic construction of city-scale Sina Weibo time series data.

4.3.3 Flood Event Outbreak Detection

Although all Weibo posts had been classified using the BERT model, and only those related to rainstorms and floods were retained, there still exists a relatively high risk of misclassification, since the BERT model makes judgments based solely on individual posts and lacks the ability to recognize collective behavioral patterns. First, the classification accuracy of the BERT model is not 100%, and misclassifications may occur. Second, a single Weibo post often reflects individual subjectivity and randomness, lacking clear event-indicative features. If flood disasters were inferred directly from individual posts, it would likely generate a large number of false alarms. For a given city, only when a large number of users post flood-related messages within a short time window can it be reliably inferred that the city is indeed experiencing a flood event. Therefore, by integrating rainstorm event information, the detection of urban flood occurrences can be achieved with greater accuracy.

After capturing rainstorm events at the city scale and assigning all rainstorm- and flood-related Weibo posts to their respective cities, it becomes possible to achieve spatiotemporal matching between rainstorm events and city-level Weibo data. This study further conducts flood event outbreak detection based on the occurrence periods of each rainstorm event and the cities affected. To this end, the Latent Dirichlet Allocation (LDA) topic model is employed to perform topic detection on Weibo texts posted during the rainstorm periods in each city, analyzing whether the public posts describe flood events. If the topics detected by LDA include flood-related keywords, the city is considered to have experienced a flood outbreak during the rainstorm event. Furthermore, if at least one city covered by a rainstorm event is determined to have a flood outbreak, the rainstorm event itself is regarded as having triggered a flood event. LDA is currently the most widely used probabilistic topic modeling algorithm[52], capable of automatically identifying, organizing, and extracting latent topics and their descriptive words (topic keywords) from large-scale text data. LDA assumes that documents in the corpus are generated from a mixture of latent topics, each of which is represented as a probability distribution over a set of words. To

infer the topic distribution for each document and the word distribution for each topic, LDA primarily relies on the following three key formulas:

(1) The probability distribution formula of the topic-word matrix, which represents the joint probability of a word w and a topic z. The core idea is to first select a topic based on the topic distribution $P(z|\alpha)$, and then generate a word according to the word distribution under that topic $P(w|z,\beta)$:

$$P(w,z|\alpha,\beta) = P(w|z,\beta) \times P(z|\alpha) \qquad (4)$$

(2) The prior probability calculation formulas for the parameters in the Dirichlet distribution. Equations (5) and (6) represent, respectively, the probability that document m belongs to topic k, and the probability that topic k contains the feature word t:

$$\theta_{m,k} = \frac{n_m^{(k)}+\alpha}{\sum_{k=1}^{K}\left(n_m^{(k)}+\alpha\right)} \qquad (5)$$

$$\phi_{m,k} = \frac{n_k^{(t)}+\beta}{\sum_{t=1}^{|V|}\left(n_k^{(t)}+\beta\right)} \qquad (6)$$

(3) The probability distribution calculation formula derived from Gibbs sampling, which is the core of the Gibbs sampling algorithm and is used to iteratively update the topic assignment of each word:

$$P\left(z_j = k \middle| w, z_\varepsilon\right) = \frac{n_{m\varepsilon}^{(k)}+\alpha_k}{\sum_{k=1}^{K}\left(n_{m\varepsilon}^{(k)}+\alpha_k\right)} \times \frac{n_{k\varepsilon}^{(t)}+\beta_t}{\sum_{t=1}^{|V|}\left(n_{k\varepsilon}^{(t)}+\beta_t\right)} \qquad (7)$$

In the formula, w and z represent a word and a topic, respectively; $\theta_{m,k}$ denotes the probability of topic k in document m; $\phi_{m,k}$ represents the probability of feature word t in topic k; α and β are the hyperparameters of the Dirichlet distributions for $\theta_{m,k}$ and $\phi_{m,k}$, respectively; $n_m^{(k)}$ is the number of feature words assigned to topic k in document m; $n_k^{(t)}$ denotes the number of occurrences of feature word t in topic k; $|V|$ represents the size of the vocabulary (the total number of distinct words in the corpus); K is the total number of topics; $n_{m\varepsilon}^{(k)}$ is the number of words in document m not assigned to topic k; and $n_{k\varepsilon}^{(t)}$ is the number of occurrences of feature word t not assigned to topic k.

After inputting all rainstorm- and flood-related Weibo posts during each city's rainstorm event period as the corpus into the LDA model, LDA outputs the topic distribution of the text dataset and the word distribution for each topic. In related studies, it is common to represent each topic by selecting the top mmm keywords with the highest probabilities[52,**Error! Reference source not found.**] (Xu et al., 2020; Paul et al., 2025). In this study, a single topic (k = 1) was set for the Weibo dataset during each city's rainstorm period, and the top 10 Chinese words with the highest probabilities were selected as topic keywords. To identify flood-related information from these 10 keywords, four Chinese terms—"洪" (flood), "涝" (waterlogging), "淹" (submerge), and "积水" (water accumulation)—were chosen as characteristic discriminators of flood event outbreaks based on Chinese language usage. If any of these four words appeared among the 10 topic keywords detected by LDA, the city was considered to have experienced a flood disaster during the rainstorm event. Furthermore, considering that some Chinese words containing "洪" or "涝" may actually refer to place names (e.g., "Honghu City (洪湖市)," "Jinghong City (景洪市)," "Shehong City (射洪市)," "Hongya County (洪雅县)") or disaster prevention terms (e.g., "flood control (防洪)," "waterlogging prevention (防涝)"), which do not indicate actual flood occurrences, this study excluded such terms during the discrimination process to reduce false positives in flood event detection.

4.4 Spatiotemporal Tracking of Flood Footprints

Since the trajectory of a rainstorm may pass through multiple cities, a single rainstorm event can cover several cities. Therefore, flood disasters occurring in multiple cities triggered by the same rainstorm event can be regarded as belonging to the same flood event. As shown in Supplementary Figure 4, to track the

spatiotemporal footprints of floods and analyze the frequency and evolution characteristics of flood disasters in depth, this study classifies flood events into three types: localized floods, simultaneous multi-city floods, and cascading wave floods.

(1) Localized floods refer to flood disasters that occur in only a single city during a rainstorm event, regardless of how many cities the rainstorm covers (Supplementary Figure 4(a)). The disaster impact is confined to that one city.

(2) Simultaneous multi-city floods occur when multiple cities experience flood disasters on the same day during the same rainstorm event (Two or more cities experience the outbreak of flooding on the same day), exhibiting multi-point simultaneous characteristics (Supplementary Figure 4(b)).

(3) Cascading wave floods refer to flood disasters that occur sequentially across multiple cities within the time span of a single rainstorm event, showing clear spatiotemporal propagation (Supplementary Figure 4(c)). That is, after the initial day with affected city(ies), additional cities experience floods over the following days.

Since simultaneous multi-city and cascading wave flood events cover multiple cities, their impact scope is larger, the likelihood of large-scale flooding is higher, and they are more easily detected by satellite observations as well as gaining more attention from the public and government authorities.

Acknowledgments: This research was funded by the National Natural Science Foundation of China (grant number 42077438) and the Fundamental Research Funds for the Central Universities (grant number CCNU25JC004, CCNU25JCPT028).

Author contributions

H. Gu: Conceptualization, Writing–original draft, Investigation & Formal analysis. J. Xiao: Conceptualization, Writing–review & editing & Supervision. C. Zhang, S. Xiao, Z. Niu, & F. Yu: Resources & Writing–review &editing. D. Shen: Conceptualization, Writing–review & editing, Supervision, Project administration & Funding acquisition.

Competing interests

The authors declare no competing interests.

Data availability

The datasets generated during the current study are available from the figshare repository (https://doi.org/10.6084/m9.figshare.29561732.v3). GSMaP precipitation data used in the current study are available via the website (http://sharaku.eorc.jaxa.jp/GSMaP/index.htm). Sina Weibo data used in the current study are available via the website (https://www.weibo.com).

Code availability

The code used for the current study is available from the figshare repository (https://doi.org/10.6084/m9.figshare.29561918.v4).

References：

1. Haslinger, K., Breinl, K., Pavlin, L., Pistotnik, G., Bertola, M., Olefs, M., Greilinger, M., Schöner, W.,

Blöschl, G., 2025. Increasing hourly heavy rainfall in Austria reflected in flood changes. Nature 639, 667-672. https://doi.org/10.1038/s41586-025-08647-2

