Cartography and Geographic Informa

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tcag20

Algorithmic uncertainties in g media data for disaster manage

Debayan Mandal, Lei Zou, Joynal Abedin, Bing Lin & Heng Cai

CARTOGRAPHY AND GEOGRAPHIC INFORMATION SCIENCE
2024, VOL. 51, NO. 4, 565–582
https://doi.org/10.1080/15230406.2023.2286385

Algorithmic uncertainties in geolocating social media data for disaster management

Debayan Mandal, Lei Zou, Joynal Abedin, Bing Zhou, Mingzheng Yang, Binbin Lin and Heng Cai

Department of Geography, College of Art & Sciences, Texas A&M University, USA

ABSTRACT

The rapid development of information and communications technology has turned individuals into sensors, fostering the growth of human-generated geospatial big data. In disaster management, geospatial big data, mainly social media data, have opened new avenues for observing human responses to disasters in near real-time. Previous research relies on geographical information in geotags, content, and user profiles to locate social media messages. However, less than 1% of users geotag their messages, leaving geolocating users through user profiles or message content addresses increasingly crucial. This paper evaluates and visualizes the margin of error incurred when using user profiles or message-mentioned addresses to geolocate social media data for disaster research. Using Twitter data during the 2017 Hurricane Harvey as an example, this research assessed the inconsistencies in predicting users' locations in various administrative units during each disaster phase using three geolocating strategies. The results reveal that the similarities between geotags, and user profile locations decrease from 94.07% to 64.56%, 43.9%, 31.82%, 27.05%, and 26.7% as the geographical scale changes from country to state, county, block group, 1-kilometer, and 30-meter levels. These similarities are overall higher than the agreements between locations derived from geotags and tweet content. The geolocation consistencies among the three methods remain stable across disaster phases. The impacts of uncertainties in geolocating Twitter data for disaster management applications were further unraveled. The findings offer valuable insights into the trade-off between spatial scale and geolocation accuracy and inform the selection of appropriate scales when applying different geolocating strategies in future social media-based investigations.

ARTICLE HISTORY
Received 1 February 2023
Accepted 14 November 2023

KEYWORDS
Geolocating; spatial uncertainty; social sensing; Twitter; disaster management

1. Introduction

Social sensing refers to the use of geospatial big data, e.g. social media, mobile phone, taxi trajectory, and smart card data, to reveal socioeconomic characteristics (Liu et al., 2016). It supplements traditional remote sensing and survey by providing in-depth insight into human behaviors, spatial interactions, and place semantics of the society at an unprecedented scale in near real-time. One type of readily accessible and ample social sensing data is social media data, i.e. data collected from Twitter (Twitter was renamed X in July 2023; this article uses Twitter to avoid confusion), Facebook, Weibo, Instagram, and other social media platforms. Social media usage coupled with Global Navigation Satellite Systems (GNSS)-enabled portable devices has become an indispensable part of daily life worldwide, which turns every user into a sensor capturing a direct snapshot of human activities at different places (Zhou et al., 2022). Therefore, various groups like the government, non-governmental organizations (NGO), companies, and researchers have started using such user-generated data to access the information flow among societies and explore its applications in multiple fields, e.g. crisis management, disaster relief, political sway, and religious and economic trends (Graham et al., 2014; Kryvasheyeu et al., 2016). Most recently, since the outbreak of the COVID-19 pandemic in 2020, a rise has been seen in using social media data for tackling COVID-19-related investigations and other epidemiology studies (Cinelli et al., 2020; Tsao et al., 2021; Wong & Wong, 2021).

To analyze and compare social media activities across space, data collected from social media need to be associated with locations. For example, Twitter messages, referred to as tweets, can commonly be geolocated through three attributes, (a) geotagged points or places reported by the device GNSS (referred to as *tweet from*), (b) locations mentioned in tweet content (*tweet about*), and (c) user profile addresses (*user from*), and each method has limitations. Geotagged locations are considered ground truth of user locations, but geotagged tweets form a small percentage of the data procured. A recent investigation statistically analyzed global Twitter data in 2020 and found that only

CONTACT Lei Zou lzou@tamu.edu

0.04% were geotagged (Lin et al., 2022). Hence, spatial analysis relying on geotagged tweets may overlook most of the information.

In the absence of geotagged data, most researchers turn to geo-enriched data to access a reasonable sample size. Geo-enrichment of Twitter data usually consists of the following methods. If any address is mentioned in the tweet (*tweet about*), the address is used to represent the user's location. Nevertheless, the address mentioned in a tweet does not necessarily reflect the user's current location. Alternatively, the "*user from*" location in the user profile can be considered. However, not every user updates their profile information promptly. If one user creates a profile and moves to another place without updating the profile address, the user profile still shows the previous address erroneously as their current location. Hence, leveraging "*user from*" or "*tweet about*" addresses to geolocate urgent and location-sensitive disaster-related tweets, e.g. rescue requests or onsite damage reports, may misguide first responders in emergency rescue and resource distribution missions. Concerns regarding the consequences of similar algorithmic uncertainty in harvesting and processing geospatial big data have been raised and elaborated upon (Kwan, 2016). However, the reliability of predicting user locations with the "*tweet from*", "*user from*", or "*tweet about*" information at different scales and how such uncertainties may affect disaster management leveraging social media is yet to be investigated thoroughly.

Therefore, the purpose of this study is to quantify the algorithmic uncertainty of analyzing geolocated tweets for disaster research and management when "*tweet about*", "*user from*", and "*tweet from*" information is used. We collected Twitter data during the 2017 Hurricane Harvey, one of the deadliest hurricanes in U.S. history, causing billions of dollars of economic loss and intensive discussions on Twitter. This study has two specific objectives: (1) to develop a framework for gauging and visualizing agreements of representing user locations through the "*user from*" and "*tweet about*" methods with the "*tweet from*" information, in different disaster phases at multiple geographical scales; (2) to showcase the implications of uncertainties in geolocating social media analysis for disaster research and management.

It is worth noting that while this study utilizes Twitter as its primary data source, the uncertainty analysis methods and subsequent insights developed are platform-agnostic and can be adapted to various other social media platforms that are rich in data for disaster management. Although Twitter has undergone several changes regarding data availability, its broad reach and influence remain very relevant. We also posit that during natural hazards, platforms like Twitter may provide crucial and often more accessible channels of communication for both affected individuals and rescue organizations (Hu & Wang, 2020a; Hu & Wang, 2020b; Mihunov et al., 2020).

This paper is organized as follows. Section 2 reviews the literature on how scholars have used different geolocating strategies to analyze social media data for disaster research. This section also highlights the primary need and novelty of our research, which aims to reveal and correct the prevalent misuse of multiple geolocation methods in academic work and disaster management scenarios. In Section 3, Hurricane Harvey, Twitter data collection and processing, as well as the analysis methods are introduced. Section 4 details the results obtained – with error distance and administrative boundary differences, along with possible inferences. Section 5 wraps up the paper with a discussion about the implications of this work and how it can act as a Rosetta stone for future case studies and research designs when applying different geolocating methods for social media data analytics.

2. Literature review

2.1. Geolocating strategies used in social media data analysis

Since the emergence of social media, the popularity of incorporating social media data and platforms into disaster research and practices has grown. In previous studies, social media data have been extensively used in both man-made and natural disaster events, e.g. wildfires (Sachdeva & McCaffrey, 2018), earthquakes (Earle et al., 2010), mass shootings (Rusho et al., 2021; Sutton et al., 2008), hurricanes (Mihunov et al., 2020; Zou et al., 2018), etc. This section summarizes various investigations using social media for disaster management and their geolocating strategies, as shown in Table 1.

The primarily used attribute for locating social media data is geotags. A geotag ("*tweet from*" in Twitter) contains point-based precise latitude and longitude or a polygon-based bounding box with four pairs of coordinates. Geotagged social media data are considered ground-truth locational data that can render accurate results at a fine spatial resolution of 10 meters (Jurdak et al., 2015; McNeill et al., 2017). Therefore, this method has been repeatedly used in disaster management studies. Among all the social media data sources, Twitter is the most widely used data source, particularly due to its ease of availability. Reynard and Shirgaokar (2019) inquired about the usefulness of geotagged Twitter data for resource allocations during 2017 Hurricane Irma in Florida. This investigation demonstrated that

Table 1. Examples of social media analysis investigations with various geolocating methods.

Article	Data	Study Area	Spatial Resolution	Application (Case Studies)
Geolocating Method: Geotag				
(Maas, 2019)	Facebook	India & USA	City	Disaster response mapping (2018 Cyclone Gaja, 2017 Tubbs Fire)
(Aulov & Halem, 2012)	Flickr	Gulf of Mexico, USA	Point	Oil spill trajectory monitoring (2010 Deepwater Horizon Oil Spill)
(Havas et al., 2017)	Twitter	Napa, California, USA	City	Mapping earthquake footprints (2014 Napa Earthquake)
(Mendoza et al., 2019)	Twitter	Chile	Municipality	Earthquake response & damage estimation (2010 Chile Earthquake)
(Wang et al., 2015)	Weibo	Beijing, China	City	Emergency response (2012 Beijing Rainstorm)
(Reynard & Shirgaokar, 2019)	Twitter	Florida, USA	Point	Disaster Response Planning (2017 Hurricane Irma)
Geolocating Method: User Profile				
(Premamayudu, 2019)	Twitter	India	Point	Detecting disaster-affected areas (Multiple Events)
(Zou et al., 2019)	Twitter	USA	County	Disparities in disaster-related Twitter use (2017 Hurricane Harvey)
(Carley et al., 2016)	Twitter	Indonesia	Country	Improving emergency early warning systems (Multiple Events)
(Karimiziarani et al., 2021)	Twitter	USA	County	Hazard risk assessment (2017 Hurricane Harvey)
Geolocating Method: Message Content				
(Zhou et al., 2022)	Twitter	Texas, USA	Point	Locating rescue requests on social media (2017 Hurricane Harvey)
(Kumar & Singh, 2018)	Twitter	Iran, Mexico, Iraq, Philippines, Algeria, the United States, and Peru	Country	Extracting earthquake affected locations (Multiple Events)
(Zahra et al., 2017)	Twitter	Italy, Myanmar	Country	Digital divide in disaster social media uses (2016 Myanmar & Italy Earthquakes)
(Yousefinaghani et al., 2019)	Twitter	Bulgaria & Philippines	Country	Detecting avian influenza outbreak (2017–2018)

training machine learning algorithms to extract location-based sentiment, damage, and rescue information from geotagged tweets during the early disaster preparedness period is beneficial to formulate effective disaster responses. Apart from hurricanes, geotagged Twitter data were used in other disasters. Mendoza et al. (2019) estimated the spatial distribution of damaged buildings after an earthquake in Chile by combining earthquake intensities at the municipality level with the frequency of geotagged tweets. Resch et al. (2017) examined the use of Twitter data during the 2014 Napa earthquake in California, United States. A total of 1,012,650 geotagged tweets falling within Napa from August 16–31, 2014, were collected. The analysis revealed that Twitter data are faster to identify earthquake footprints compared to traditional datasets.

