Assessing The Quality of Water Depth Derived From Volunteered Geographic Information (VGI) For Flood Monitoring. A Case Study of Hurricane Harvey In Harris County, Texas

Abdullatif Alyaqout (a.alyaqout@ku.edu.kw)
Kuwait University https://orcid.org/0000-0002-8341-3198
T. Edwin Chow
Texas State University San Marcos: Texas State University
Alexander Savelyev
Texas State University San Marcos: Texas State University

Research Article

Keywords: Risk, Flood depth, Crowdsourcing, Social media, Water surface elevation, Big data analytics

Posted Date: July 21st, 2021

DOI: https://doi.org/10.21203/rs.3.rs-644550/v1

Abstract

The primary objectives of this study are to 1) assess the quality of each volunteered geographic information (VGI) data modality (text, pictures, and videos), and 2) evaluate the quality of multiple VGI data sources, especially the multimedia that include pictures and videos, against synthesized water depth (WD) derived from remote sensing (RS) and authoritative data (e.g. stream gauges and depth grids). The availability of VGI, such as social media and crowdsourced data, empowered the researchers to monitor and model floods in near-real-time by integrating multi-sourced data available. Nevertheless, the quality of VGI sources and its reliability for flood monitoring (e.g. WD) is not well understood and validated by empirical data. Moreover, existing literature focuses mostly on text messages but not the multimedia nature of VGI. Therefore, this study measures the differences in synthesized WD from VGI modalities in terms of (1) spatial and (2) temporal variations, (3) against WD derived from RS, and (4) against authoritative data including (a) stream gauges and (b) depth grids. The results of the study show that there are significant differences in terms of spatial and temporal distribution of VGI modalities. Regarding VGI and RS comparison, the results show that there is a significant difference in WD between VGI and RS. In terms of VGI and authoritative data comparison, the analysis revealed that there is no significant difference in WD between VGI and stream gauges, while there is a significant difference between the depth grids and VGI.

Introduction

Natural hazards are considered a major source of catastrophic damages and losses both to the population and to the economy at a global scale. It has been estimated that hazards during the first 12 years in the 21st century caused up to US$1.7 trillion in losses to the economy and more than 1.4 million population casualties around the globe (Montz, Tobin, and Hagelman 2017). About 30% of the world's land area is flood-prone, which affects approximately 80% of the world's population (Dilley et al. 2005). In the U.S., floods are the costliest hazard in term of human lives, property damage, and economic losses (Strömberg 2007). About 40% of hazardous events occurring between 1900 and 2015 in the U.S. were major floods, which makes them the leading cause of natural disaster losses (United States Geological Survey 2016; Cigler 2017). The losses can surpass billions of dollars, where the cumulative costs of 16 weather events in the US in 2017 was estimated to exceed US$300 billion These losses can destroy many businesses as well. According to Federal Emergency Management Agency (FEMA 2015), about 40% of small and individual businesses do not reopen following a disaster. Two of the most destructive flood events recorded in the U.S. were Hurricane Katrina in 2005 and Superstorm Sandy in 2012, with damages costing more than US$16.3 billion and US$8.6 billion respectively (FEMA 2018). Urban areas are particularly subjected to flood risk as land cover change influences the magnitude and frequency of floods in terms of stormwater infiltration, watershed drainage, urban climate change, and stream channel alteration (Marsh and Grossa 2005).

Rapid flood monitoring is essential to disaster response and relief, emergency planning, and future mitigation. To monitor flood, there has been an emerging interest in using Volunteered Geographic Information (VGI) as a source of data to improve situational awareness and real-time support for emergency management (Smith et al. 2015). VGI is user-generated geographic content through crowdsourcing platforms (Goodchild 2007). In flood monitoring, Water Depth (WD) measurement is more commonly conducted by official authorities or agencies using Remote Sensing (RS), stream gauges, or Water Height Mark (WHM) (Schnebele et al. 2014). As a non-authoritative data source, VGI sources can be derived from crowdsourced projects and/or social media. Crowdsourcing is typically initiated, mobilized, and undertaken by volunteer(s) with specific vision towards how the data should be created,