2. Van Dijk, A.I.J.M., Beck H.E., Boergens E., de Jeu R.A.M., Dorigo W.A., Edirisinghe C., Forootan E., Guo E., Güntner A., Hou J., Mehrnegar N., Mo S., Preimesberger W., Rahman J., Rozas Larraondo P., 2025. Global Water Monitor 2024, Summary Report, 2025. Published by Global Water Monitor Consortium, available at www.globalwater.online.
3. Jongman, B., Ward, P.J., and Aerts, J.C.J.H., 2012. Global exposure to river and coastal flooding: Long term trends and changes. Global Environmental Change, 22, 823-835.
4. Gao, S., Huang, Y., Zhang, S., Han, J., Wang, G., Zhang, M., Lin, Q., 2020. Short-term runoff prediction with GRU and LSTM networks without requiring time step optimization during sample generation. Journal of Hydrology 589, 125188. https://doi.org/10.1016/j.jhydrol.2020.125188
5. Kellermann, P., Schröter, K., Thieken, A., Haubrock, S.-N., Kreibich, H., 2020. The object-specific flood damage database HOWAS21. https://doi.org/10.5194/nhess-2019-420
6. Wang, Y., Yang, S., Zhang, L., Cao, Y., Yin, Y., 2022. Comparative analysis and outlook of three global databases for meteorological disasters. Climate Change Research 18, 253-260. https://doi.org/10.12006/j.issn.1673-1719.2021.067
7. Moriyama, K., Sasaki, D., Ono, Y., 2018. Comparison of Global Databases for Disaster Loss and Damage Data. Journal of Disaster Research 13, 1007-1014. https://doi.org/10.20965/jdr.2018.p1007
8. Shen, G., Hwang, S.N., 2019. Spatial–Temporal snapshots of global natural disaster impacts Revealed from EM-DAT for 1900-2015. Geomatics, Natural Hazards and Risk 10, 912-934. https://doi.org/10.1080/19475705.2018.1552630
9. Tellman, B., Sullivan, J., Kuhn, C., Kettner, A., Doyle, C., Brakenridge, G., Erickson, T., Slayback, D., 2021. Satellite imaging reveals increased proportion of population exposed to floods. Nature 596, 80-86. https://doi.org/10.1038/s41586-021-03695-w
10. Linderson, S., Raffetti, E., Rusca, M., Brandimarte, L., Mard, J., Di Baldassarre, G., 2023. The wider the gap between rich and poor the higher the flood mortality. Nature Sustainability, 6, 995-1005
11. De Groeve, T., 2007. Global Disaster Alert and Coordination System: More Effective and Efficient Humanitarian Response. Hydrographic Institute of the Republic of Croatia, Split (Croatia).
12. Kuenzer, C., Guo, H., Huth, J., Leinenkugel, P., Li, X., Dech, S., 2013. Flood Mapping and Flood Dynamics of the Mekong Delta: ENVISAT-ASAR-WSM Based Time Series Analyses. Remote Sensing 5, 687-715. https://doi.org/10.3390/rs5020687
13. Tong, X., Luo, X., Liu, S., Xie, H., Chao, W., Liu, S., Liu, S., Makhinov, A., Makhinova, A., Jiang, Y., 2018. An approach for flood monitoring by the combined use of Landsat 8 optical imagery and COSMO-SkyMed radar imagery. ISPRS Journal of Photogrammetry and Remote Sensing 136, 144-153. https://doi.org/10.1016/j.isprsjprs.2017.11.006
14. Klemas, V., 2015. Remote Sensing of Floods and Flood-Prone Areas: An Overview. Journal of Coastal Research 31, 1005-1013. https://doi.org/10.2112/JCOASTRES-D-14-00160.1
15. Yan, K., Di Baldassarre, G., Solomatine, D., Schumann, G., 2015. A review of low-cost space-born data for flood modelling: topography, flood extent and water level. Hydrological Processes 29. https://doi.org/10.1002/hyp.10449
16. Sanyal, J., Lu, X.X., 2004. Application of Remote Sensing in Flood Management with Special Reference to Monsoon Asia: A Review. Natural Hazards 33, 283-301.

https://doi.org/10.1023/B:NHAZ.0000037035.65105.95

17. Schumann, G., Bates, P., Horritt, M., Matgen, P., Pappenberger, F., 2009. Progress in integration of remote sensing-derived flood extent and stage and hydraulic models. Reviews of Geophysics 47. https://doi.org/10.1029/2008RG000274
18. Brodzik, M., Long, D., Hardman, M., 2018. Best Practices in Crafting the Calibrated, Enhanced-Resolution Passive-Microwave EASE-Grid 2.0 Brightness Temperature Earth System Data Record. Remote Sensing 10, 1793. https://doi.org/10.3390/rs10111793
19. Zeng, Z., Gan, Y., Kettner, A., Yang, Q., Zeng, C., Brakenridge, R., Hong, Y., 2019. Towards high resolution flood monitoring: An integrated methodology using passive microwave brightness temperatures and Sentinel synthetic aperture radar imagery. Journal of Hydrology 582, 124377. https://doi.org/10.1016/j.jhydrol.2019.124377
20. Misra, A., White, K., Nsutezo, S., Straka III, W., Lavista Ferres, J., 2025. Mapping global floods with 10 years of satellite radar data. Nature Communications 16. https://doi.org/10.1038/s41467-025-60973-1
21. Matosak, B., Gella, G., Lang, S., 2025. SenForFlood: A New Global Dataset for Flooded Area Detection. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLVIII-M-7-2025, 97-102. https://doi.org/10.5194/isprs-archives-XLVIII-M-7-2025-97-2025
22. Liu, Y., Liu, X., Gao, S., Gong, L., Kang, C., Zhi, Y., Chi, G., Shi, L., 2015. Social Sensing: A New Approach to Understanding Our Socioeconomic Environments. Annals of the Association of American Geographers 105, 1-19. https://doi.org/10.1080/00045608.2015.1018773
23. Li, J., he, Z., Plaza, J., Li, S., Jinfen, C., Wu, H., Wang, Y., Liu, Y., 2017. Social Media: New Perspectives to Improve Remote Sensing for Emergency Response. Proceedings of the IEEE PP, 1-13. https://doi.org/10.1109/JPROC.2017.2684460
24. Zhao, Q., Chen, Z., Liu, C., Luo, N., 2018. Extracting and classifying typhoon disaster information based on volunteered geographic information from Chinese Sina microblog. Concurrency and Computation: Practice and Experience 31, e4910. https://doi.org/10.1002/cpe.4910
25. Chae, J., Thom, D., Jang, Y., Kim, S., Ertl, T., Ebert, D., 2014. Public Behavior Response Analysis in Disaster Events Utilizing Visual Analytics of Microblog Data. Computers & Graphics 38, 51-60. https://doi.org/10.1016/j.cag.2013.10.008
26. Hao, H., Wang, Y., 2020. Leveraging Multimodal Social Media Data for Rapid Disaster Damage Assessment. International Journal of Disaster Risk Reduction 51, 101760. https://doi.org/10.1016/j.ijdrr.2020.101760
27. Sadiq, R., Akhtar, Z., Imran, M., Ofli, F., 2022. Integrating remote sensing and social sensing for flood mapping. Remote Sensing Applications: Society and Environment 25, 100697. https://doi.org/10.1016/j.rsase.2022.100697
28. Saleem, H., Xu, Y., Ruths, D., 2014. Effects of Disaster Characteristics on Twitter Event Signature. Procedia Engineering 78, 165-172. https://doi.org/10.1016/j.proeng.2014.07.053
29. Fu, S., Lyu, H., Wang, Z., Hao, X., Zhang, C., 2022. Extracting historical flood locations from news media data by the named entity recognition (NER) model to assess urban flood susceptibility. Journal of Hydrology 612, 128312. https://doi.org/10.1016/j.jhydrol.2022.128312
30. Karmegam, D., Sivakumar, R., Mappillairaju, B., 2021. Near real time flood inundation mapping using social media data as an information source: a case study of 2015 Chennai flood. Geoenvironmental Disasters 8. https://doi.org/10.1186/s40677-021-00195-x

31. Li, Y., Osei, F., Hu, T., Stein, A., 2022. Urban flood susceptibility mapping based on social media data in Chengdu city, China. Sustainable Cities and Society 88, 104307. https://doi.org/10.1016/j.scs.2022.104307
32. Yang, Z., Nguyen, L., Stuve, J., Cao, G., Jin, F., 2017. Harvey flooding rescue in social media. https://doi.org/10.1109/BigData.2017.8258166
33. Scotti, V., Giannini, M., Cioffi, F., 2020. Enhanced flood mapping using synthetic aperture radar (SAR) images, hydraulic modelling, and social media: A case study of Hurricane Harvey (Houston, TX). Journal of Flood Risk Management 13. https://doi.org/10.1111/jfr3.12647
34. Shoyama, K., Cui, Q., Hanashima, M., Sano, H., Usuda, Y., 2020. Emergency flood detection using multiple information sources: Integrated analysis of natural hazard monitoring and social media data. Science of The Total Environment 767, 144371. https://doi.org/10.1016/j.scitotenv.2020.144371
35. de Bruijn, J., Moel, H., Jongman, B., de Ruiter, M., Wagemaker, J., Aerts, J., 2019. A global database of historic and real-time flood events based on social media. Scientific Data 6, 311. https://doi.org/10.1038/s41597-019-0326-9
36. Li, Y., Zhao, S., 2022. Floods losses and hazards in China from 2001 to 2020. Climate Change Research 18, 154-165. https://doi.org/10.12006/j.issn.1673-1719.2021.196
37. Shen, D., Gu, H., Chen, W., Zhang, C., Xiao, S., Zhang, S., 2025. How Many Floods Have Occurred in China in The Past Decade? A Perspective from Social Media, Earth's Future 13, e2024EF004775.
38. Wang, Z., 2023. Analysis of Extreme Precipitation Weather Process in Henan during July 17-22, 2021. Journal of Geoscience and Environment Protection 11(4), 72-85.
39. Stampoulis, D., Anagnostou, E., 2012. Evaluation of Global Satellite Rainfall Products over Continental Europe. Journal of Hydrometeorology 13, 588-603. https://doi.org/10.1175/JHM-D-11-086.1
40. Gu, H., Shen, D., Xiao, S., Zhang, C., Bai, F., Yu, F., 2024. Evaluation of Daily and Hourly Performance of Multi-Source Satellite Precipitation Products in China's Nine Water Resource Regions. Remote Sensing 16. https://doi.org/10.3390/rs16091516
41. Kong, F., Liu, F., Wang, Y., Fang, J., Lu, L., 2018. Spatial Relationship between Flooding Frequency and Climate Change based on Climate Change Regionalization in China from 1961 to 2010. Journal of Catastrophology 32, 35-42.
42. Xie, P., Yatagai, A., Chen, M., Hayasaka, T., Fukushima, Y., Liu, C., Yang, S., 2007. A Gauge-Based Analysis of Daily Precipitation over East Asia. Journal of Hydrometeorology 8(3), 607-626. https://doi.org/10.1175/JHM583.1
43. Smith, L., Liang, Q., James, P., Lin, W., 2015. Assessing the utility of social media as a data source for flood risk management using a real-time modelling framework: Assessing the utility of social media for flood risk management. Journal of Flood Risk Management 10(3), 370-380. https://doi.org/10.1111/jfr3.12154
44. Resch, B., Usländer, F., Havas, C., 2017. Combining machine-learning topic models and spatiotemporal analysis of social media data for disaster footprint and damage assessment. Cartography and Geographic Information Science 45(8), 1-15. https://doi.org/10.1080/15230406.2017.1356242
45. Huang, X., Wang, C., Li, Z., 2018. A Near Real-Time Flood Mapping Approach by Integrating Social Media and Post-event Satellite Imagery. Annals of GIS 24, 113-123. https://doi.org/10.1080/19475683.2018.1450787
46. Yu, Z., Yang, X., Yu, X., Jiang, X., 2025. Capability and biases of satellite precipitation products in monitoring extreme rainstorms along the northern slope of the Kunlun Mountains. Arid Zone Research,