Similarly, geotagged data from other social media sites have been used for disaster studies. Wang et al. (2015) examined density-based distribution patterns of 706,835 posts on Weibo (a Chinese microblogging platform) during the 2012 Beijing Rainstorm to find clusters for the distribution of resources during emergency response. Maas (2019) exploited the use of geotagged Facebook data during Cyclone Gaja and Tubbs fire. They formulated five disaster maps that can be of use to first responders in places where Facebook penetration is high, i.e. Facebook population, movement, power availability, network coverage, and displacement. Moreover, Aulov and Halem (2012) used the oil spill images captured by geotagged Flickr data to estimate the oil spill rate and diffusion. However, these studies were limited by a lack of data. Less than 1% of data collected from social media are geotagged as many users turn off location-sharing due to privacy concerns (Graham et al., 2014). Meanwhile, social media companies have implemented more strict policies for accessing and sharing users' location data. Amidst all these, Twitter has removed the feature of precise geotagging since 2019 (Porter, 2019).

One alternative attribute to geolocate social media data is user profile locations (*user from* in Twitter) (Hecht et al., 2011). Researchers have attempted to leverage this method for disaster management studies. Zou et al. (2019) evaluated the county-level public awareness and sentiment toward Hurricane Harvey using 224,327 tweets geolocated by user profiles from August 17 to 12 September 2017. They reported that communities with higher disaster-related Twitter activities had better socioeconomic conditions and were closer to the hurricane track during the preparedness period. Carley et al. (2016) scrutinized the viability of leveraging social media to improve early warning systems and response plans in Indonesia, based on the spatial activity of Twitter users geolocated by the "*user from*" method. Karimiziarani et al. (2021) leveraged user profiles to geolocate Twitter activities during the 2017 Hurricane Harvey and analyzed the processed tweets to evaluate county-level hazard risk. Similarly, Premamayudu (2019) showed that affected areas from different types of natural disasters could be identified based on the profile locations of users whose tweets were relevant to disaster events. However, Cheng et al. (2010) reveal that most user profiles do not contain fine-scaled location information to support spatial analysis. The study shows that in a random sample of one million Twitter users, only 26% provide a city name in their user profile locations, with the others either being overly general (state/country) or nonsensical texts. Furthermore, the reliability of this method relies on how frequently users update their information.

Another feature of geolocating social media is the content-mentioned locations. It can be extracted through machine learning and deep learning models. Kumar and Singh (2018) proposed a convolutional neural network (CNN) model to extract location names. An F1 score of 0.92 was obtained, when the CNN model was applied to earthquake-related tweets collected from 20 October 2017, to 15 March 2018, in various countries. Lingad et al. (2013) evaluated the effectiveness of different Named Entity Recognition (NER) tools, e.g. Stanford NER, OpenNLP, Yahoo! PlaceMaker, and TwitterNLP in recognizing locations from Twitter data, and inferred that the Stanford NER performs the best. Zahra et al. (2017) explored the granularity of locations parsed from tweets posted during the 2016 Italy and Myanmar earthquakes and thus uncovered the differences in disaster reporting behaviors between Asian and European users. Zhou et al. (2022) built a rule-based model to extract locations from rescue request tweets during Hurricane Harvey. Yousefinaghani et al. (2019) used content from tweets in Bulgaria and the Philippines to detect Avian Influenza outbreaks in real-time using 209,000 tweets over the period from July 2017 to November 2018.

The process of converting user profile-mentioned or content-mentioned address names or descriptions into coordinates is toponym resolution. This process plays a vital role when the "*user from*" or "*tweet about*" method is employed. It can be achieved using lookup tables or geocoding services. Chen et al. (2014) matched the user profile locations with the place names from the Topologically Integrated Geographic Encoding and Referencing (TIGER) dataset to geolocate Twitter data and determine religious distributions in the United States. Backstrom et al. (2010) used TIGER data to geocode Facebook users and investigate their displacement. Meanwhile, studies like Zou et al. (2018), Carley et al. (2016), Karimiziarani et al. (2021), and Kumar and Singh (2018) have leveraged the available geocoding services, e.g. Google and Open Street Maps, to geolocate social media data. Previous efforts compared publicly available geocoding services offered by Centrus, ESRI, Geocoder.us, Geolytics, Google Maps, Yahoo Maps, and the University of Southern California, and reported that the Google geocoding service yields the best performance (Goldberg et al., 2013; Swift et al., 2008).

2.2. *Conflated geolocating strategies usage*

Many past studies have utilized the three geolocating methods interchangeably, an issue that has often been overlooked. With geotagged tweets comprising less than 1% of the Twitter databases, studies relying solely on these messages may encounter challenges such as limited sample size and scale inconsistency. This fact illuminates a crucial limitation in the field of research that relies heavily on social media data, particularly in the fields of disaster response and management. This work contributes to the field by shedding light on the motivations behind users specifying addresses in various attributes and highlights the degrees of agreeability amongst different geolocating methods.

The utilization of multiple geolocating methods to increase the amount of usable social media data for analysis has been a common approach of "geo-enrichment" in existing literature. This is particularly relevant when considering the reproducibility and replicability of GIScience research. The inferences are dependent on accurate and detailed geolocation data, and the choice of geolocation method can significantly impact the results of a study. By systematically examining and distinguishing the differences between these methods, we can significantly enhance the robustness and reliability of future research, contributing to the overall credibility of GIScience research. This section summarizes various investigations using social media for disaster management which have used a combination of geolocating strategies, as shown in Table 2.

Numerous studies in the past have demonstrated the combined usage of geotags and tweet content to infer insights about a wide array of disaster investigations. Belcastro et al. (2021) proposed a model known as SEDOM-DD to detect critical local cascading damage events during disasters. The model was applied to tweets from the Barletta and Peru earthquakes in Italy and Peru respectively that occurred between May 21 and 26, 2019. The data used in this study were geolocated using geotags and tweet content. The study concluded that SEDOM-DD can enhance emergency response by identifying critical sub-events during a disaster. Huang et al. (2022) devised a new text clustering approach named EDEE to detect early emergency events (for multiple disasters – Flood, Typhoon, Tornado, Rainstorm, Snowstorm, Earthquake, Landslide, Collapse, Forest Fire) from social media. The method was applied to a collection of 890,938 Weibo posts collected during the Xiangshu earthquake in Yancheng City, China, and geolocated using geotags and the content of the posts. The results demonstrated that EDEE could develop a cloud service system for the early detection of emergency events from social media. The effects of machine learning algorithms in understanding neighborhood equity from Twitter data were explored by Zhai et al. (2020) during Hurricane Florence. They utilized a dataset of 2.2 million tweets from September 10 to 20, 2018, in North and South Carolina, USA, which were geolocated utilizing the geotags and tweet content. Using various models, they discovered that while tweets from poorer neighborhoods are more negative, the predominantly black neighborhoods, on the other hand are less likely to be negative, thereby exhibiting different sentiment patterns, highlighting social and economic inequities in disaster response. Havas et al. (2017) introduced a novel system architecture for emergency management service (E2mC) by harnessing social media and crowdsourcing analysis in near real-time. This research geolocated Twitter data relevant to the 2014 Napa Earthquake and the 2016 Central Italy Earthquake using geotags and tweet content to demonstrate social media's ability to provide early information on disaster impact and ongoing updates that can support satellite imagery interpretation and task coordination throughout the event.

In parallel, several studies have utilized a combination of geotags and user profile locations to analyze social media data. Skripnikov et al. (2021) utilized Twitter activity to assess the community impacts of the Florida red tide event from 2017 to 2019. By examining 18,082 tweets geolocated using both geotags and user profile locations across the Pasco, Hillsborough, Pinellas, Manatee, and Sarasota regions, they gauged the sentiments and perceptions of local communities about the disaster. Kryvasheyeu et al. (2016) leveraged 52.55 million Twitter posts, using both geotagging and user profile locations for geolocation, to rapidly assess disaster damage during Hurricane Sandy. The study found significant associations between Twitter activities and disaster damages at multiple spatial scales based on the geo-enriched database. Wang et al. (2021) analyzed over 1.6 million tweets during Hurricane Isaac to measure disaster resilience. The posts were geolocated using geotags and user profile locations. Analysis results reveal a significant correlation between tweet activity and sentiment computed from the enriched database and resilience indexes during natural disasters, implying the potential of estimating disaster resilience with social media data. Yin et al. (2012) employed data mining and natural language processing techniques to derive

Table 2. Examples of social media analysis investigations combining geolocating methods.

Article	Disaster	No. of Posts (Source)	Time	Study Area	Scale	Application
Geolocating Methods: Geotag and Tweet Content						
(Belcastro et al., 2021)	Barletta and Peru earthquake	10,029,395 (Twitter)	05/21/2019 and 05/26/2019	Barletta (Italy) and Peru (Peru)	City	Disaster Subevent identification
(Huang et al., 2022)	Xiangshui earthquake	890,938 (Weibo)	03/21/2019 (For the Case Study)	Yancheng City (China)	City	Proposed Early Detection of emergency events framework
(Zhai et al., 2020)	Hurricane Florence	2.2 million (Twitter)	09/10/2018 to 09/20/2018	North Carolina and South Carolina (USA)	State	Assessing situational awareness with neighborhood equity
(Havas et al., 2017)	Napa (2014), Central Italy Earthquake (2016)	1,012,650 and 152,062 (Twitter)	08/16/2014 to 08/21/2014 and 08/24/2016 to 08/26/2016	North San Francisco Bay (USA), Central Italy (Italy)	Point	Emergency response architecture (E2mC)
Geolocating Methods: Geotag and User Profile						
(Kryvasheyeu et al., 2016)	Hurricane Sandy	52.55 million (Twitter)	10/15/2012 to 11/12/2012	50 most populous urban areas according to the 2010 U.S. Census (USA)	State, National	Disaster Damage Analysis
(Yin et al., 2012)	Eight major disasters, both natural and manmade, in 2010–2011.	66 million (Twitter)	March 2010 to February 2011	Australia and New Zealand	Country	Situation awareness information extraction
(Skripnikov et al., 2021)	Florida red tide event (2017 to 2019)	18,082 (Twitter)	05/15/2018 to 05/15/2019	Pasco, Hillsborough, Pinellas, Manatee, Sarasota (USA)	County	Community perceptions analysis
(Wang et al., 2021)	Hurricane Isaac (August 2012)	1,686,851 (Twitter)	08/21/2012 to 09/17/2102	146 counties in Louisiana and Mississippi (USA)	County	Twitter activity and Sentiment analysis

situational awareness information from 66 million Twitter messages. This large dataset, geolocated using geotags and user profile locations, pertained to multiple disasters and crises in Australia and New Zealand since March 2010. Their analysis utilized burst detection, text classification, and online clustering methodologies, demonstrating the significant role social media plays in augmenting emergency awareness.