collected and used. People can share their updated statuses as well as their geographic location, which comes in handy during disaster events (Smith et al. 2015). In a flood event, both social media and crowdsourced data can provide real time information on flood conditions and associated damage, which is an advantage over traditional methods (e.g. RS and stream gauges) that might take a long time for data collection and processing (Li et al. 2017). In a disaster when time is critical and human lives are at stake, all sources of relevant information about a flood and its potential harm should be considered and collected. Since traditional methods are limited in terms of the spatial (e.g. stream gauges limited spatial distribution) and temporal coverage (e.g. RS data may not be available during floods or cloud-cover prevent the water body extraction process) (Li et al. 2017), VGI sources can provide an alternative, and perhaps supplementary to conventional sources, insights to be used in flood monitoring. Although VGI can be acquired from multiple sources, it requires a lot of pre-processing and the data quality is uncertain. Therefore, it is important to evaluate the accuracy and representativeness of such data for flood monitoring. The primary objectives of this study are to assess the quality of each VGI data modality (text, pictures, and videos) and to evaluate the quality of multiple VGI sources against authoritative and non-authoritative data for WD mapping. Hence, the research questions pursued are as follows:

1. Are there any significant differences in kernel density of VGI related to WD across various data modalities?
2. Are there any significant differences in the temporal distribution of VGI related to WD across various data modalities?
3. Are there any significant differences in synthesized WD between VGI and RS?
4. Are there any significant differences in synthesized WD between VGI and:

a. Authoritative stream gauge data?

b. Authoritative modeled depth grids?

Background

In this section, we will examine previous studies using various data sources for flood analysis and WD mapping. These data sources include RS (Sect. 2.1) and various forms of VGI (Sect. 2.2).

2.1 Remote sensing and flood disasters

Multiple RS techniques were previously deployed to detect presence of water bodies using its spectral reflectance characteristics (e.g. Gregg and Casey 2004; Karaska et al. 2004). In general, these previous studies focus on RS-based flood analysis along two main directions: (1) developing and comparing water extraction indices using the spectral characteristics of RS images (Amarnath 2014; Feyisa et al. 2014) and (2) using spatial sharpening algorithms to enhance the spatial resolution of moderate spectral bands in RS data to extract water bodies in finer spatial scale for more accurate and detailed outputs (Du et al. 2016).

One of the developed indices, the Normalized Difference Surface Water Index (NDSWI), combined Short-Wave Infrared (SWIR), Near-Infrared (NIR), and green bands from Landsat5 TM to detect water bodies (Amarnath 2014), while another index, an Automated Water Extraction Index (AWEI), similarly used the spectral bands of the NDSWI with the addition of the blue band for water body detection using Landsat 5 TM (Feyisa et al. 2014). The developed indices were compared against existing indices, e.g. the Normalized Difference Vegetation Index (NDVI), the Normalized Difference Water Index (NDWI) (Amarnath 2014), and the Modified Normalized Difference Water

Index (MNDWI) and the Maximum Likelihood (ML) classifier (Feyisa et al. 2014). Compared with the use of NIR, the results using SWIR showed better capabilities for detecting water bodies (Amarnath 2014). The developed algorithms were able to distinguish between water and non-water pixels and provided higher accuracy in classifying edge pixels, a mix of land and water, than the other compared techniques (Feyisa et al. 2014). However, RS data is limited by the availability of cloud-free images to detect water bodies during unstable atmospheric condition when flood occurs and the temporal coverage of it may only provide a snapshot of the entire period of flooding. Also, detecting water bodies using these indices does not quantify the water depth required in flood analysis. Integrating the delineated water bodies with DEM is required to derive that.

Beside developing water indeces, band spatial sharpening algorithms –– the second direction in RS-based flood analysis outlined earlier ––is used to provide the users with an approach to enhance the spatial resolution of water bodies extraction when using moderate spatial resolution bands (e.g. SWIR). Despite the availability of multiple algorithms for spatial sharpening in the literature, the user should have high-level skills and knowledge to select and perform the suitable sharpening algorithm. To leverage the capabilities of Sentinel-2 for water body extraction, Du et al. (2016) used different pan-sharpening techniques to downscale the spatial resolution of the SWIR from 20 m to 10 m. The pan-sharpening techniques used were Principal Component Analysis (PCA), Intensity Hue Saturation (IHS), High Pass Filter (HPF), and the À Trous Wavelet Transform (ATWT). Two spectral water indices were used to extract water features including the NDWI and the MNDWI. The results show that applying pan-sharpening techniques enhanced the spatial resolution of extracted water bodies from Sentinel-2 images.