42(10), 1777-1790. https://link.cnki.net/doi/10.13866/j.azr.2025.10.03

47. Pang, Z., Zhang, Y., Shi, C., Gu, J., Yang, Q., Pan, Y., Wang, Z., Xu, B., 2023. A Comprehensive Assessment of Multiple High-Resolution Precipitation Grid Products for Monitoring Heavy Rainfall during the "7.20" Extreme Rainstorm Event in China. Remote Sensing 15, 5255. https://doi.org/10.3390/rs15215255
48. Skok, G., Bacmeister, J., Tribbia, J., 2013. Analysis of Tropical Cyclone Precipitation Using an Object-Based Algorithm. Journal of Climate 26, 2563-2579. https://doi.org/10.1175/JCLI-D-12-00135.1
49. Davis, C., Brown, B., Bullock, R., 2006. Object-Based Verification of Precipitation Forecasts. Part I: Methodology and Application to Mesoscale Rain Areas. Monthly Weather Review 134, 1772-1784. https://doi.org/10.1175/MWR3145.1
50. Skok, G., Tribbia, J., Rakovec, J., Brown, B., 2009. Object-Based Analysis of Satellite-Derived Precipitation Systems over the Low and Midlatitude Pacific Ocean. Monthly Weather Review 137, 3196-3218. https://doi.org/10.1175/2009MWR2900.1
51. Devlin, J., Chang, M., Lee, K., Toutanova, K., 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. https://doi.org/10.48550/arXiv.1810.04805
52. Paul, P.C., Begum, A., Ahmed, M.T., Chakraborty, D., Rahman, M., 2025. Combining BERT with LDA: Improved Topic Modeling in Bengali Language. IAENG International Journal of Computer Science 52, 383-393.
53. Jawahar, G., Sagot, B., Seddah, D., 2019. What does BERT learn about the structure of language?. In Proceedings of the 57th annual meeting of the association for computational linguistics, pp. 3651-3657. https://doi.org/10.18653/v1/P19-1356
54. Fang, W., Luo, H., Xu, S., Lu, Z., Ye, C., 2020. Automated text classification of near-misses from safety reports: An improved deep learning approach. Advanced Engineering Informatics 44, 101060. https://doi.org/10.1016/j.aei.2020.101060
55. Gardazi, N., Daud, A., Malik, M., Bukhari, A., Alsahfi, T., Alshemaimri, B., 2025. BERT applications in natural language processing: a review. Artificial Intelligence Review 58, 166. https://doi.org/10.1007/s10462-025-11162-5
56. Xu, L., Xue, Z., Huang, H., 2020. Short text semantic feature extension and classification based on LDA. IOP Conference Series: Materials Science and Engineering 715, 012110. https://doi.org/10.1088/1757-899X/715/1/012110

Figure Captions

Figure 1: Conceptual Diagram of Social Sensing for Spatiotemporal Flood Footprint Tracking.

Figure 2: Statistics of Rainstorm- and Flood-Related Sina Weibo Posts Classified by BERT from 2012 to 2024. **a** Number of rainstorm- and flood-related Weibo posts in cities across China. **b** Annual variation line chart of rainstorm- and flood-related Weibo posts. The blue-circle line and red-rectangle line represent the total number of posts and the number of rainstorm- and flood-related posts, respectively. **c** Scatter plot of rainstorm- and flood-related Weibo posts in cities across China.

Figure 3: Statistics of Rainstorm Events in China from 2012 to 2024.
a Spatial distribution of rainstorm object counts across China. **b** Spatial distribution of rainstorm event counts at the city level. **c** Daily temporal distribution of rainstorm events in China. **d** Line chart showing daily variation in the number of rainstorm events in China.

Figure 4: Statistical Chart of Rainstorm Objects, Rainstorm Events, and Flood Events in China from 2012 to 2024.

Figure 5: Statistics of Flood Events in China from 2012 to 2024.
a Spatial distribution of flood event counts across Chinese cities. **b** Monthly trend line chart of the three types of flood events. **c** Box plots of flood event duration (blue) and number of affected cities (red). Box plots show the median (white line) and interquartile range (Q1-Q3); whiskers extend to the minimum and maximum values. **d** Scatter plots of flood event duration and number of affected cities.

Figure 6: Spatiotemporal Footprint of Flood Event No. 967 (Partial).
a Changes in rainstorm objects and flood-affected areas from June 18 to 20 at 12:00 noon. **b** Social media images of flooding from multiple affected cities. **c** Line chart showing the variation in Weibo post counts for multiple cities experiencing flood disasters.

Figure 7: Statistical Results of Flood Event Detection by Remote Sensing Satellites.
a The line chart of satellite-detected flood events. **b** The line chart of satellite-undetected flood events

Figure 8: Case study of Flood Event No. 898.
a Line chart of the number of rainstorm- and flood-related Weibo posts, along with the available satellite imagery within the specified time range. **b**–**g** show the reference water bodies, flood water bodies, and flood inundation extent extracted from visible-light and microwave imagery, respectively: **b** and **e** – reference water bodies; **c** and **f** – flood water bodies; **d** and **g** – flood inundation extent and inundation points identified from Weibo posts.

Tables

Table 1. Comparison of Flood Event Counts Between EM-DAT, DFO Databases, and Flood Events Extracted by This Study (Val = Validated Flood Events)

Year	Our	Our (Val)	EM-DAT	EM-DAT/ Our (Val)	Our (Val) /EM-DAT	DFO	DFO/ Our (Val)	Our (Val) /DFO
2012	67	59	12	8	16	9	4	7
2013	116	94	14	9	14	8	4	7
2014	84	70	13	8	16	6	4	9
2015	109	97	12	12	37	3	1	11
2016	125	101	12	11	31	4	4	23
2017	104	92	12	8	20	4	2	4
2018	129	107	8	6	42	3	2	2
2019	118	96	6	2	10	3	3	6
2020	141	113	5	3	78	3	3	8
2021	96	85	4	3	10	2	2	3
2022	86	66	4	2	4	0	\	\
2023	73	54	8	3	4	0	\	\
2024	82	60	4	2	2	0	\	\
Total	1330	1094	114	77	284	45	29	80

Table 2. Comparative Statistics of Flood Event Observations Based on Optical and Microwave Remote Sensing Imagery

Year	Our	RS	Sentinel-1	Sentinel-2/Landsat
2012	59	2	0	2
2013	94	7	0	7
2014	70	4	0	4
2015	97	15	14	4
2016	101	60	51	25
2017	92	60	56	22
2018	107	79	78	29
2019	96	81	80	41
2020	113	99	97	35
2021	85	66	63	40
2022	66	20	19	11
2023	54	36	34	22
2024	60	16	16	6
Total	1094	545	508	248

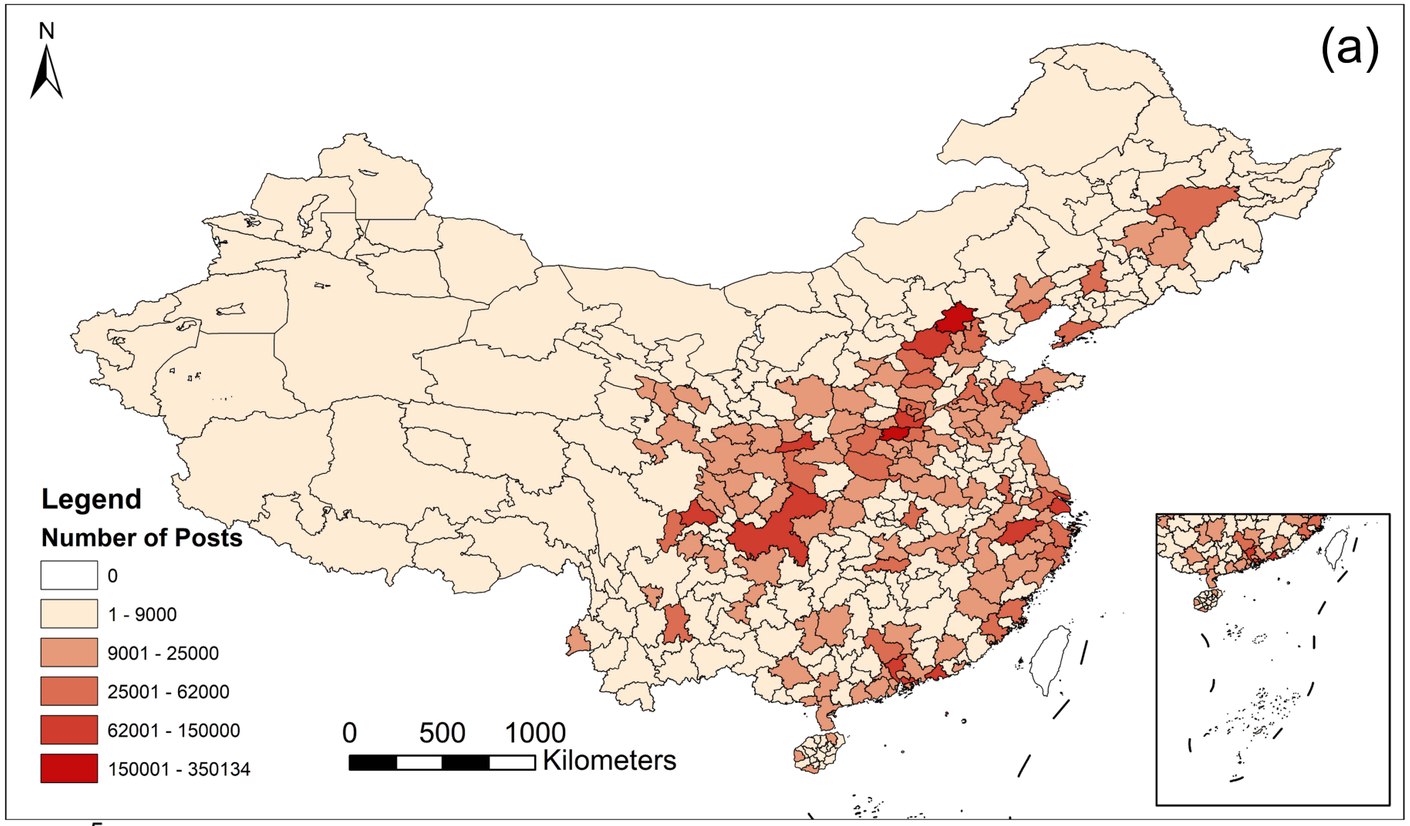
Editorial summary:

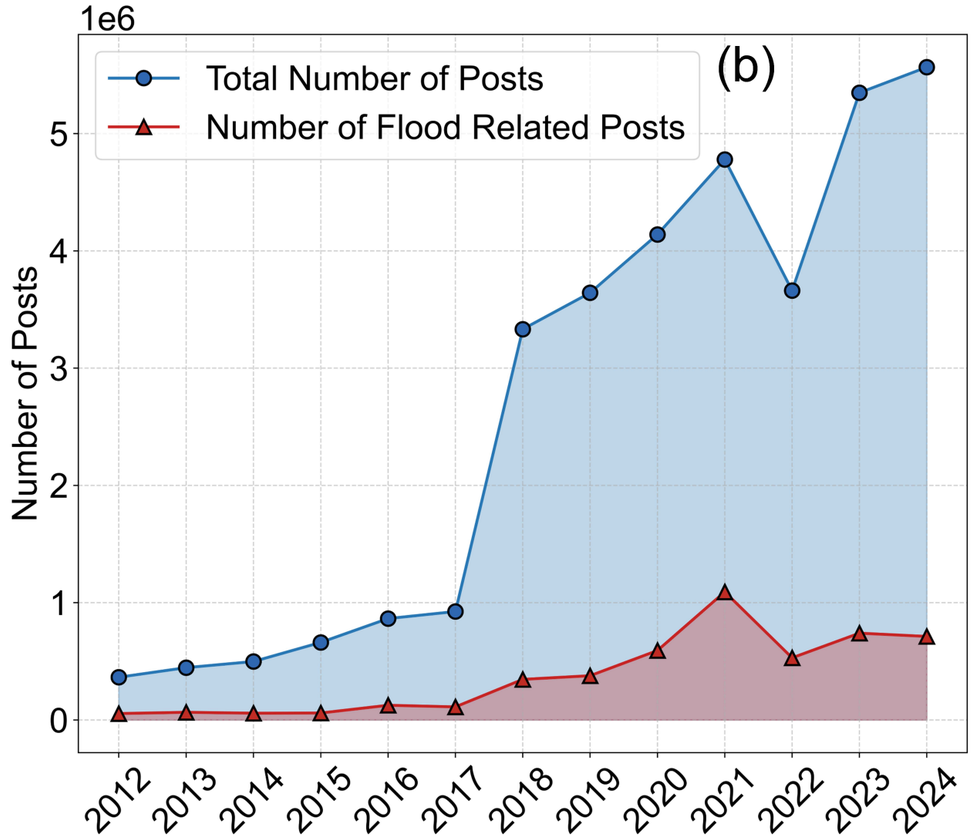
In China, about half of the flood events were unobservable by remote sensing, capturing mostly riverine flooding in open areas, while social media fills gaps in densely populated urban settings, according to a social sensing framework that combines satellite precipitation mapping and Sina Weibo posts.

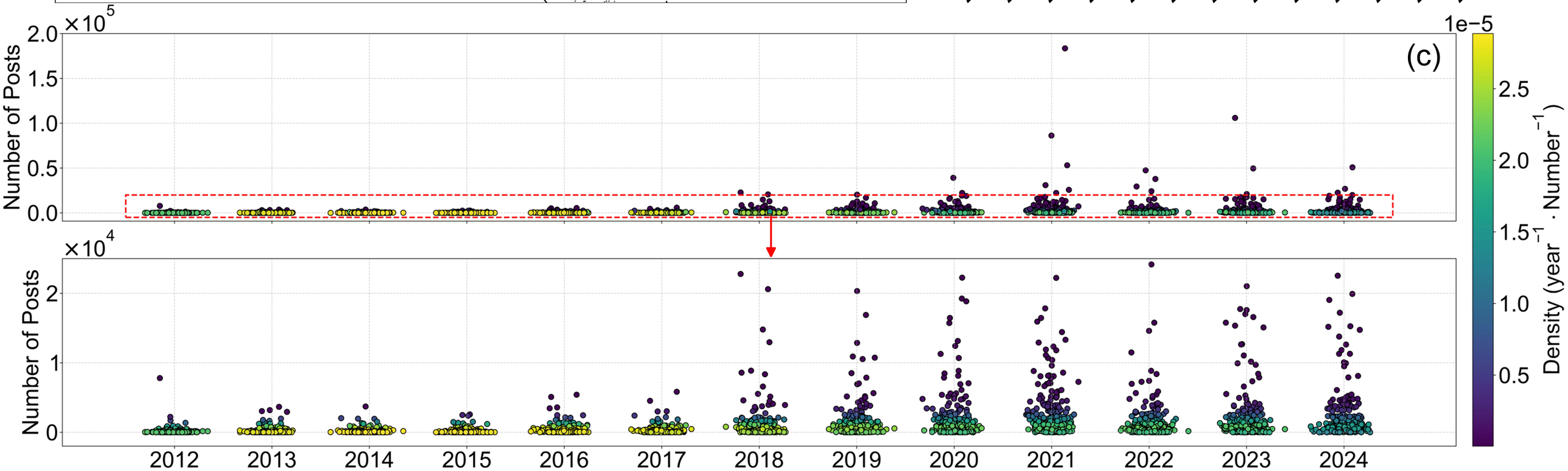
Peer review information:

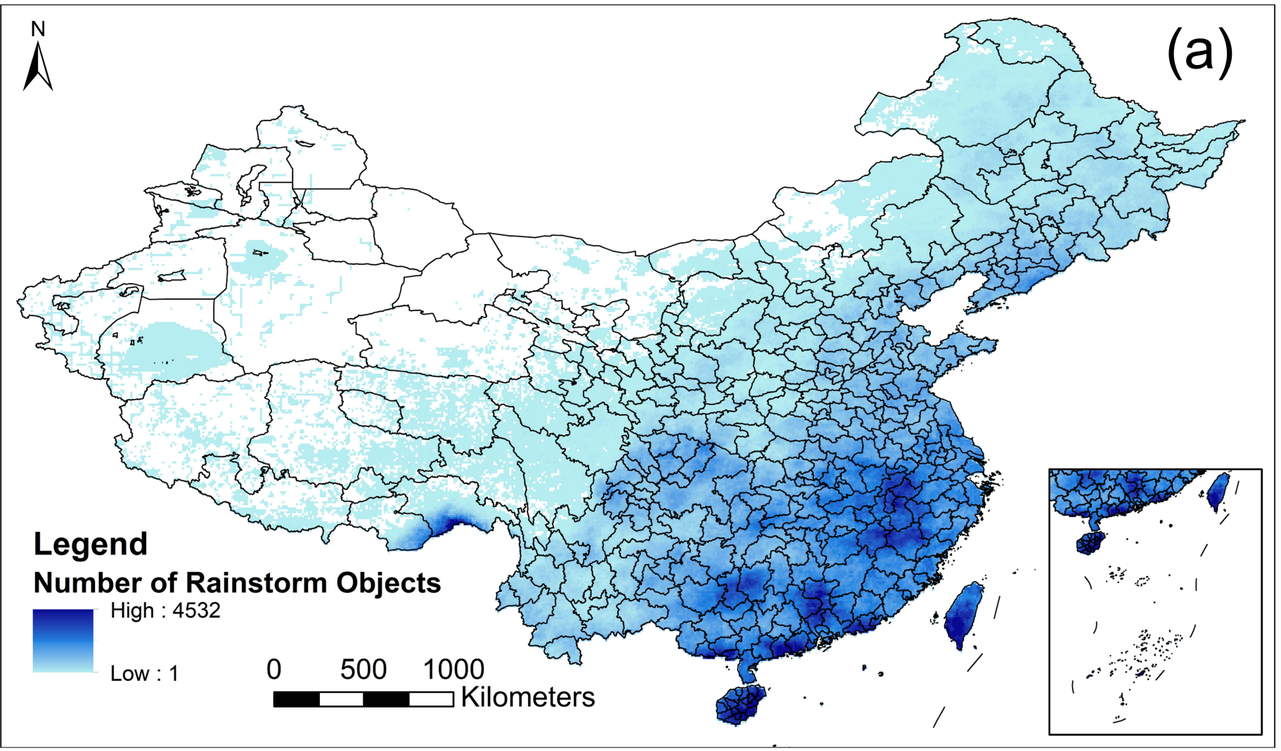
Communications Earth and Environment thanks Chenghan Yang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: Rajarshi Das Bhowmik and Martina Grecequet. A peer review file is available.