Despite the importance of geolocating social media data for disaster management, the potential uncertainties of using different geolocating methods interchangeably or simultaneously have not been explored thoroughly. Therefore, there is an urgent need to address the following research questions to select appropriate geolocating strategies in location-based social media analytics and applications: (1) does uncertainty in geolocating social media data decrease with increasing the spatial scale of the analysis? (2) Is there a certain spatial scale that the error factor would be admissible for analysis? In this study, spatial scale refers to the level of administrative units being examined, rather than the granularity or resolution of the data being used. Also, a universally accepted standard for an admissible level of uncertainty when using geocoded Twitter data for analyzing disaster impacts and supporting disaster management is not agreed upon. The acceptable level of uncertainty depends on the specific context and objectives of the research. For example, studies focusing on localized impacts may require a higher degree of precision, e.g. street, or individual levels (and thus lower uncertainty), than those examining broader regional trends at the community level. Answering the second research question necessitates a comprehensive examination of the consistencies of locations obtained from different geolocating methods. This analysis is poised to address the research gap through a case study examining the agreements of three commonly used methods in geolocating disaster-related Twitter data during the 2017 Hurricane Harvey. The findings could prevent potential judgment errors arising from combining different types of location attributes.

3. Data and methods

3.1. Hurricane harvey

Hurricane Harvey was a catastrophic event that came upon the Caribbean and Latin American countries before its landfalls in Texas and Louisiana, United States, in August 2017 (Figure 1). It was a category-4 hurricane on its first landfall in Texas and lasted 15 days from August 17 to 2 September 2017 (Mooney, 2021). As a wet tropical cyclone, Harvey caused over 40 inches of rainfall within a week in the city of Houston and its surrounding areas, resulting in unprecedented flooding in most Texas and Louisiana coastal zones and a total of $125 billion in damage, the second costliest tropical cyclones in the United States (National Oceanic and Atmospheric Administration, 2018). Many Houston residents were caught unprepared for the heavy precipitation and sought social media to access disaster-related information, stay connected with families and friends, request help, and report local damages. Therefore, Harvey is an ideal case study for analyzing the disagreements of locations derived from the massive social media data through multiple methods.

3.2. Twitter data collection and preprocessing

Figure 2 depicts the Twitter data collection, preprocessing, and analysis workflow in this investigation. Harvey-related tweets were retrieved by a list of predefined keywords through the Twitter Enterprise Application Programming Interface (API), which can access the complete historical Twitter data. The Twitter dataset used in our study, which corresponds to Hurricane Harvey, spans from 17 August 2017, through 7 September 2017. The keyword list is [*harvey, hurricane, storm, flood, houston, txtf* (Texas Task Force), *coast guard, uscg* (U.S. Coast Guard), *houstonpolice, cajun navy, fema* (Federal Emergency Management Agency), *rescue*]. Any tweet containing one or several keywords in the list was identified as Harvey-related and collected in the initial database, resulting in 47 million tweets. Next, we examined the attributes of each collected tweet and kept those having geotags for the subsequent analysis. A total of 588,401 tweets were retrieved.

The filtered data were geolocated through "*tweet about*" and "*user from*" approaches. The user profile locations were extracted from tweet details. Among geotagged tweets, 503,898 (85.64%) had user profile locations attached to them. To ensure the accuracy and quality of the data, we implemented a three-step procedure to identify and remove bot-generated tweets. First, the "Device" field of each tweet was examined thoroughly for the presence of the term "bot" or the use of commonly used applications by corporations like TweetDeck, Percolate, Media Studio, etc. Second, the User field was scrutinized. Finally, accounts exhibiting a high disparity between the number of followers and friends were manually reviewed to identify any potential signs of being bots. This rigorous screening process ensured that the remaining data were of high relevance and quality for further analysis.

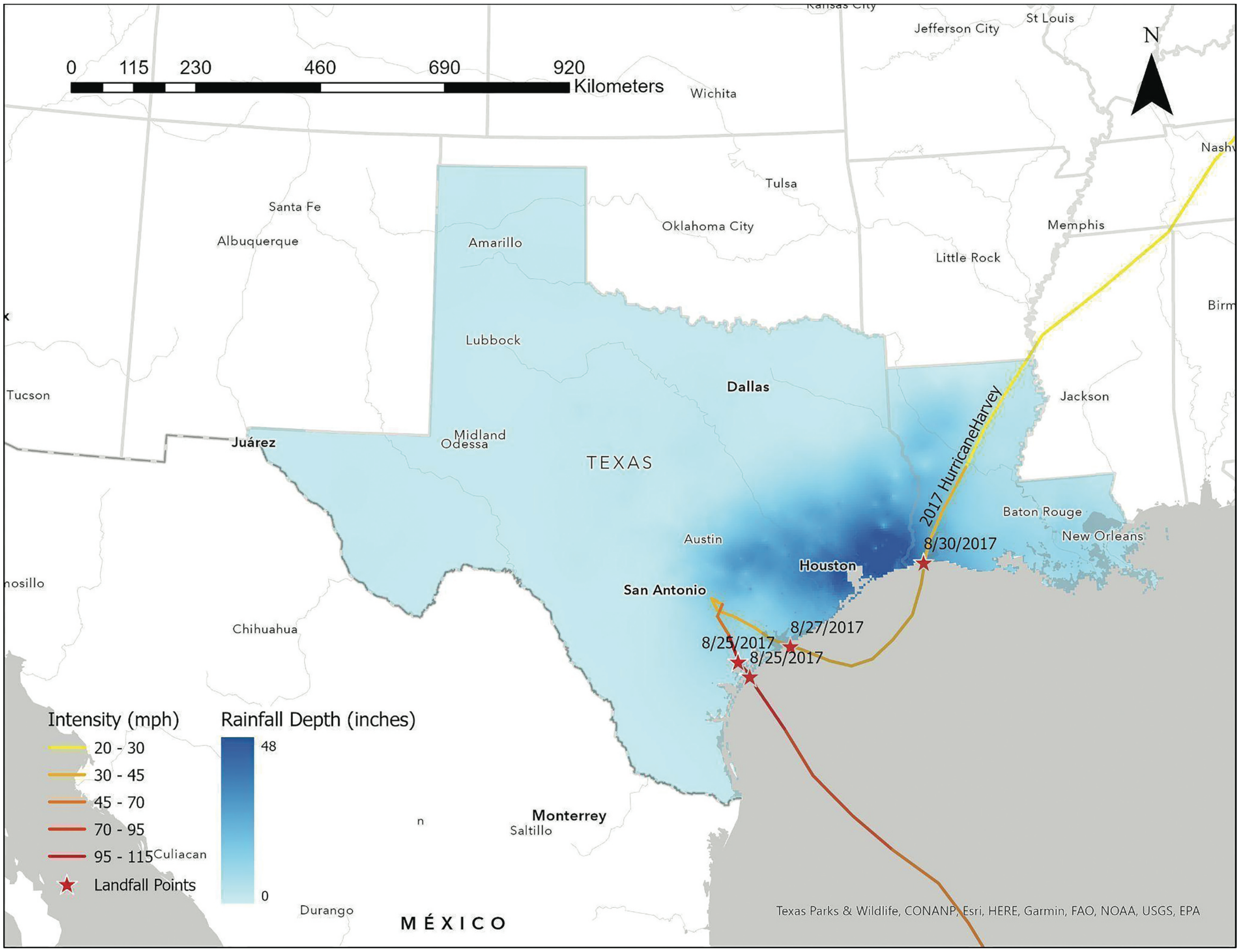

Figure 1. Track and rainfall depth of 2017 Hurricane Harvey.

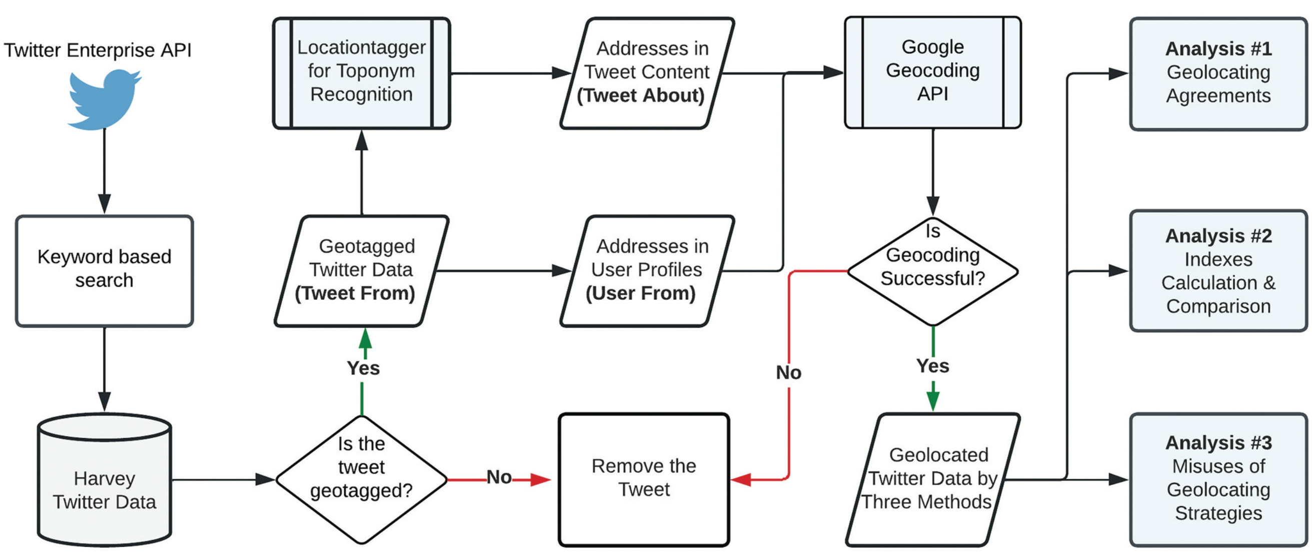

Figure 2. The workflow of Twitter data collection, preprocessing, and analysis.

Concurrently, a python package "locationtagger" (Soni, 2020) was used to recognize tweet-mentioned addresses. The "locationtagger" package is designed to recognize and filter location terms from Named Entities (NEs) extracted through Named Entity Recognition (NER) methods, such as those used in Natural Language ToolKit (nltk) and Spacy. Prior to the extraction of location terms, the package performs several text pre-processing steps, including the removal of HTML tags, multiple spaces, and non-ASCII characters, to enhance the accuracy of location identification.