The work on enhancing water body extraction methods for floods analysis continues to grow. These efforts are likely to be complemented by the availability of VGI during disasters may provide supplementary information to leverage WD analysis and modeling, especially to overcome the limitations of RS data in terms of spatial and temporal coverage. Thus, assessing the quality of both sources (VGI and RS) and explore whether they share similarities or not is essential prior to using them for future scenarios.

2.2 VGI for flood monitoring

Most studies using VGI (e.g. Guan and Chen 2014; Li et al. 2017) focus on text, one modality of VGI, to extract and map WD after a flood event. Relatively fewer work emphases on the combination of multiple modalities (e.g. pictures and videos) for flood analysis and WD extraction (e.g. Schnebele et al. 2014; Fohringer et al. 2015). In this section, we will review the previous work on using text modality (Sect. 2.2.1), multimedia modalities (Sect. 2.2.2), and integration of multiple data modalities for flood analysis (Sect. 2.2.3). Finally, we examine how the results of using VGI sources for flood analysis are validated against authoritative data (Sect. 2.2.4).

2.2.1 Analyzing text messages for flood monitoring

Focusing mainly on the sheer volume and semantics of text, several studies attempt to quantify and model the spatial relationship between the kernel density of tweets with the hazardous area. Guan and Chen (2014) related disaster-related ratio (DRR) with storm surge and wind damage in four cities. They have examined DRR relationship with hurricane damage data in three spatial categories—coastline proximity, urban areas proximity, and storm surge and wind damage at county level. Similarly, Li et al. (2017) explored the correlation between flood possibility index (FPI) within flooded areas by using geotagged flood-related tweets and USGS water gauges to extract water height points (WHPs) from DEM. Both studies found that urban areas with large populations tended to have higher social media activity (Guan and Chen, 2014) and that people tended to tweet more about the flood when they were closer to the flooded areas (Li et al. 2017). In their study, Li et al. (2017) found that about 10% of

flood-related tweets are within the inundated areas identified by the USGS, which agrees with findings of Guan and Chen (2014) where there was an overall spatiotemporal positive correlation between the relevance of tweets (the number of tweets related to the event) and the damage from the hurricane.

However, the limited availability of text tweets in less populated areas to generate WHPs led the model to underestimate parts of the flooded areas (Li et al. 2017) and further research is needed regarding the spatiotemporal uncertainties (Guan and Chen 2014). This finding seems to hold true across natural hazards, including flooding (Li et al. 2017), storm surge and wind damage (Guan and Chen, 2014).

Besides the density of hazards-related tweets, recent studies also explored the use of qualitative classification (De Albuquerque et al. 2015), as well as semantic and sentiment analysis (Deng et al. 2016; Shalunts et al. 2014) for flood monitoring where the volume of relevant tweets increased within flooded areas proportionately to the temporal pattern of flood events. In a bi-lingual study focusing on both English and German, De Albuquerque et al. (2015) leveraged the usage of text information to identify on-topic tweets, and classified them to seven-coded themes (volunteer actions, media reports, traffic conditions, first-hand observations, official actions, infrastructure damage and other). Their findings were consistent with Li et al. (2017) and Guan and Chen (2014) where on-topic tweets are more likely to occur near affected (flooded) areas ($\leq$ 10 km away from the affected areas), in catchments with high relative water levels (+ 0.75m), as defined by the difference between daily maximum water level and average flood water level throughout the study period.