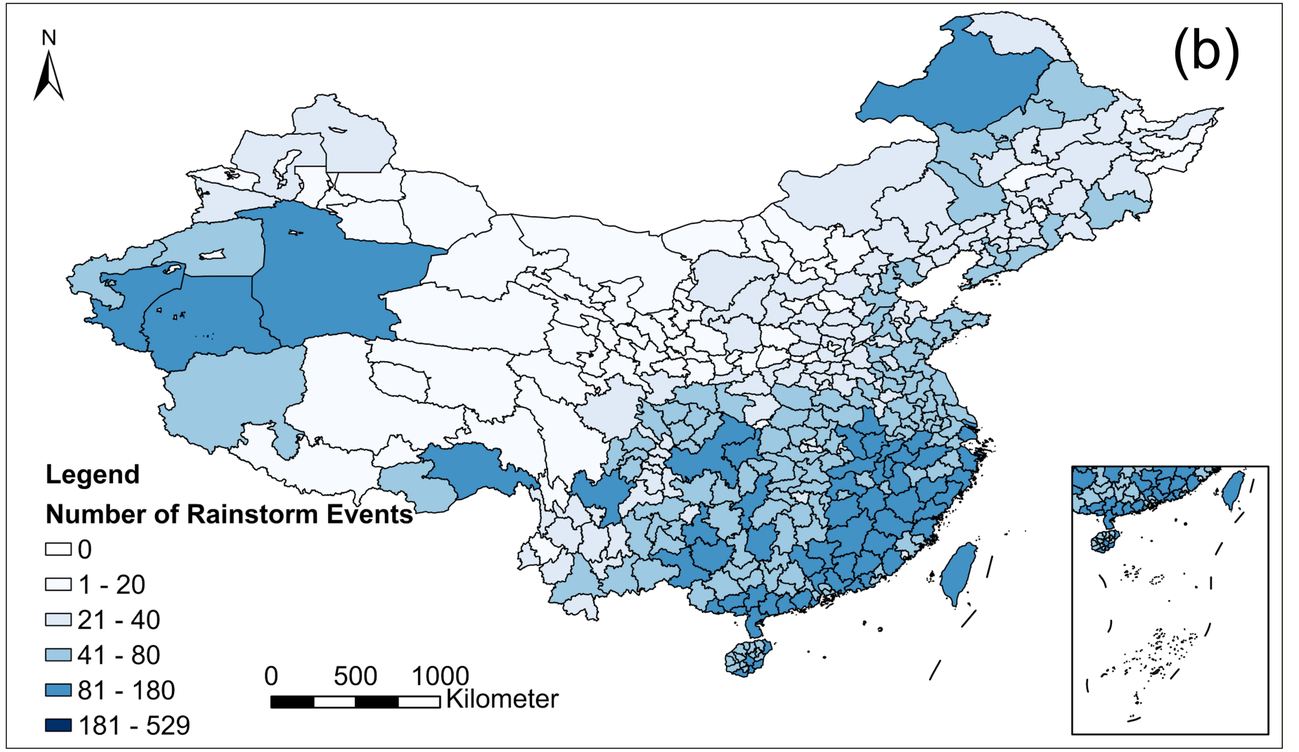

6,018 rainstorm events
#urban waterlogging#
content: waist-deep water flooded the streets, turning them into a vast sea.
#torrential downpour#
content: multiple vehicles were washed away, and residences were flooded by backflow.
90+ million Weibo posts
#clear skies after the rain#
content: bright sunshine and balmy weather.
#Flash flood#
content: a flash flood triggered by torrential rain destroyed sections of the roadway.
#flood emergency response#
content: two people were swept away by floodwaters and remain missing.
1094 flood events

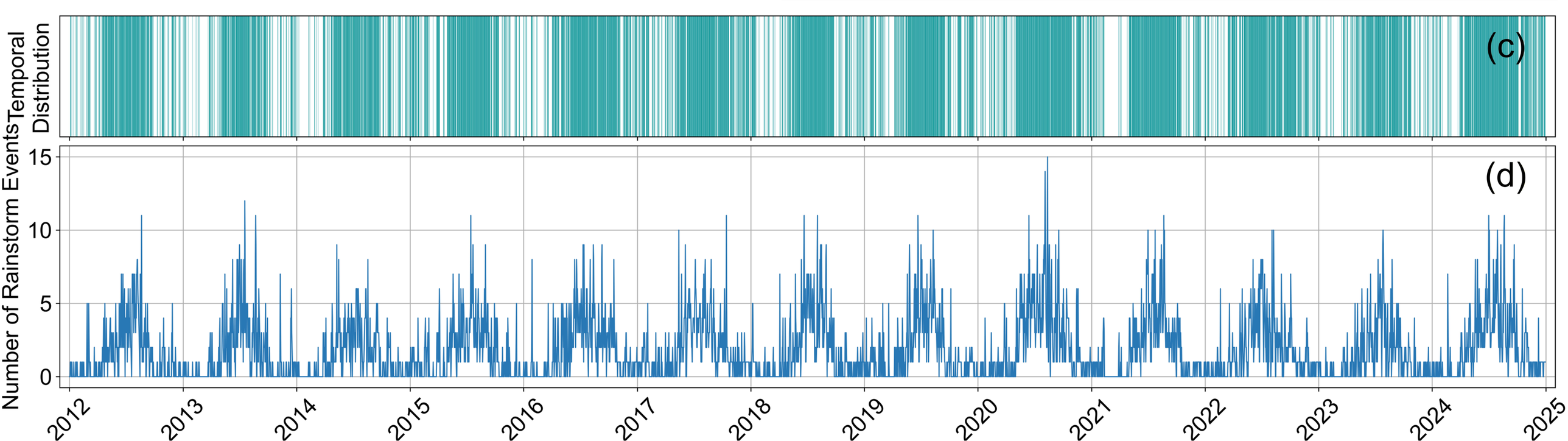

(a)
N
Legend
Number of Posts
0
1 - 9000
9001 - 25000
25001 - 62000
62001 - 150000
150001 - 350134
0 500 1000
Kilometers

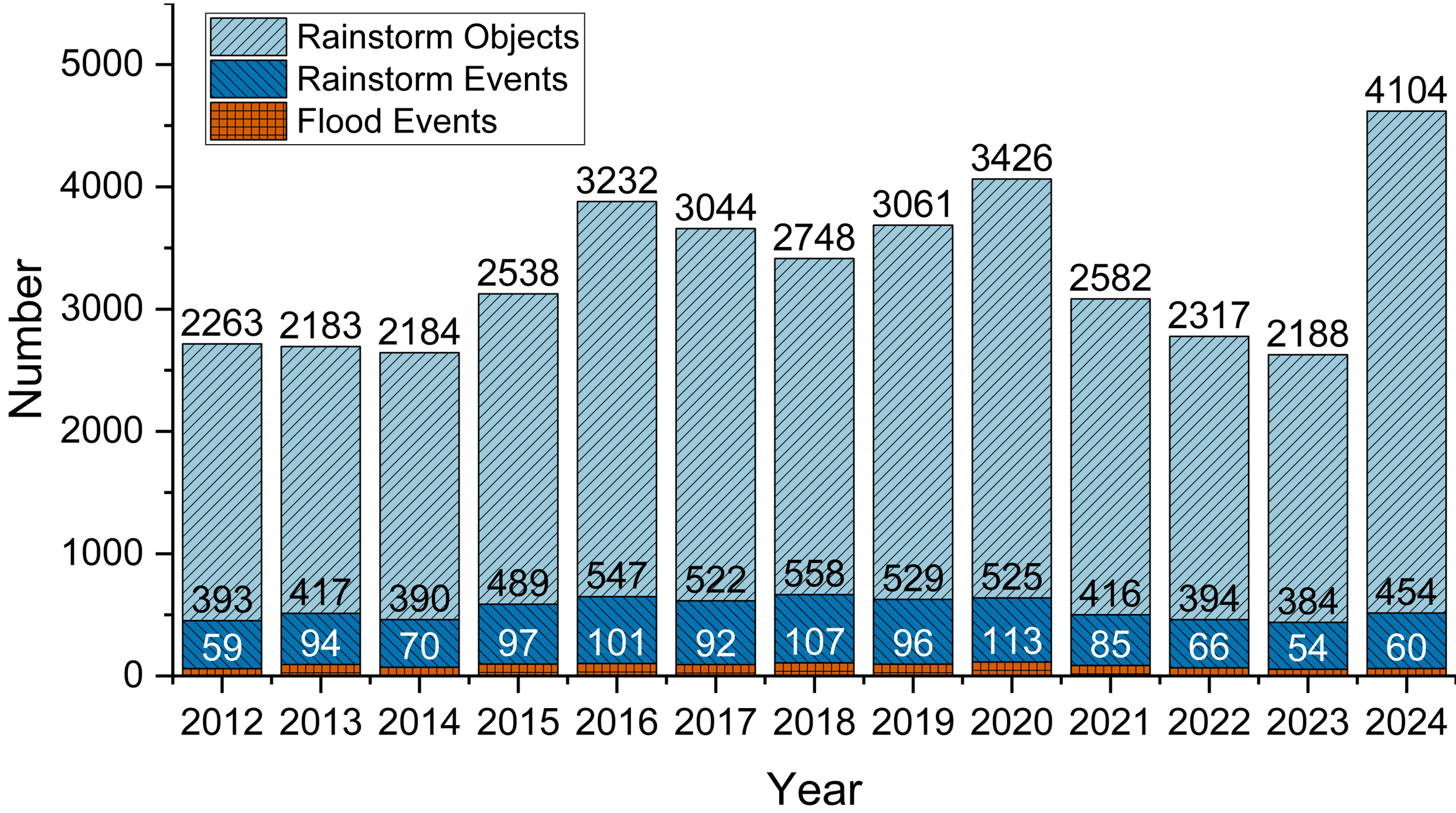

(b)
1e6
Total Number of Posts
Number of Flood Related Posts
Number of Posts
0
1
2
3
4
5
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024