The identification of location terms includes two main steps. The first step utilizes the pycountry library to identify country names within the NEs, which does not require additional geographical data parsing. The second step relies on an extensive database of city and region names, Geolite2, to detect city or region names within the NEs. The Geolite2 data are freely accessible, and the package matches the identified NEs with this dataset to verify if they are city or region names. As a versatile tool, locationtagger can extract location terms from texts, and it also identifies the relationships between the detected location entities. It is worth noting that the package allows for data updates to ensure that the geographical information used for location identification is accurate and comprehensive. While the locationtagger package provides significant advantages in ease of use and specific focus on location entities, it may have limitations, particularly when applied to complex NLP tasks. Evaluating its performance using the CoNLL test dataset, we obtained a precision score of 0.5415 and an F1 score of 0.5068. These scores suggest that the package provides moderate performance on this standard NLP benchmark, but results may vary depending on the specific use case and nature of the data.

Keywords like "harvey" and "hurricane" are relevant to the disaster event rather than location names, so they were excluded from the tweet content before inputting Twitter data for toponym recognition. Within geotagged messages, 153,434 (26.08%) had parsable locational information in their content. From these, a fraction comprised of retweets, bot-generated posts, and news-based tweets, specifically 87 (0.05%), 1,592 (1.03%), and 100 (0.07%) respectively. However, these were subsequently removed to enhance the authenticity and validity of the data pool. The user profile locations and parsed text addresses were then geocoded using the Google geocoding API, which offers 40,000 free geocoding requests per month for each developer's account. The reliability and accuracy of geocoding plays a vital role in our research, and for this purpose, we have utilized Google geocoding service. This service is renowned for its reliability, with a 99.90% Uptime Service Level Agreement (SLA) and impressive processing speed, handling up to 50 requests per second and geocoding 10,000 addresses in around 3 minutes. The geocoding accuracy of Google's service is presented through a location-type field. This categorizes the results into several levels of precision: ROOFTOP which provides a precise geocode at the street address level; RANGE_INTERPOLATED for an approximate, interpolated position on a street; GEOMETRIC_CENTER offering the geometric center of a result such as a polyline or a polygon; and APPROXIMATE that gives a general approximation of location.

3.3. Methods of analysis

Three types of analyses were conducted to compare the three geolocating methods: assessing geolocating agreements, Twitter-based index comparison, and demonstrations of misuses of geolocating strategies in disaster research and practice. The first analysis is to evaluate the agreements of geolocating Twitter messages by *tweet from*, *user from*, and *tweet about* methods. The geotagged *tweet from* coordinates reflect the locations where users post messages on Twitter, so they were considered the reference in the first analysis. Initially, we calculated the displacement distance between coordinates derived from the geotag and the other two methods for each tweet using the Haversine formula (Equations (1) and (2)):

$$a = Haversine\,(L_2 - L_1) + Cos\,(L_1)\,Cos\,(L_2)\,Haversine\,(Lo_2 - Lo_1) \quad (1)$$

$$D = R * \left[atan2\left(\sqrt{a}, \sqrt{1-a}\right)\right] \quad (2)$$

where R is the Earth radius, D is the displacement distance between points 1 and 2, L_1 and L_2 are their latitudes, Lo_1 and Lo_2 are their Longitudes, and Haversine (θ) is calculated as $\sin^2(\theta/2)$. We computed the percentage of agreements among three geolocating methods by different spatial scales, i.e. perfect match (a tweet is geolocated to the same point by the two methods), 30-meter and 1-kilometer (the distance between a tweet's locations obtained by the two methods is no more than 30 meters and 1 kilometer). In addition, many previous case studies aggregated social media data by administrative boundaries to analyze the association between social media activities and socio-economic or political features (Kryvasheyeu et al., 2016; Zou et al., 2018). Therefore, we spatially joined tweet locations derived from the three geolocating methods with block groups, counties, states, and countries and evaluated their identicality.

The geolocating agreements of different methods at various spatial scales in the pre-, during-, and post-disaster phases were also examined. The three phases have different human interactions as reactions to immediate events. The pre-disaster phase consists of the populace getting ready to face the disaster, the response period focuses on the time-sensitive rescue and repair after the disaster hit, and the recovery period is when communities are returning to the pre-disaster conditions (Zou et al., 2019). In previous efforts, the geolocated social media data have been intensively used in all three phases, including calculating disaster awareness based on discussions on social media in the pre-disaster phase to identify under-prepared neighborhoods, searching for victims sending help requests on social media in the during-disaster phase to conduct rescue operations, and identifying local damages reported on social media in the post-disaster phase to facilitate recovery. In this work, the periods are differentiated as follows as suggested by Zou et al. (2019): August 17 to 24 August 2017, as the preparedness period, August 25 to September 2 2017, as the response period, and September 3 to September 7 2017, as the recovery phase.

The second analysis is to measure the uncertainty in calculating Twitter-based indexes when using *user from* and *tweet about* strategies. We compared two commonly used indexes, Ratio and Sentiment. Ratio, also defined as situational awareness or discussion intensity, is calculated as the average number of tweets posted per capita in an area during a period by Equation 3 (Wang et al., 2021). In our analysis, we calculated a ratio based on the population at the administrative level. This approach, while having its limitations, has been found to be effective in reflecting the dynamics and responsiveness of the affected population. It is important to clarify that Twitter users do not represent a cross-section of the general population. The demographic skew toward younger age groups in the Twitter population is acknowledged (Wang et al., 2021). As such, the ratio indexes that demonstrate higher values can be more representative of a younger population that resides closer to the disaster epicenter, as per the findings of Wang et al. (2021). Therefore, the calculations should be interpreted considering these biases. The sentiment index is the average sentiment score of target tweets in an area in a duration (Equation 4) (Zou et al., 2018). This study chose the Valence Aware Dictionary and sEntiment Reasoner (VADER) sentiment analysis tool (Hutto & Gilbert, 2014), a lexicon and rule-based python package which produces sentiment scores. It has been proven a reliable tool for analyzing sentiment in Twitter data (Wang et al., 2022). This sentiment analysis tool is specifically attuned to the nuances of social media text. It applies specific rules to adjust the valence of words based on their context, handling polarity reversal, intensity changes, and degree modifiers. The final sentiment score assigned by VADER is a composite score, normalized between −1 (most negative) and + 1 (most positive), providing a robust measure of the sentiment conveyed in a text. We calculated the Twitter Ratio and Sentiment indexes at the county level, the most used spatial scale in location-based social media analysis, using data derived from the three geolocating methods and examined their Pearson correlations.

$$Ratio = \frac{\#Disaster - Related\ Tweets}{Population} \quad (3)$$

$$Sentiment\ Index = Mean\,(Sentiment\ Scores) \quad (4)$$

The last analysis is to demonstrate the potential consequences when erroneously applying "*tweet from*", "*tweet about*", or "*user from*" locational information for disaster management. We specifically focused on Tweets requesting rescue or reporting damages. Two sets of keywords were used to identify rescue-related and damage-related tweets, including [*harveyform, harveyrescue, harveysos, cajunnavy*] and [*damage, blackout, power outage*]. The keyword-filtration yielded 1234 rescue requests and 2416 damage reports. The tweet content and *tweet about* locations were manually verified and updated for greater precision. Some of these tweets were removed due to their different context from Hurricane Harvey, e.g. political opinions, memes, citing previous hurricanes. The locations mentioned in the links, attached images, or videos were only considered if there was minimal location information in the tweet text. Prior to applying the "locationtagger" Python Package, we performed several pre-processing steps to enhance toponym recognition accuracy. We began by decoding specific acronyms such as "Txwx" and "houwx", signifying Texas weather and Houston weather, respectively. Next, we manually assigned geographical coordinates to specific place names, primarily institutions or notable locations within specific counties (e.g. "Ridge Point High School" in Fort Bend County and "Arkema Chemical Plant" in Crosby County). Spelling errors within these names were identified and rectified, ensuring the accuracy of location data. Following these pre-processing efforts, we assessed and visualized the displacement of tweets mentioning locations within Texas, a subset of which is displayed in Table 3. To maintain user privacy, numbers and links within the tweet examples have been substituted with "9999" and "url_link" respectively. The displacement of those messages with the *tweet about* location in Texas was evaluated and visualized. Some of the filtered tweets are displayed in Table 3. The numbers and links in tweet examples have

Table 3. Examples of rescue and damage-related tweets.

Categories	Tweets
Rescue Requests	We have stranded animals in need. The waters are coming. #HarveySOS #HarveyRescue
	Need rescue boat at *9999* Canford Ct Kingwood, TX 77345 United States #HarveySOS #HarveyRescue #CajunNavy
	9999 Kingwood Place – retirement home needing rescue - only boats can reach #HarveyRescue #harveysos
Damage Reports	Enormous damage along #Horsepen Creek #Houston #Copperfield #HurricaneHarvey @ Copperfield *url_link*
	This is nothing considering the fact that the flood damage alone is approximated to be $30 billion dollars *url_link*
	Finding the most damaged areas in #houston #harvey @tzuchiusa @ Briar Forest, Houston *url_link*

been replaced by "9999" and "url_link" to avoid disclosing users' privacy.

4. Results

4.1. Geolocating agreements by geographical scales

Figure 3 summarizes the agreements of employing different geolocating strategies by geographical scale. The agreement between geolocated tweets by the "*user from*" and "*tweet from*" methods at the country level is 94.07%, the highest among all tested scales. The agreements decrease from 64.56% to 43.9%, 31.82%, 27.05%, and 26.7% as the geographical scale changes from the state to the county, block group, 1-kilometer, and 30-meter levels. The percentage of perfect matches is the lowest at 24.82%. A similar trend is observed from the agreements between the "*tweet about*" and "*tweet from*" methods – the highest at the country level with a value of 81.57% but below 10% at the state level and finer geographical scales. The results demonstrate that the "*user from*" method is more similar to the "*tweet from*" approach than the "*tweet about*" approach in geolocating Twitter data. When predicting users' locations where they post messages on Twitter during disasters, the "*user from*" method performs well at the country and state levels. In contrast, the "*tweet about*" method shows acceptable accuracy only at the country level. Both methods do not produce satisfactory accuracies when applied on scales finer than the state level.

Figure 4 investigates spatial patterns of geolocating uncertainties by visualizing the agreements of the locations obtained by the two methods compared with the "*tweet from*" method at the state and county levels. The state-level agreements of the "*user from*" and "*tweet from*" method ranges from 16.23% to 74.64% (Figure 4a), with the highest in Texas, the prime affected state with the most tweet samples. However, the state-level consistencies of the "*tweet about*" and "*tweet from*" methods show a distinctly different pattern, with the values ranging from 0% to 52.6% (Figure 4b). Surprisingly, Texas is not among the states with the highest agreements despite being one of the affected states. Therefore, the result indicates that predicting user locations at the state level through the "*user from*" method is more reliable than the "*tweet about*" approach and is most effective in disaster-affected areas.