Other researchers aimed to analyze shared semantics during a disaster to evaluate the risk and damages both spatially (Deng et al. 2016) and temporally (Shalunts et al. 2014). During Typhoon Haiyan in 2013, Deng et al. (2016) used Weibo (a Twitter-equivalent in China) API to extract text information based on pre-defined keywords. They set forth an index-system model for situational awareness and damage assessment before, during, and after disasters through semantic analysis. Likewise, Shalunts et al. (2014) proposed a new tool (SentiSAIL) for sentiment analysis during the Central Europe floods in 2013. They obtained and classified the tweets into positive, negative, mixed, and neutral sentiment. After analyzing the spatiotemporal variation of the crowds sentiments, the degree of risk varies from one place to another before the hurricane made landfall (Deng et al. 2016) and the trend of positive, negative, and neutral sentiments was nearly consistent with the temporal pattern of the flood, with negative sentiment dominating (Shalunts et al. 2014).

2.2.2 Analyzing multimedia in flood monitoring

Compared to text, multimedia (e.g. pictures and videos) contains more graphical and contextual details regarding the geographic phenomena under investigation. Like text, multimedia forms can be geotagged or not. While it is straightforward to extract the location of a geotagged photo, users may also identify the surrounding landmarks or objects to infer the approximate location of the picture/video that does not have an explicit geotag. Until recently, the availability of multimedia in social media or crowdsourced projects remained an untapped resource for researchers on disaster analysis and modeling to harnest flood-related data.

Several studies explored the use of multimedia to extract and analyze flood information (Fohringer et al. 2015) and for damage assessment (Schnebele et al. 2014). To extract and analyze flood information, a step-by-step model (PostDistiller) was developed for flood information collection from social media (PostCrowler), storing (PostStorage), and visualization (PostExplorer) to estimate and map inundated areas (Fohringer et al. 2015). In extension to PostDistiller, Schnebele et al. (2014) combined authoritative data (e.g. FEMA storm surge) and fused non-authoritative data (e.g. aerial photographs and videos) for rapid assessment of road damage during floods.

When comparing against authoritative data, multimedia from VGI demonstrated a good agreement in general (Schnebele et al. 2014). For example, the accuracy of water level extracted from VGI could be within decimeter, a range desirable for rapid flood analysis (Fohringer et al., 2015). Nevertheless, extracting water information from the multimedia is a time-consuming process, especially when done manually, and the uncertainty of the location of the post could limit the flood results (Fohringer et al. 2015).

2.2.3 Integration between multiple data modalities

Most work on flood analysis using VGI relies solely on one data modality, whether text, pictures, or videos. We were only able to identify two studies that use multiple data modalities of VGI (Schnebele, Cervone, and Waters 2014) or integrated them with RS data (Cervone et al. 2015) for hazards analysis. In their study, Schnebele et al. (2014) combined Civil Air Patrol (CAP) aerial photos with YouTube videos in their approach. In a similar vein, Cervone et al. (2015) integrated RS imagery with social media data during hazard events for damage assessment. Cervone et al. (2015) used Twitter data to select suitable satellite images related to the floods by searching for ten tweets within a 100 km^2 related to the event as a threshold to identify hot spots of floods. Then, they integrate the social media data, including Twitter and Flickr, with satellite images by interpolating their kernel density to evaluate the damage and predict which roads were more likely to be damaged and impassable. The integration between multiple data helped filling the gaps of spatiotemporal data availability or coverage during disasters (Schnebele et al. 2014) and showed effictiveness of Twitter, as a source for identifying hotspots –– which was used as a proxy to identify road damage locations –– through relevant keywoords, in estimating road closures (Cervone et al. 2015). Despite the integration of multiple VGI forms in several studies, the scarcity of research in this field presents a need to explore the uncertainty of data integration between VGI and conventional geospatial data in flood analysis.

2.2.4 Validation against authoritative data

In flood modeling, authoritative data—those distributed by official authorities or agencies (e.g. stream gauge data from USGS or inundation maps from FEMA), provides reference data for validation to measure the accuracy of flood information derived from VGI. We have identified two studies that validated their results against authoritative data for quality assessment. The first of these studies shows an overall 83% match between the modeled inundation map from twitter and the official maps produced by the USGS (Li et al. 2017), while another study found good agreement among the non-authoritative data, the FEMA flooded extent and the CAP aerial photos (Schnebele et al. 2014). Assuming that the authoritative data are of higher quality, there was a concern mentioned by the literature regarding the uncertainty of the spatial accuracy of non-authoritative data. The concern involves localisation where geocoded observations are localized through pre-defined addresses or enabled through GPS signals for more precise locations (Poser and Dransch 2010). Thus, further validation of non-authoritative data is required.