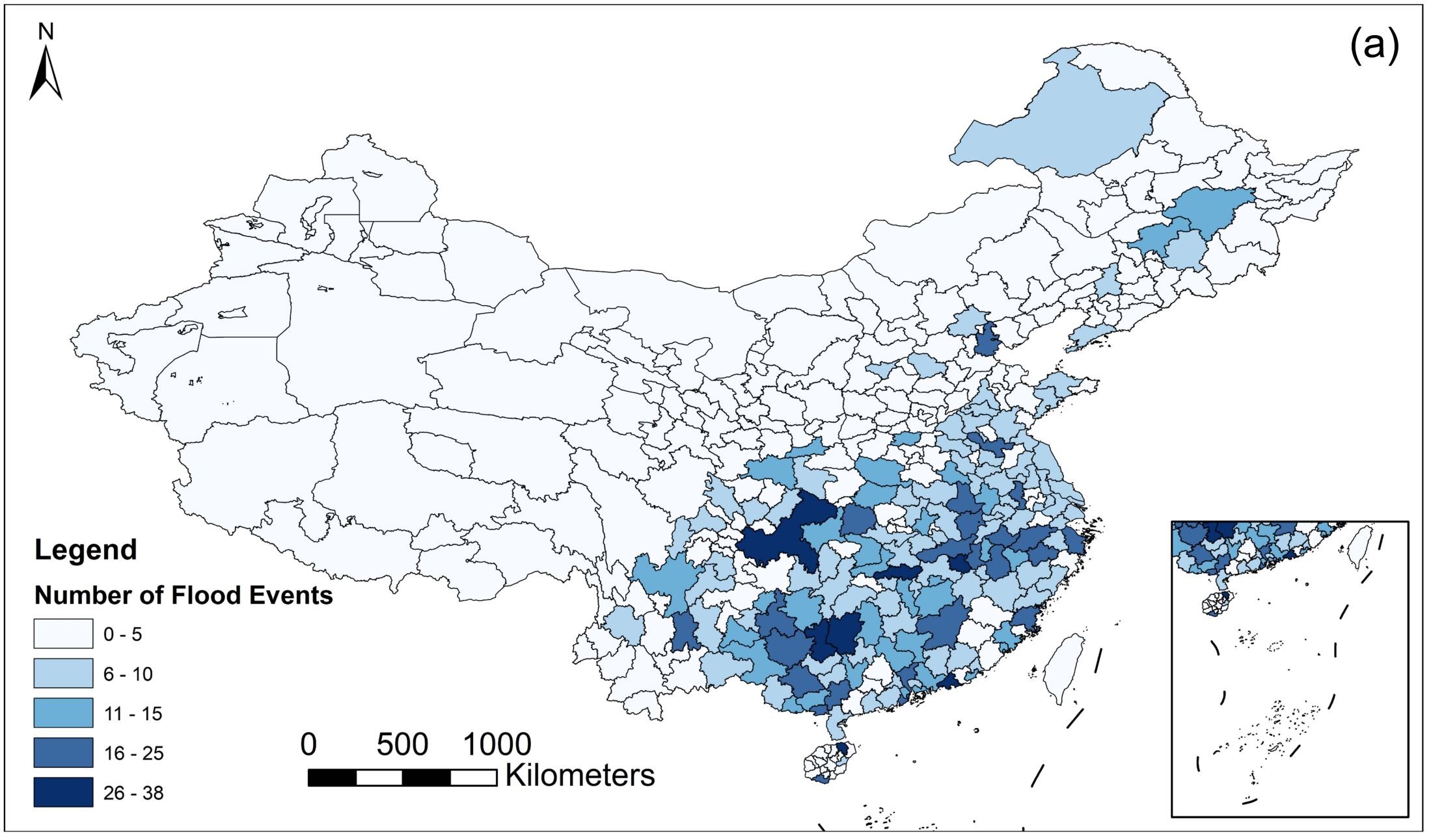

(c)
×10^5
Number of Posts
0.0
0.5
1.0
1.5
2.0
×10^4
Number of Posts
0
1
2
1e−5
Density (year^−1 · Number^−1)
0.5
1.0
1.5
2.0
2.5
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024

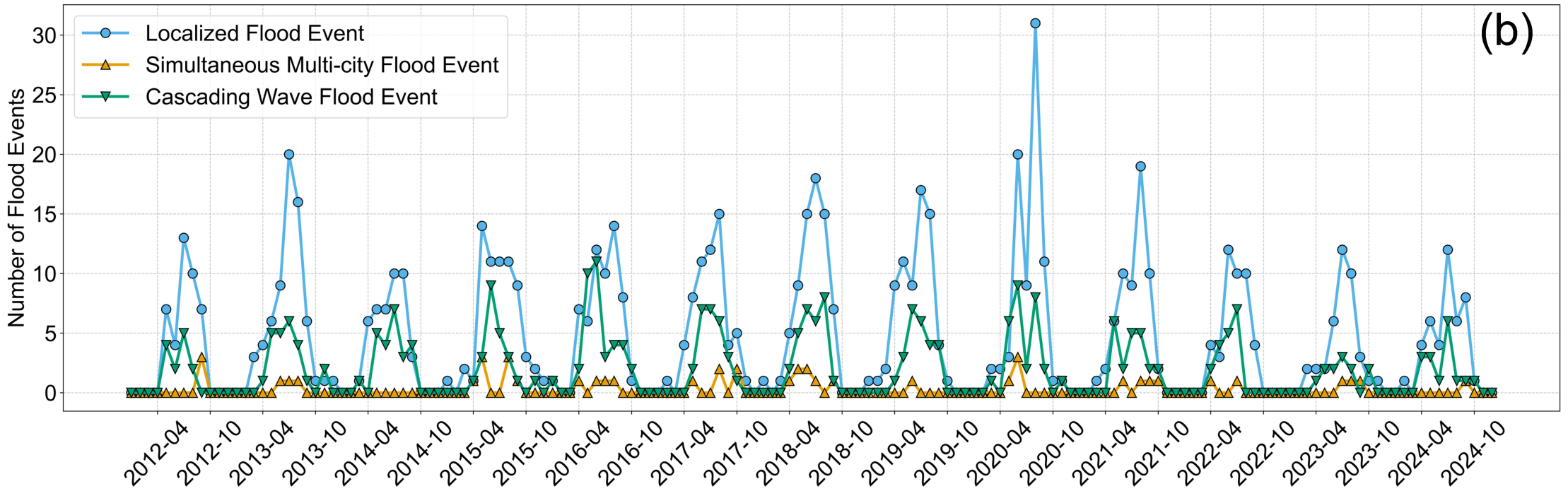

N
(a)
Legend
Number of Rainstorm Objects
High : 4532
Low : 1
0 500 1000
Kilometers

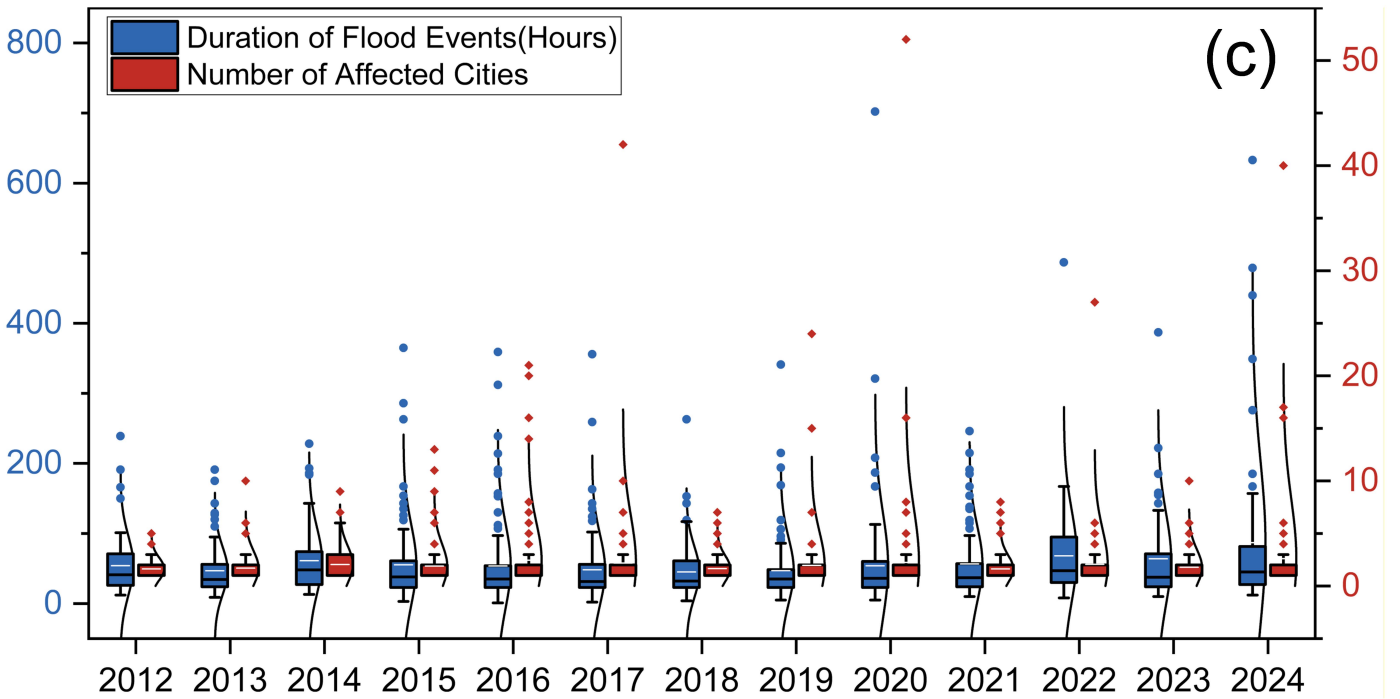

N
(b)
Legend
Number of Rainstorm Events
0
1 - 20
21 - 40
41 - 80
81 - 180
181 - 529
0 500 1000
Kilometer

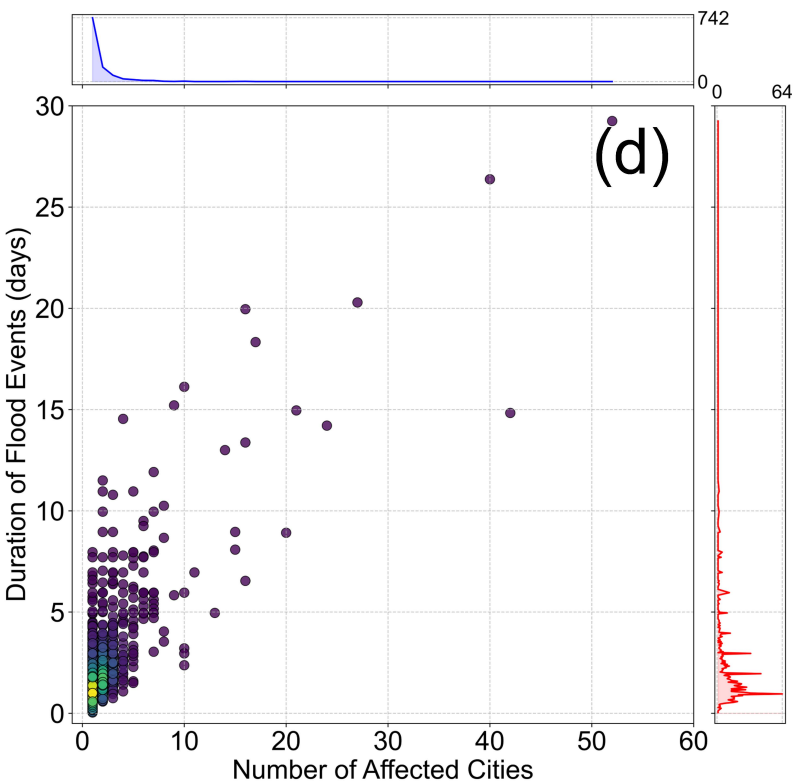

Temporal Distribution
(c)
Number of Rainstorm Events
(d)
0
5
10
15
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025


Rainstorm Objects
Rainstorm Events
Flood Events
Number
Year
0
1000
2000
3000
4000
5000
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2263
2183
2184
2538
3232
3044
2748
3061
3426
2582
2317
2188
4104
393
417
390
489
547
522
558
529
525
416
394
384
454
59
94
70
97
101
92
107
96
113
85
66
54
60