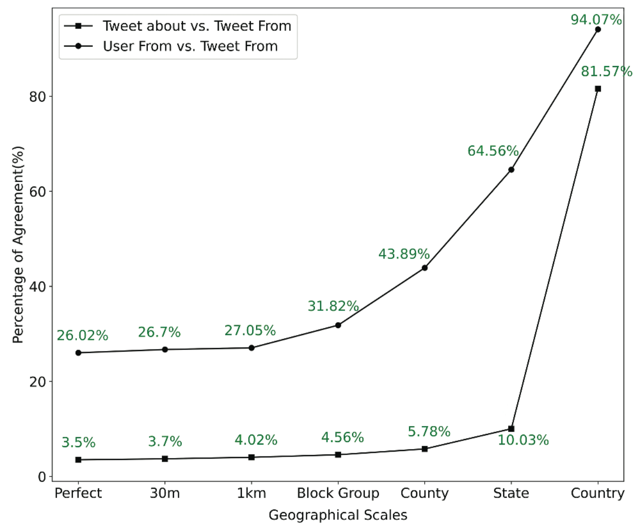

Figure 3. Geolocating agreements between the "*tweet from*" and the "*user from*" and "*tweet about*" methods at different geographical scales.

The county-level agreement analysis reveals more details about the uncertainty of employing different geolocating strategies to predict Twitter users' locations. Counties with less than ten geotagged tweets were removed to avoid the small number fallacy, which occurs when reaching an inductive generalization based on insufficient samples. A total of 1552 out of 3243 counties were included in this analysis. The county-level agreements between "*tweet from*" and "*user from*" locations range from 0% to 64% (average value: 43.9%), with a distribution lop-sided to below fifty percent, including Harvey-affected counties in Texas (Figure 4c). Counties with high agreements were primarily located in Texas and Louisiana states. The geolocating consistencies between the "*tweet about*" and "*tweet from*" methods range from 0% to 10.2%, with a mean value of 5.78%. The similarities between the "*tweet about*" and "*tweet from*" methods were the highest in the counties in Texas and Washington. Generally, the overall county-level agreements were

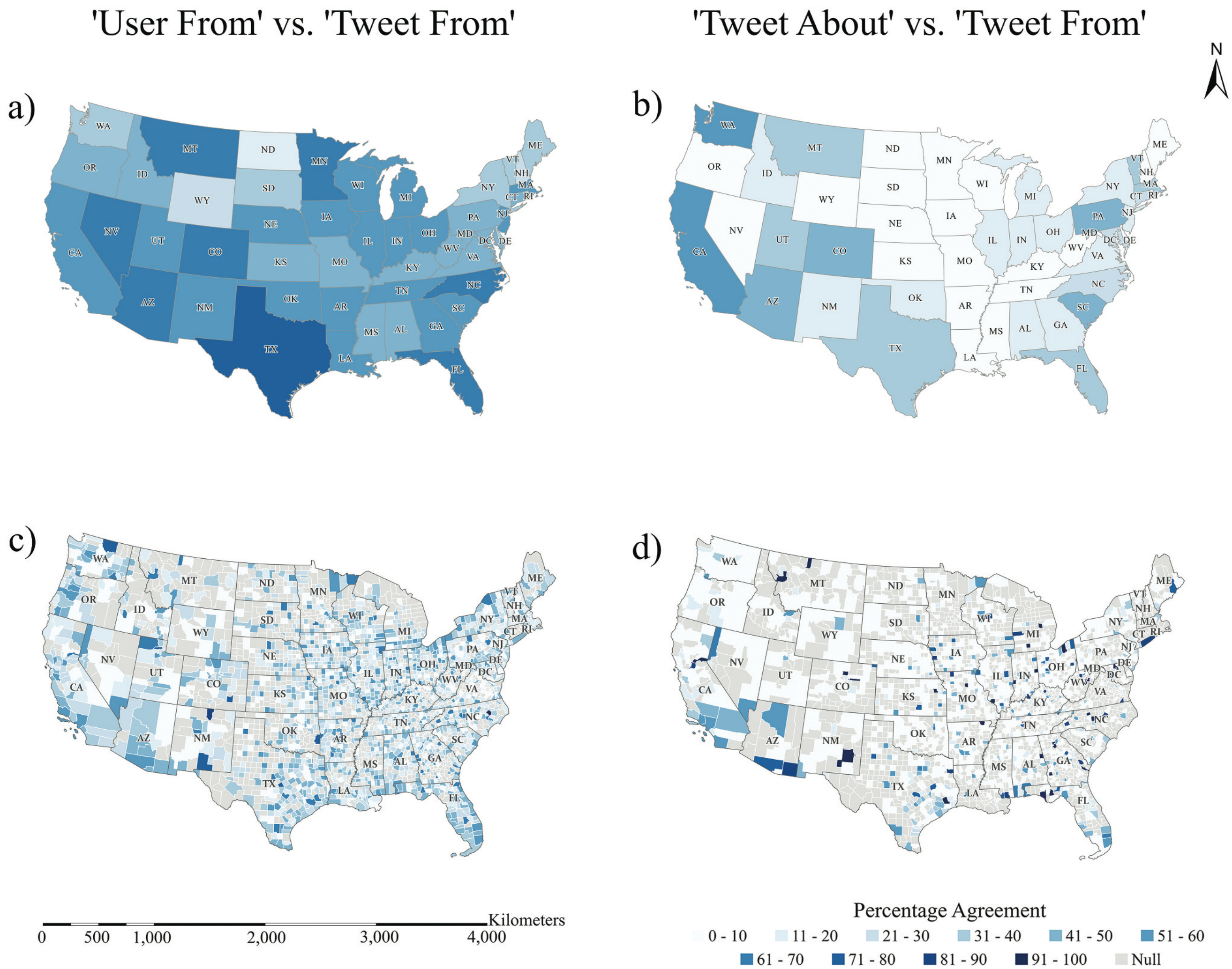

Figure 4. Spatial patterns of geolocating agreements between different methods at state and county levels: a) agreement between "*tweet from*" and "*user from*" methods at the state level, b) agreement between "*tweet about*" and "*tweet from*" methods at the state level, c) agreement between the "*tweet from*" and "*user from*" methods at the county level, and d) agreement between the "*tweet about*" and "*tweet from*" methods at the county level.

higher between "*tweet from*" and "*user from*" than between "*tweet about*" and "*user from*" locations.

Figure 5 illustrates the differences in geolocating agreements by different methods segregated through the three phases of disaster management. The agreements between "*tweet from*" and "*user from*" locations were the highest in the preparedness phase at the country, state, and county levels, with values of 95.1%, 62.5%, and 40.3%. In the perfect match, 30-meter, and 1-kilometer analyses, the agreements were the highest in the response phase, with values of 28%, 28.2%, and 29%. The geolocating similarities were lowest in the recovery phase at all geographical scales. When comparing "*tweet from*" and "*tweet about*" locations, the consistencies were the highest in the response phase and lowest in the recovery phase at most of the geographical scales. However, the agreements at the state and finer geographical scales were less than 10%, and the variations by disaster phases were minor.

The spatial autocorrelation of the percentage agreement distribution at the county level for both the comparisons was assessed using Moran's *I*. This technique helps determine whether there is a significant clustering or dispersion pattern in the data. For the "*user from*" and "*tweet from*" method comparison, Moran's Index was calculated as 0.009623, with an expected index of −0.000499, variance of 0.000013, and a z-score of 2.824787. This resulted in a p-value of 0.004731, demonstrating a significant spatial autocorrelation (clustered). Similarly, for the "*tweet about*" and "*tweet from*" comparison, Moran's I was computed as 0.019625, with an expected index of −0.000461, variance of

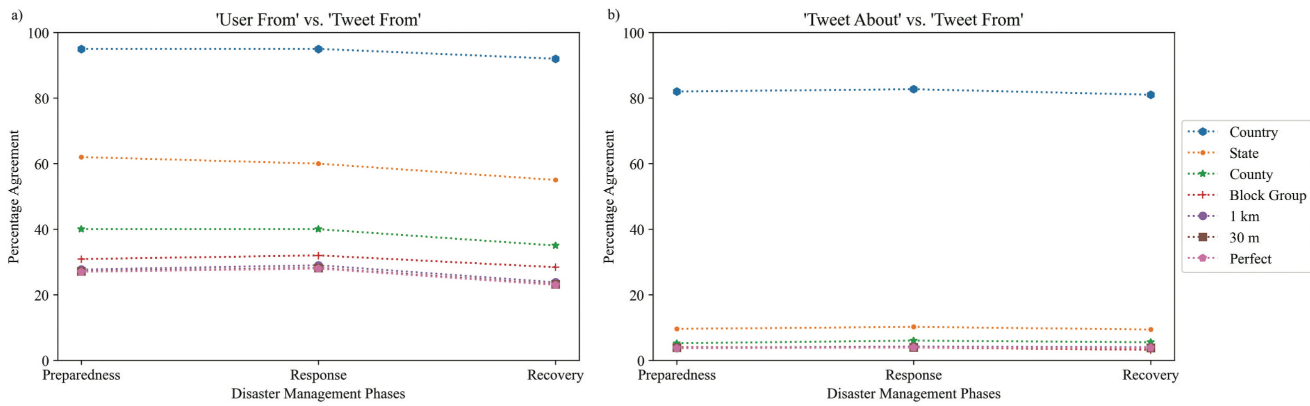

Figure 5. The percentage agreements of geolocating results by a) "*user from*" and b) "*tweet about*" compared with "*tweet from*" methods in different disaster phases.

0.000002, and a z-score of 13.257198. This resulted in a p-value less than 0.001, signifying a high level of significant spatial autocorrelation (clustered). Therefore, although the patterns may not seem significant from visual interpretation, statistical analysis reveals the existence of significant spatial patterns in the percentage agreement distribution at the county level.