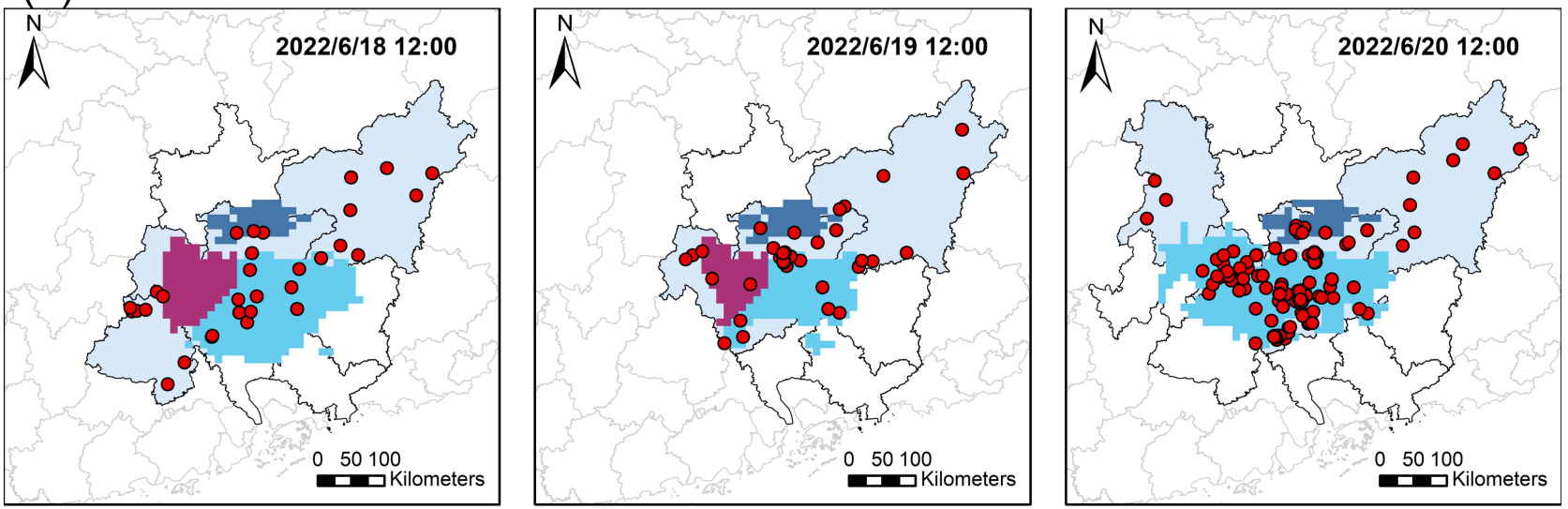

N
(a)
Legend
Number of Flood Events
0 - 5
6 - 10
11 - 15
16 - 25
26 - 38
0 500 1000
Kilometers


(b)
Localized Flood Event
Simultaneous Multi-city Flood Event
Cascading Wave Flood Event
Number of Flood Events
0
5
10
15
20
25
30
2012-04
2012-10
2013-04
2013-10
2014-04
2014-10
2015-04
2015-10
2016-04
2016-10
2017-04
2017-10
2018-04
2018-10
2019-04
2019-10
2020-04
2020-10
2021-04
2021-10
2022-04
2022-10
2023-04
2023-10
2024-04
2024-10


(c)
Duration of Flood Events(Hours)
Number of Affected Cities
0
200
400
600
800
0
10
20
30
40
50
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024


(d)
742
0
0
64
Duration of Flood Events (days)
Number of Affected Cities
0
5
10
15
20
25
30
0
10
20
30
40
50
60

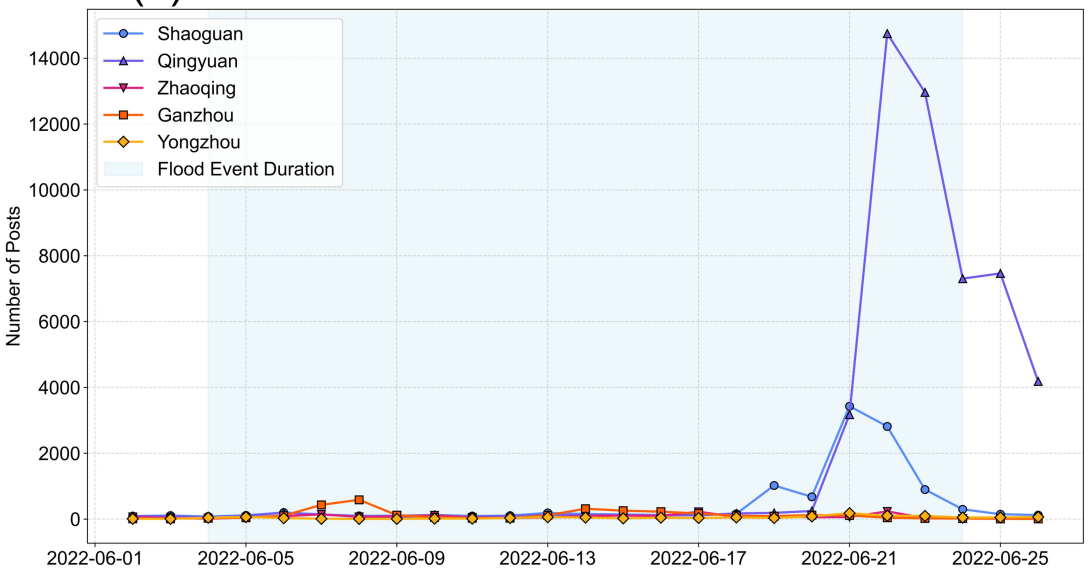
(a)

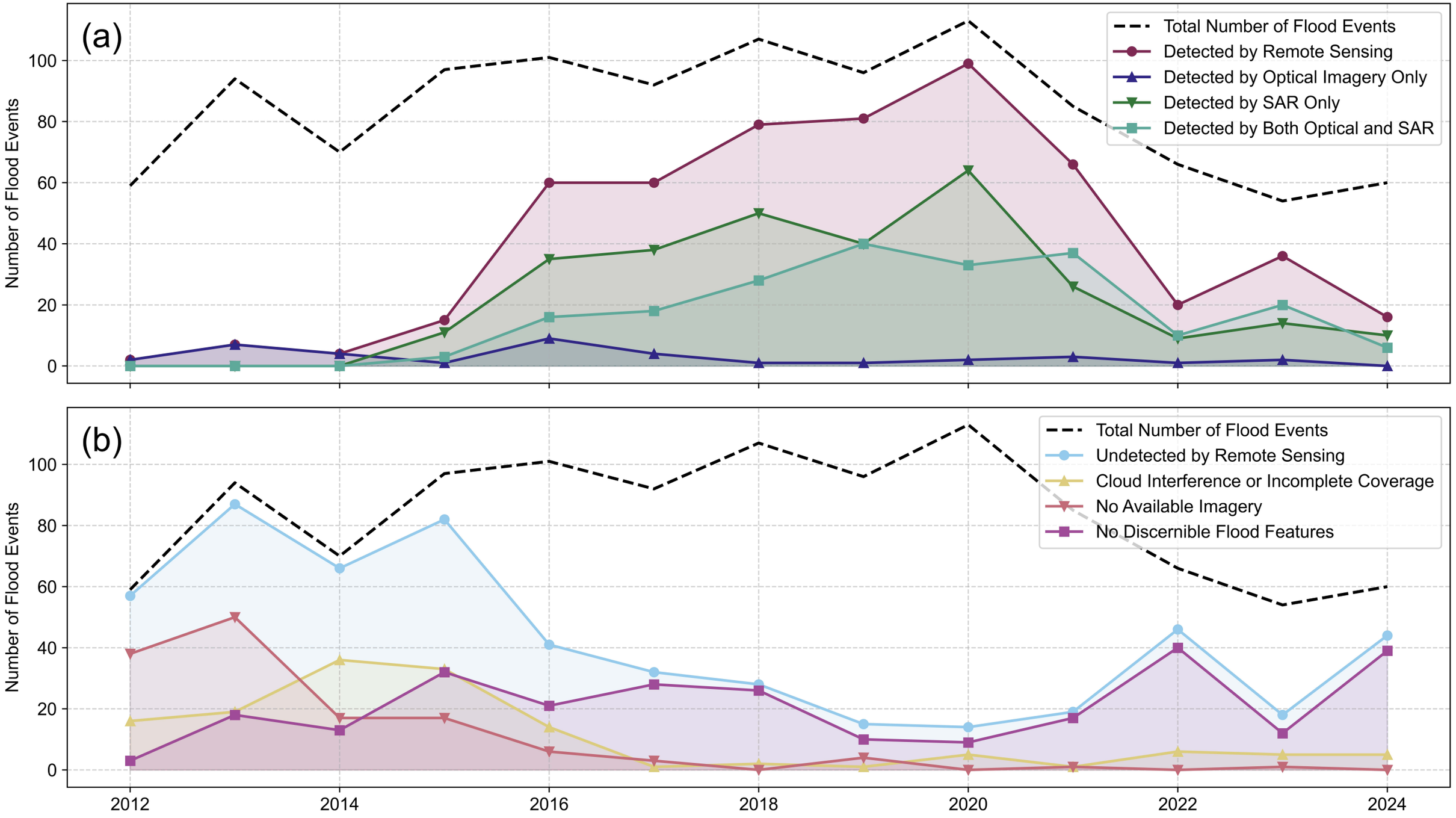


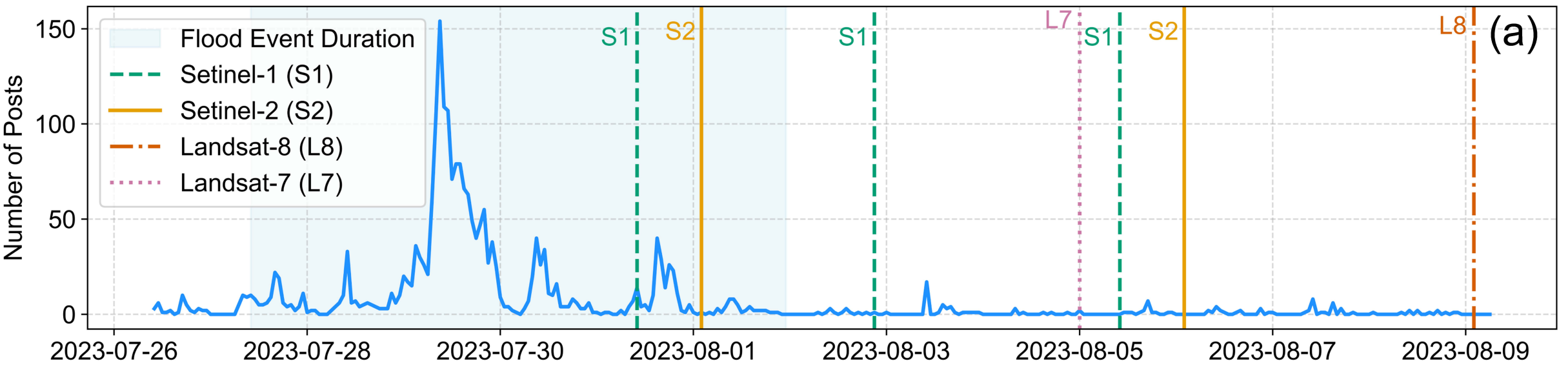
Location of Flood Post | Rainstorm Object 1 | Rainstorm Object 2 | Rainstorm Object 3 | Flood-affected Cities | Rainstorm-affected Cities