4.2. Twitter-based indexes comparison

Figure 6 shows correlations between the county-level Ratio and Sentiment indexes based on "*tweet from*" with both indexes computed from "*user from*" and "*tweet about*" methods. Runnels County was left out as an outlier with a high Ratio value. This county has an abnormally high Ratio value because disaster responders in this county used social media for live rescue updates during Hurricane Harvey and posted many Harvey-related tweets geotagging the county. This produced a higher Harvey-related discussion intensity compared to other counties. As indicated in Figure 6a, the "*tweet from*" Ratio indexes are highly correlated with the "*user from*" Ratio indexes with a Pearson coefficient value of 0.92. However, the correlation coefficient between the "*tweet from*" and "*tweet about*" Ratio indexes is significantly lower at 0.50. The results reveal that employing the "*user from*" method to geolocate social media data for calculating the Ratio index when geotagged data are unavailable is a feasible approach when the analysis scale is at the county level. However, "*tweet from*" and "*tweet about*" strategies cannot be used interchangeably to derive the Ratio index. Discretion should be taken in choosing the method of obtaining the location for the purpose of ratio calculation. It is worth noting that the Ratio index value ranges are different in Figures 6a) and 6b). This is because the number of tweets that had both "*tweet from*" and "*user from*" is different from the number of tweets that had both "*tweet from*" and "*tweet about*" when tweets were aggregated into county-level to calculate the Ratio index. The second group is considerably lower in number, leading to a much smaller value range. The correlation analysis of the county-level Sentiment scores derived from tweets geolocated by different strategies uncovers a more significant discrepancy if the three methods are used interchangeably. In the majority of the cases both the "*user from*" and "*tweet about*" methods fail to yield Sentiment Index values similar to the results from the "*tweet from*" method with Pearson correlation coefficients of 0.12 and 0.01, respectively.

4.3. Misuses of geolocating strategies

This section depicts how rescue requests and user-reported damages that are posted in real-time on social media may be displaced when choosing incorrect geolocating strategies. As this investigation applied keyword-based rules to identify rescue-related messages on Twitter, both the tweets asking for rescue and providing help or information to people needing rescue were retrieved and included. When people request rescue for themselves and others or report local damages on social media, they usually provide their addresses or victims' locations in the text content instead of geotags (Zhou et al., 2022). Therefore, the "*tweet about*" addresses of those messages provide more vital geographical information about people needing help and are considered as reference locations. Figure 7 geographically visualizes the displacements of rescue-related and damage-related tweets as travel paths from "*tweet about*" (green) to "*tweet from*" (red) and "*user from*" (red) locations.

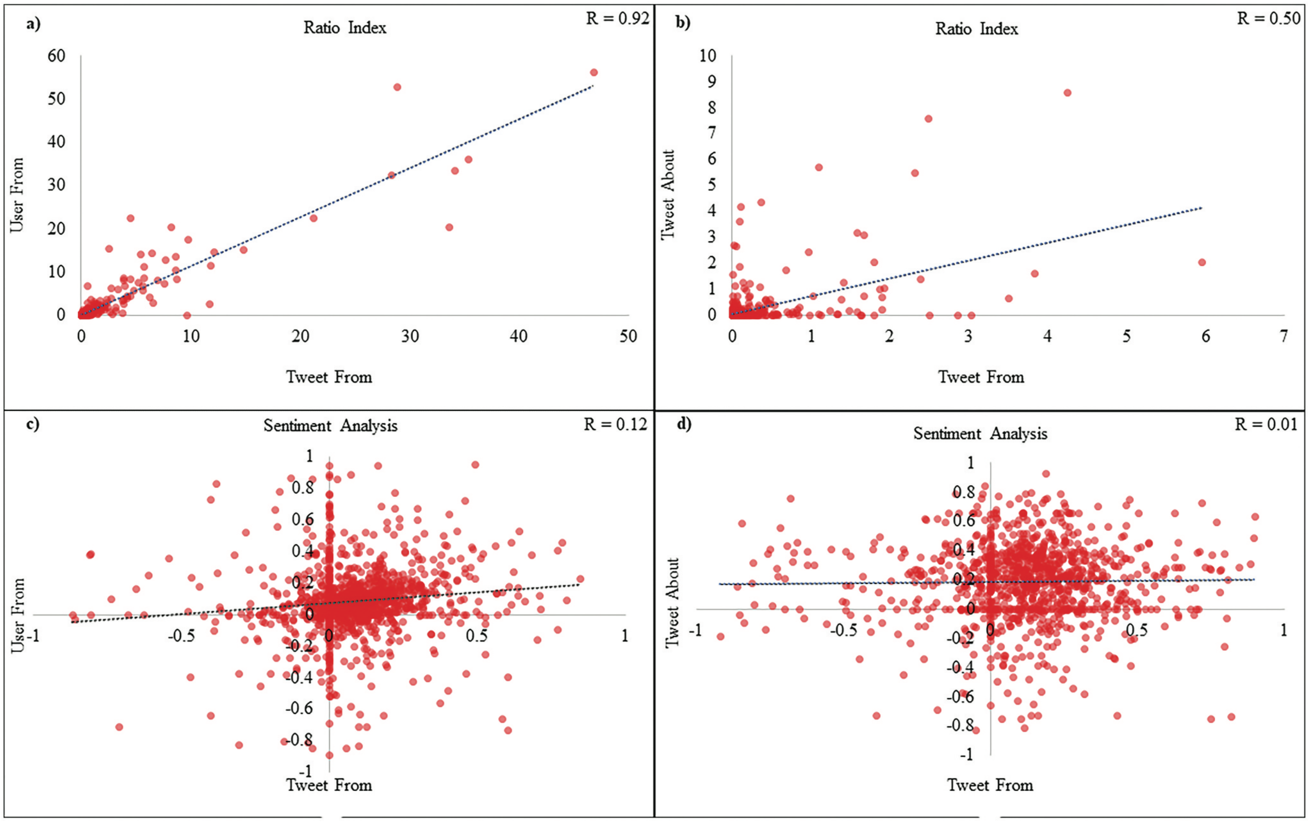

Figure 6. Correlations between the county-level Ratio and sentiment indexes based on "tweet from" with "user from" and "tweet about" methods.

The manual validation retrieved 1164 (94.3%) rescue-related tweets with locational information out of the 1234 tweets returned by the keyword filtration. Figure 7a) exemplifies one rescue request tweet with its "*tweet about*" location in Harris County but being geolocated to Los Angeles County and Guilford County by "*tweet from*" and "*user from*" attributes. Figure 7c) & e) illustrate all rescue-related tweets having Texas addresses mentioned in the content and their travel paths from "*tweet about*" coordinates to locations determined by the "*tweet from*" and "*user from*" methods, respectively. Most of the tweet-mentioned addresses were clustered in Harris County, where the city of Houston locates, followed by Aransas County (Rockport) and Nueces County (Corpus Christi), where Hurricane Harvey made its first and second land-falls. Among them, 9.87% were geolocated to the same county by all three methods. When comparing "*tweet from*" and "*tweet about*" addresses in the rescue-related messages, only 20.16% and 30.92% were geolocated to the same county and state. Meanwhile, 13.92% and 31.31% of tweets were spatially joined with the same county and state by the "*tweet about*" and "*user from*" approaches.

A total of 1033 (42.76%) out of 2416 tweets were manually validated as damage-related and contained geographic information in the content. Figure 7b) shows the travel path of one damage report located in Harris County based on its "*tweet about*" location but being displaced to Oklahoma County and Pottawatomie County in its "*tweet from*" and "*user from*" locations. Figure 7d) & e) display the travel paths of all damage report tweets from their Texas "*tweet about*" addresses to "*tweet from*" and "*user from*" locations, respectively. The three methods geolocated 11.03% of the damage-related messages in the same county. The agreements of geolocating damage reports from social media using the "*tweet from*" and "*user from*" methods compared with the "*tweet about*" results are 37.10% and 32.12% at the county level and 39.65% and 45.18% at the state level, higher than the similarities of rescue-related tweets geolocated by different algorithms.

These significant mismatches of geolocating rescue-related and damage-related tweets by different geolocating methods are concerning. The majority of these displaced messages were geolocated in a different county or state by the "*tweet from*" and "*user from*" approaches. This is because many rescue-related or damage-related messages were posted by people who lived outside of the affected areas seeking help for their friends or families or sharing

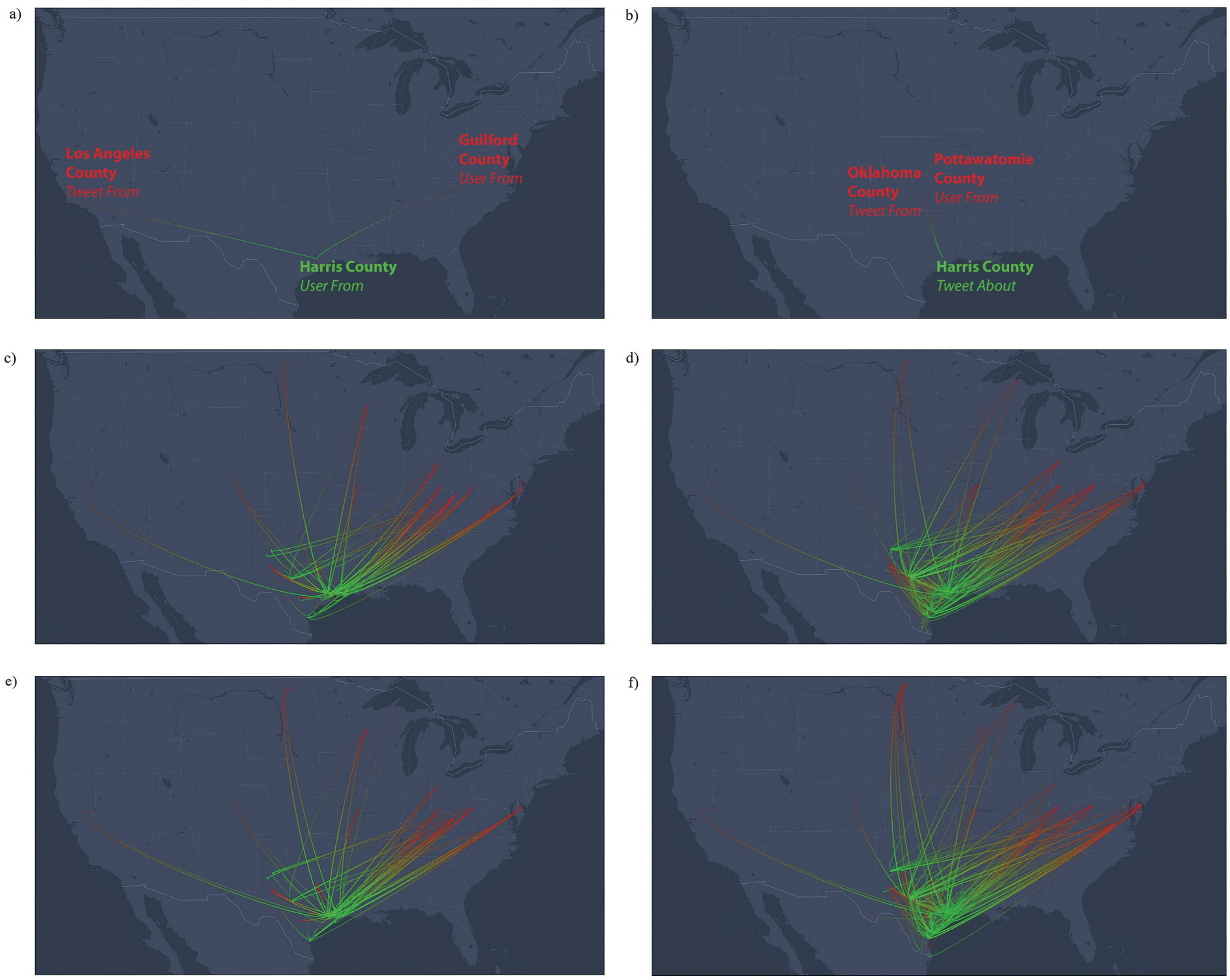

Figure 7. Travel paths of: a) a rescue tweet example; b) a damage tweet example; c) all rescue requests from "*tweet about*" (green) to "*tweet from*" (red) locations; d) all damage reports from "*tweet about*" addresses (green) to "*tweet from*" (red) locations; e) all rescue requests from "*tweet about*" addresses (green) to "*user from*" (red) locations; f) all damage reports from "*tweet about*" addresses (green) to "*user from*" (red) locations.

local information obtained from other resources at the time of Hurricane Harvey. Therefore, using these locations interchangeably during real-time disaster response will provide an erroneous value misguiding first responders in rapidly and accurately locating disaster victims or damaged infrastructures and sending help, thus creating more confusion in times of peril than good. The results highlight the need for a more sophisticated and inclusive mechanism for identifying appropriate geolocating strategies in location-based social media analytics. While our study has provided important insights into the geolocation uncertainty in disaster-related tweets, we acknowledge that our research could have been further enriched by comparing these results with location assessments of non-disaster-related tweets. Such an analysis could help discern if there are unique characteristics to disaster-related tweets.

5. Discussion and conclusion

This research aims to quantify and visualize the algorithmic uncertainties of geolocating social media data using different methods and attributes and their impacts on analyzing social media for disaster management. Using Twitter data during the 2017 Hurricane Harvey in the United States as an example, we compared the agreements of associating tweets with locations by the "*tweet about*", "*user from*," and "*tweet from*" methods through three analyses.

The first analysis evaluated the agreements of the three geolocating methods by geographical scales and disaster phases. Observations reveal that the three geolocating algorithms cannot be used interchangeably or concurrently in most analysis scales. The similarities between the "*user from*" and "*tweet* *from*" methods are overall higher than the agreements between the "*tweet about*" and "*tweet from*" approaches. The percentages of agreements are the highest at the country level (94.07% and 81.57%) and decrease rapidly as the geographical scale becomes finer. States in the affected areas present the highest level of agreement in geolocating results (74.64% in Texas at the state level) from the "*user from*" and "*tweet from*" methods compared to the non-affected regions. However, this phenomenon does not hold when analyzing the spatial patterns of consistencies of associating tweets with locations by "*tweet about*" and "*tweet from*" information (52.6% in California is the highest at the state level). The geolocating consistencies between the "*user from*" and "*tweet from*" results are highest in the preparedness and response phases and slightly lower in the recovery phase. On the other hand, the agreements of the geolocating results from the "*tweet about*" and "*tweet from*" strategies are the highest in the response phase and lowest in the recovery phase.

The second analysis computed two frequently used Twitter-based indexes based on tweets geolocated by the three approaches to reveal the disparities of deriving community-level public discussion intensity and sentiment toward hazardous events caused by geolocating uncertainties. Results show that the agreement between the Ratio index values calculated by tweets geolocated by "*user from*" and "*tweet from*" is plausible at the county level ($r = 0.92$), but the sentiment index values' correlations are insignificant among data generated by the three geolocating strategies. Finally, we quantified and visualized the displacement of rescue-related and damage-related messages on Twitter when choosing an irrational geolocating algorithm. It highlights the need to select appropriate geolocating strategies to meet the demand for different research and practice goals when processing and analyzing social media data.

While the higher level of agreement for broader administrative levels, such as "County" to "State" to "Country," is a known artifact of the geocoding process, our study's primary aim is not to reiterate this point. Rather, we focus on quantifying the uncertainty at these various administrative scales. This quantification provides researchers with a reference for understanding the potential limitations and allowances when dealing with uncertainties in geolocating social media data for disaster management purposes.

Despite the significant findings of this investigation, several limitations exist that necessitate further studies. First, this work is based on Twitter data during a single catastrophic natural hazard and focuses on Twitter use in the United States alone. Therefore, it may not necessarily reflect the algorithmic uncertainty in geolocating data obtained from different social media platforms, in other countries, or during different types of disasters. Future studies could also explore the relationship and variances in location assessments across disaster-related and non-disaster-related social media uses to enhance the scope of understanding of geolocation uncertainties in different contexts. However, the framework and analysis methods can be seamlessly applied in similar investigations and include more disaster case studies with diverse geographical and climatic contexts to give a broader picture of uncertainties in geolocating social media data. Although our study offers valuable insights with potentially broad applications for social media analysis, it is crucial to recognize the distinct characteristics inherent to each platform. As such, these unique attributes should be carefully considered when employing spatial analyses in different social media contexts.

Second, the disaster-, rescue-, and damage-related Twitter data were collected and filtered using keywords, which may potentially cause another type of algorithmic uncertainty. Although existing literature argues that keyword-based filtration is solid in selecting event-related social media messages (Zou et al., 2022), advances in natural language processing (NLP), e.g. the emerging pre-trained language models, have been demonstrated as effective alternatives yielding more reliable data classification and collection from social media platforms (Wang et al., 2020; Zhou et al., 2022). Future research can benefit from those NLP models in obtaining and processing social media accurately.

Third, the recent changes to Twitter API usage policies may limit the scope of future research. The data used in this study was collected in 2017. However, starting in February 2023, Twitter API has required a paid subscription for basic, pro, and enterprise access levels. As the volume and duration of data collection could be curtailed under the new paid model, there might be potential impacts on the availability of Twitter data for disaster-related research and management. Thus, it is crucial for the research community to consider these aspects when planning future studies using Twitter data for disaster management.

Lastly, this study does not examine the uncertainty brought in by the selection of data processing toolsets, i.e. toponym recognition ("locationtagger" Python Package) and toponym resolution (Google Geocoding API) tools. While packages like "locationtagger" significantly streamline the process of location extraction from tweets, they are not without limitations. Challenges include handling less known, newly named locations, as well as ambiguous location references. To further automate the process, it would be beneficial to continuously update and expand the geographic databases used by these packages. Moreover, enhancing their ability to handle misspellings and colloquial terms, as well as improving contextual analysis, could increase the accuracy of location identification, reducing the necessity for manual data cleaning. Additional work evaluating the performance differences among popular toponym recognition modules and toponym resolution tools in geolocating social media could unravel the algorithmic uncertainty of location-based social media mining for disaster and other applications from more dimensions.

Acknowledgments

We extend our sincere appreciation to the anonymous reviewers whose constructive comments have greatly improved the quality of this manuscript.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the NSF under Grant No. [1762600] (P.I.: Dr. Nina Lam) and [2318206].

ORCID

Debayan Mandal http://orcid.org/0000-0002-9529-830X
Joynal Abedin http://orcid.org/0000-0001-6203-0959
Bing Zhou http://orcid.org/0000-0003-1106-1370

Data availability statement

Data will be available upon request. It will be hosted in a private GitHub Repo: https://github.com/rohan-debayan/DeidentifiedHarveyTweets.git

References

Aulov, O., & Halem, M. (2012). Human sensor networks for improved modeling of natural disasters. *Proceedings of the IEEE*, *100*(10), 2812–2823. https://doi.org/10.1109/JPROC.2012.2195629

Backstrom, L., Sun, E., & Marlow, C. (2010). Find me if you can: Improving geographical prediction with social and spatial proximity. *Proceedings of the 19th International Conference on World Wide Web - WWW*, *10*, 61. https://doi.org/10.1145/1772690.1772698

Belcastro, L., Marozzo, F., Talia, D., Nappi, M., & Trunfio, P. (2021). Using social media for sub-event detection during disasters. *Journal of Big Data*, *8*(1), 79. https://doi.org/10.1186/s40537-021-00467-1

Carley, K. M., Malik, M., Landwehr, P. M., Pfeffer, J., & Kowalchuck, M. (2016). Crowd sourcing disaster management: The complex nature of twitter usage in Padang Indonesia. *Safety Science*, *90*, 48–61. https://doi.org/10.1016/j.ssci.2016.04.002

Cheng, Z., Caverlee, J., & Lee, K. (2010). You are where you tweet: A content-based approach to geo-locating twitter users. *Proceedings of the 19th ACM International Conference on Information and Knowledge Management - CIKM*, *10*, 759. https://doi.org/10.1145/1871437.1871535

Chen, X., Vorvoreanu, M., & Madhavan, K. P. C. (2014). Mining social media data for understanding students' learning experiences. *IEEE Transactions on Learning Technologies*, *7*(3), 246–259. https://doi.org/10.1109/tlt.2013.2296520

Cinelli, M., Cresci, S., Galeazzi, A., Quattrociocchi, W., Tesconi, M., & Murase, Y. (2020). The limited reach of fake news on twitter during 2019 European elections. *PLOS ONE*, *15*(6), e0234689. https://doi.org/10.1371/journal.pone.0234689

Costliest, U. S. (2018). *Tropical cyclones tables updated. National oceanic and atmospheric administration*. United States National Hurricane Center. https://www.nhc.noaa.gov/news/UpdatedCostliest.pdf

Earle, P., Guy, M., Buckmaster, R., Ostrum, C., Horvath, S., & Vaughan, A. (2010). OMG earthquake! Can Twitter improve earthquake response? *Seismological Research Letters*, *81*(2), 246–251. https://doi.org/10.1785/gssrl.81.2.246

Goldberg, D. W., Ballard, M., Boyd, J. H., Mullan, N., Garfield, C., Rosman, D., Ferrante, A. M., & Semmens, J. B. (2013). An evaluation framework for comparing geocoding systems. *International Journal of Health Geographics*, *12*(1), 50. https://doi.org/10.1186/1476-072X-12-50

Graham, M., Hale, S. A., & Gaffney, D. (2014). Where in the world are you? geolocation and language identification in Twitter. *The Professional Geographer*, *66*(4), 568–578. https://doi.org/10.1080/00330124.2014.907699

Havas, C., Resch, B., Francalanci, C., Pernici, B., Scalia, G., Fernandez-Marquez, J., Van Achte, T., & Wagemann, J. (2017). E2mC: Improving emergency management service practice through social media and crowdsourcing analysis in near real time. *Sensors*, *17*(12), 2766. https://doi.org/10.3390/s17122766. MDPI AG.