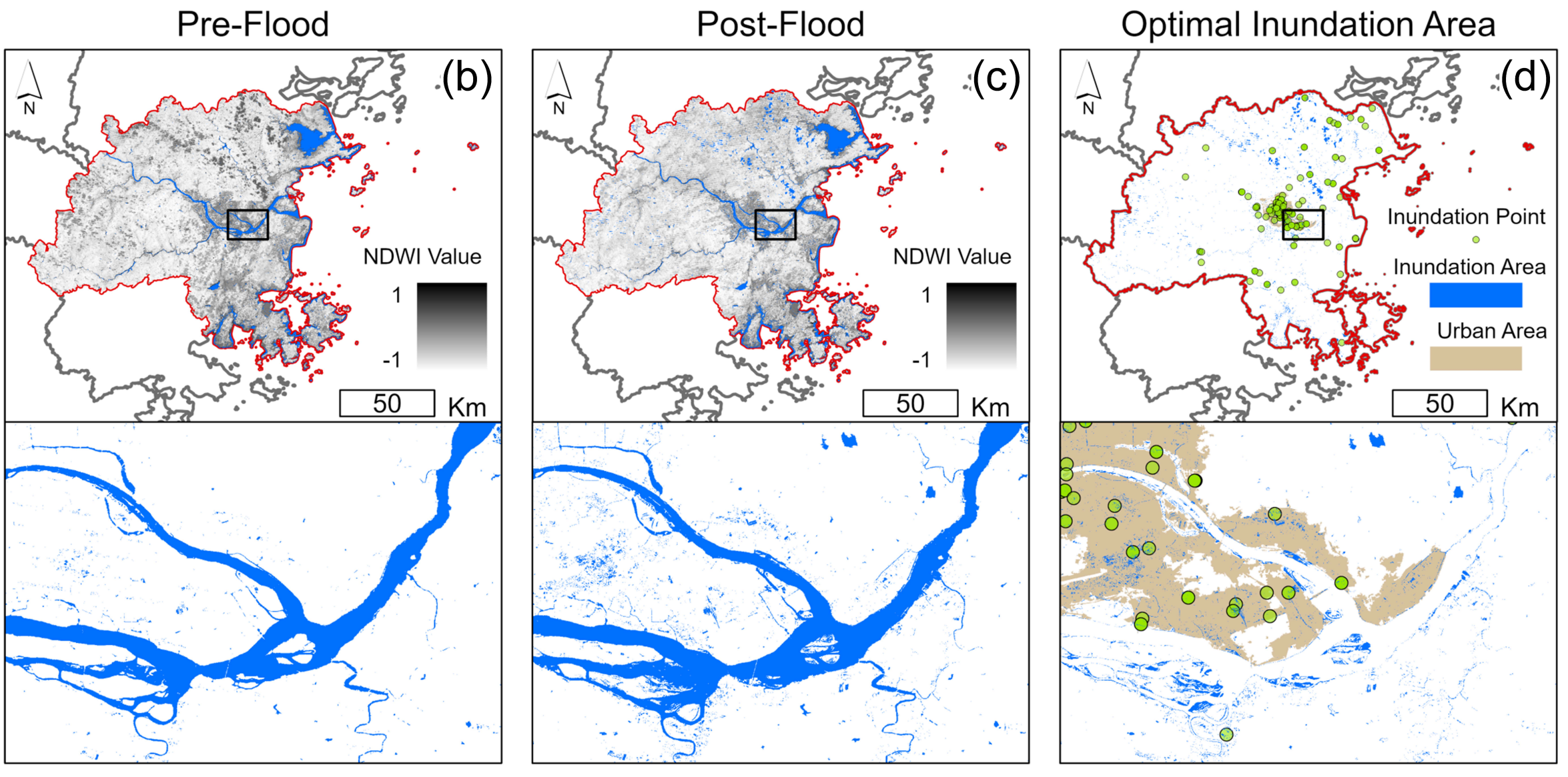
(b)

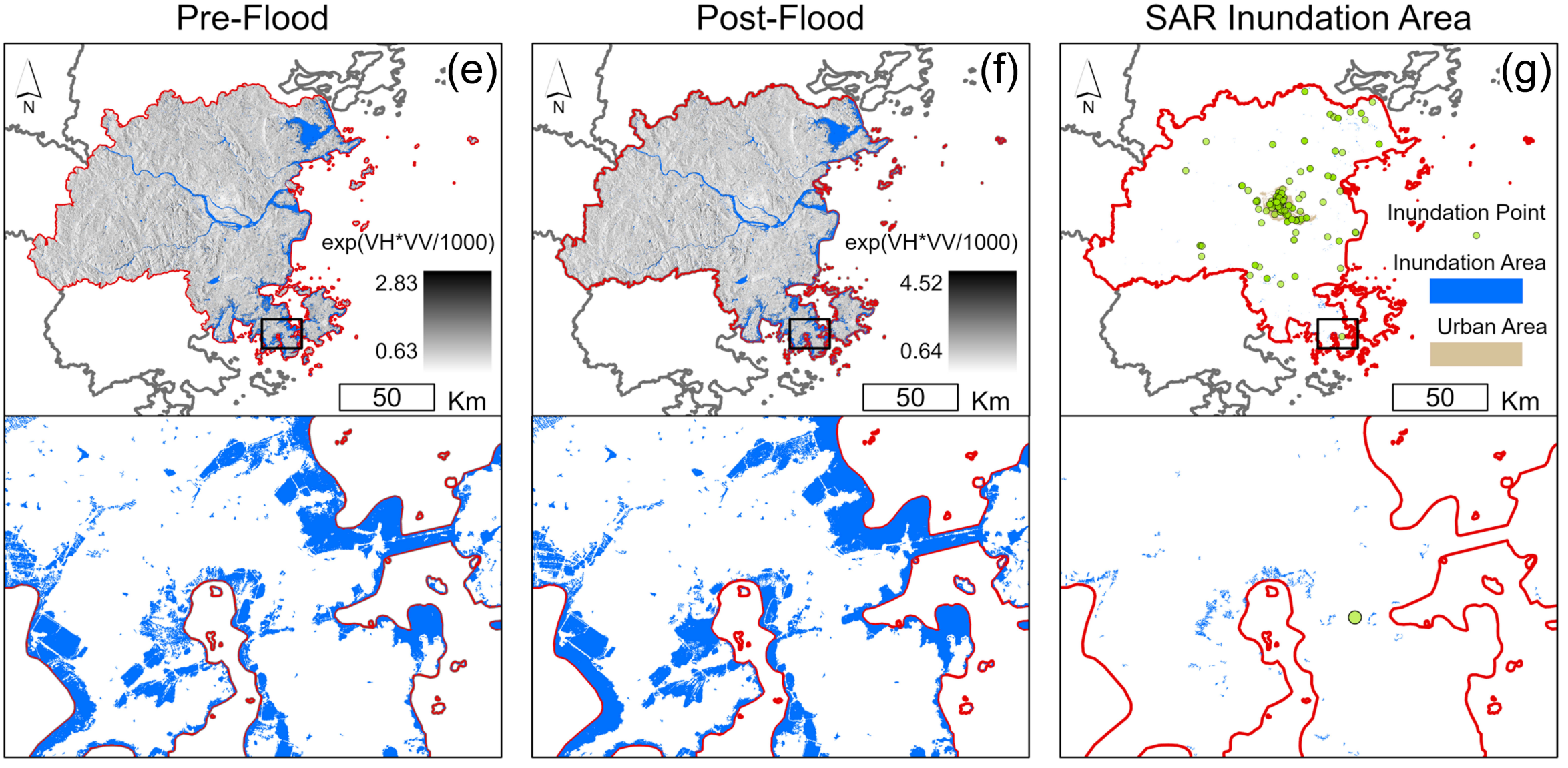
(c)




(a)
Number of Flood Events
0
20
40
60
80
100
Total Number of Flood Events
Detected by Remote Sensing
Detected by Optical Imagery Only
Detected by SAR Only
Detected by Both Optical and SAR
(b)
Number of Flood Events
0
20
40
60
80
100
Total Number of Flood Events
Undetected by Remote Sensing
Cloud Interference or Incomplete Coverage
No Available Imagery
No Discernible Flood Features
2012
2014
2016
2018
2020
2022
2024


(a)
Flood Event Duration
Setinel-1 (S1)
Setinel-2 (S2)
Landsat-8 (L8)
Landsat-7 (L7)
Number of Posts
150
100
50
0
S1
S2
S1
L7
S1
S2
L8
2023-07-26
2023-07-28
2023-07-30
2023-08-01
2023-08-03
2023-08-05
2023-08-07
2023-08-09


Pre-Flood
Post-Flood
Optimal Inundation Area
(b)
(c)
(d)
N
NDWI Value
1
-1
50
Km
Inundation Point
Inundation Area
Urban Area
Pre-Flood
Post-Flood
SAR Inundation Area


(e)
(f)
(g)
N
exp(VH*VV/1000)
2.83
0.63
4.52
0.64
50
Km
Inundation Point
Inundation Area
Urban Area