Hecht, B., Hong, L., Suh, B., & Chi, E. H. (2011). Tweets from Justin Bieber's heart: The dynamics of the location field in user profiles. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 237–246. https://doi.org/10.1145/1978942.1978976

Huang, L., Shi, P., Zhu, H., Guo, D., & Li, M. (2022). Early detection of emergency events from social media: A new text clustering approach. *Natural Hazards*, *111*(3), 851–875. https://doi.org/10.1007/s11069-021-05081-1

Hutto, C., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web & Social Media*, *8*(1), 216–225. https://doi.org/10.1609/icwsm.v8i1.14550

Hu, Y., & Wang, J. (2020a). *How do people describe locations during a natural disaster: An analysis of tweets from Hurricane Harvey*. https://doi.org/10.48550/ARXIV.2009.12914

Hu, Y., & Wang, R.-Q. (2020b). Understanding the removal of precise geotagging in tweets. *Nature Human Behaviour*, *4*(12), 1219–1221. https://doi.org/10.1038/s41562-020-00949-x

Jurdak, R., Zhao, K., Liu, J., AbouJaoude, M., Cameron, M., Newth, D., & Wu, Y. (2015). Understanding human mobility from Twitter. *PLOS ONE*, *10*(7), e0131469. https://doi.org/10.1371/journal.pone.0131469

Karimiziarani, M., Jafarzadegan, K., Abbaszadeh, P., Shao, W., & Moradkhani, H. (2021). Hazard risk awareness and disaster management: Extracting the information content of Twitter data. *Sustainable Cities and Society*, *77*, 103577. https://doi.org/10.1016/j.scs.2021.103577

Kryvasheyeu, Y., Chen, H., Obradovich, N., Moro, E., Van Hentenryck, P., Fowler, J., & Cebrian, M. (2016). Rapid assessment of disaster damage using social media activity. *Science Advances*, *2*(3). https://doi.org/10.1126/sciadv.1500779

Kumar, A., & Singh, J. P. (2018). Location reference identification from tweets during emergencies: A deep learning approach. *International Journal of Disaster Risk Reduction*, *33*, 365–375. https://doi.org/10.1016/j.ijdrr.2018.10.021

Kwan, M. P. (2016). Algorithmic geographies: Big data, algorithmic uncertainty, and the production of geographic knowledge. *Annals of the American Association of Geographers*, *106*(2), 274–282. https://doi.org/10.1080/00045608.2015.1117937

Lingad, J., Karimi, S., & Yin, J. (2013). Location extraction from disaster-related microblogs. *Proceedings of the 22nd International Conference on World Wide Web*, 1017–1020. https://doi.org/10.1145/2487788.2488108

Lin, B., Zou, L., Duffield, N., Mostafavi, A., Cai, H., Zhou, B., Tao, J., Yang, M., Mandal, D., & Abedin, J. (2022). Revealing the global linguistic and geographical disparities of public awareness to covid-19 outbreak through social media. *International Journal of Digital Earth*, *15*(1), 868–889. https://doi.org/10.1080/17538947.2022.2070677

Liu, X., Li, Q., Nourbakhsh, A., Fang, R., Thomas, M., Anderson, K., Kociuba, R., Vedder, M., Pomerville, S., Wudali, R., Martin, R., Duprey, J., Vachher, A., Keenan, W., & Shah, S. (2016). Reuters tracer: A large scale System of Detecting & verifying real-time news events from Twitter. *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, 207–216. https://doi.org/10.1145/2983323.2983363

Maas, P. (2019). Facebook disaster Maps: Aggregate insights for crisis response & recovery. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 3173–3173. https://doi.org/10.1145/3292500.3340412

McNeill, G., Bright, J., & Hale, S. A. (2017). Estimating local commuting patterns from geolocated Twitter data. *EPJ Data Science*, *6*(1), 24. https://doi.org/10.1140/epjds/s13688-017-0120-x

Mendoza, M., Poblete, B., & Valderrama, I. (2019). Nowcasting earthquake damages with Twitter. *EPJ Data Science*, *8*(1), 3. https://doi.org/10.1140/epjds/s13688-019-0181-0

Mihunov, V., Lam, N. S. N., Zou, L., Wang, Z., & Wang, K. (2020). Use of Twitter in disaster rescue: Lessons learned from hurricane harvey. *International Journal of Digital Earth*, *13*(12), 1454–1466. https://doi.org/10.1080/17538947.2020.1729879

Mooney, C. (2021, October 27). The science behind the U.S.'s strange hurricane 'drought' - and its sudden end. *The Washington Post*. Retrieved November 1, 2022, from https://www.washingtonpost.com/news/energy-environment/wp/2017/09/07/the-science-behind-the-u-s-s-strange-hurricane-drought-and-its-sudden-end/

Porter, J. (2019, June 19). *Twitter removes support for precise geotagging because no one uses it " the verge*. www.theverge.com/2019/6/19/18691174/twitter-location-tagging-geotagging-discontinued-removal

Premamayudu, B. (2019). Identification of natural disaster affected area precise location based on tweets. *International Journal of Innovative Technology and Exploring Engineering*, *8*(6), 119–123. https://doi.org/10.35940/ijitee

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*(4), 362–376. https://doi.org/10.1080/15230406.2017.1356242

Reynard, D., & Shirgaokar, M. (2019). Harnessing the power of machine learning: Can Twitter data be useful in guiding resource allocation decisions during a natural disaster? *Transportation Research, Part D: Transport & Environment*, *77*, 449–463. https://doi.org/10.1016/j.trd.2019.03.002

Rusho, M. A., Ahmed, M. A., & Sadri, A. M. (2021). Social media response and crisis communications in active shootings during COVID-19 pandemic. *Transportation Research Interdisciplinary Perspectives*, *11*, 100420. https://doi.org/10.1016/j.trip.2021.100420

Sachdeva, S., & McCaffrey, S. (2018). Using social media to predict air pollution during California wildfires. *Proceedings of the 9th International Conference on Social Media and Society*, 365–369. https://doi.org/10.1145/3217804.3217946

Skripnikov, A., Wagner, N., Shafer, J., Beck, M., Sherwood, E., & Burke, M. (2021). Using localized Twitter activity to assess harmful algal bloom impacts of karenia brevis in Florida, USA. *Harmful Algae*, *110*, 102118. https://doi.org/10.1016/j.hal.2021.102118

Soni, K. (2020). Locationtagger. *GitHub repository*, https://github.com/kaushiksoni10/locationtagger

Sutton, J., Palen, L., & Shklovski, I. (2008). "Backchannels on the front lines: Emergent uses of social media in the 2007 Southern California wildfires." *Proceedings of the 5th International ISCRAM Conference*, Washington, DC, USA.

Swift, J. N., Goldberg, D. W., & Wilson, J. P. (2008). *Geocoding best practices: Review of eight commonly used geocoding systems*. University of Southern California GIS Research Laboratory.

Tsao, S.-F., Chen, H., Tisseverasinghe, T., Yang, Y., Li, L., & Butt, Z. A. (2021). What social media told us in the time of COVID-19: A scoping review. *Lancet Digital Health*, *3*(3), e175–e194. https://doi.org/10.1016/S2589-7500(20)30315-0

Wang, J., Hu, Y., & Joseph, K. (2020). NeuroTPR: A neuro net toponym recognition model for extracting locations from social media messages. *Transactions in GIS*, *24*(3), 719–735. https://doi.org/10.1111/tgis.12627

Wang, K., Lam, N. S., Zou, L., & Mihunov, V. (2021). Twitter use in hurricane Isaac and its implications for disaster resilience. *ISPRS International Journal of Geo-Information*, *10*(3), 116. https://doi.org/10.3390/ijgi10030116

Wang, Y., Wang, T., Ye, X., Zhu, J., & Lee, J. (2015). Using social media for emergency response and Urban sustainability: A case study of the 2012 Beijing Rainstorm. *Sustainability*, *8*(1), 25. https://doi.org/10.3390/su8010025

Wang, S., Zhang, M., Huang, X., Hu, T., Li, Z., Sun, Q. C., & Liu, Y. (2022). Urban-regional disparities in mental health signals in Australia during the COVID-19 pandemic: A study via Twitter data and machine learning models. *Cambridge Journal of Regions, Economy & Society*, *15*(3), 663–682. https://doi.org/10.1093/cjres/rsac025

Wong, J., & Wong, N. (2021). The economics and accounting for COVID-19 wage subsidy and other government grants. *Pacific Accounting Review*, *33*(2), 199–211. https://doi.org/10.1108/PAR-10-2020-0189

Yin, J., Lampert, A., Cameron, M., Robinson, B., & Power, R. (2012). Using social media to enhance emergency situation awareness. *IEEE Intelligent Systems*, *27*(6), 52–59. https://doi.org/10.1109/MIS.2012.6

Yousefinaghani, S., Dara, R., Poljak, Z., Bernardo, T. M., & Sharif, S. (2019). The assessment of Twitter's potential for outbreak detection: Avian influenza case study. *Scientific Reports*, *9*(1), 18147. https://doi.org/10.1038/s41598-019-54388-4

Zahra, K., Ostermann, F. O., & Purves, R. S. (2017). Geographic variability of Twitter usage characteristics during disaster events. *Geo-Spatial Information Science*, *20*(3), 231–240. https://doi.org/10.1080/10095020.2017.1371903

Zhai, W., Peng, Z.-R., & Yuan, F. (2020). Examine the effects of neighborhood equity on disaster situational awareness: Harness machine learning and geotagged Twitter data. *International Journal of Disaster Risk Reduction*, *48*, 101611. https://doi.org/10.1016/j.ijdrr.2020.101611

Zhou, B., Zou, L., Mostafavi, A., Lin, B., Yang, M., Gharaibeh, N., Cai, H., Abedin, J., & Mandal, D. (2022). VictimFinder: Harvesting rescue requests in disaster response from social media with BERT. *Computers, Environment and Urban Systems*, *95*, 101824. https://doi.org/10.1016/j.compenvurbsys.2022.101824

Zou, L., Lam, N. S. N., Cai, H., & Qiang, Y. (2018). Mining Twitter data for improved understanding of disaster resilience. *Annals of the American Association of Geographers*, *108*(5), 1422–1441. https://doi.org/10.1080/24694452.2017.1421897

Zou, L., Lam, N. S. N., Shams, S., Cai, H., Meyer, M. A., Yang, S., Lee, K., Park, S.-J., & Reams, M. A. (2019). Social and geographical disparities in Twitter use during Hurricane Harvey. *International Journal of Digital Earth*, *12*(11), 1300–1318. https://doi.org/10.1080/17538947.2018.1545878

Zou, L., Liao, D., Lam, N. S. N., Meyer, M. A., Gharaibeh, N. G., Cai, H., Zhou, B., & Li, D. (2022). Social media for emergency rescue: An analysis of rescue requests on Twitter during Hurricane Harvey. *International Journal of Disaster Risk Reduction*, *85*, 103513. https://doi.org/10.1016/j.ijdrr.2022.103513