Materials And Methods

3.1 Study area

In 2017, Hurricane Harvey made landfall along the Texas coast on August 25th and lasted for five days, until August 29th. It was one of the most destructive hurricanes in the history of the state, regarding population damage and economic losses, with 1.4 m of rain and 209.2 km/h winds at its peak. It was estimated that the Houston area experienced 1.3 m of rainfall, which is the largest amount recorded in a single storm. More than 780,000 Texans

were evacuated, and there was no drinking water available for 61 communities during the five days of the hurricane (FEMA 2017a). The hurricane left the state with 50 people dead and damage costs estimated to reach $180 billion (Fortune 2017).

In this study, the area investigated was Harris County, located within Houston Metropolitan Statistical Area (MSA) near the east coast of Texas, U.S. (Fig. 1). Harris County is the third-largest populated county in the U.S. with an estimated population of 4.6 million in 2016 (U.S. Census Bureau 2017). During the last decade, the county experienced massive natural disasters (Hwang and Lee 2017; Floodlist 2016).

3.2 Data Sources and Preprocessing

In this study, the VGI sources include social media and crowdsourced data. To get WD observations shared by the public, Twitter data was collected during August 25th to September 1st, 2017. The data was collected using a bounding box query that fully covered the region of the study and captured between 95–100% of all geolocated tweets originating in the area, as estimated through the analysis of the throttling estimates provided by the Twitter Streaming API at the time of data collection (https://developer.twitter.com/en/docs/tweets/filter-realtime/api-reference/post-statuses-filter).The tweets combined both text and/or multimedia information.

Crowdsourced data included geotagged pictures and videos. Geotagged pictures were collected from the National Alliance for Public Safety GIS (NAPSG) Harvey picture-sharing platform (NAPSG 2017). This platform allows the users to upload pictures associated with a geographic location regarding the disaster situation in the impacted areas. Geotagged videos, the second crowdsourced data, were obtained from Homeland Infrastructure Foundation-Level Data (HIFLD 2017).

This research also utilized the inundation maps and depth grids modeled by FEMA (2017b) as an authoritative data source. Furthermore, water level records collected by the stream gauges from the USGS National Water Information System (USGS 2018b) were used to validate VGI WD data. Satellite images were collected from the USGS EarthExplorer (USGS 2018a) for VGI WD validation. The post-disaster image was collected on September 1st, 2017 by Sentinel-2. The selected water index to extract and map the WD was the MNDWI:

$$MNDWI = \frac{\rho_{green} - \rho_{swir}}{\rho_{green} + \rho_{swir}} \quad (1)$$

where ρ is the reflectance value of Sentinel-2 band 3 (green) and band 11 (SWIR). To synthesize WD from the RS images, a 1.5 m DEM derived from lidar data (acquired in 2008) from TNRIS (2017) was used. Table (2) illustrates the data used and its sources.

Table 1
Data summary

Data	Details	Source
Water Depth	Text Pictures Videos	Twitter and crowdsourced (NAPSG 2017 and HIFLD 2017)
	Stream gauges	USGS 2018b
Flooded areas outline	Modeled depth grids	FEMA 2017b
Satellite images	Sentinel-2	USGS 2018a
DEM	1.5 m DEM derived from lidar	TNRIS 2017

Our Twitter dataset was filtered to only include messages originating from Houston and precisely matching one of the keywords from the Table 2 below. Our approach was inspired by Li et al. (2017), who used Harvey-related hashtags or keywords, including the name of the event (e.g. Harvey), multiple descriptions (e.g. floods, water, etc.) or the most common hashtags during the event (e.g. based on the trending ones suggested by hashtagify.me) to compile the list of hashtags and keywords used (Table 2). For temporal classification and analysis, the time of the tweets was adjusted from Coordinated Universal Time (UTC) to Central Daylight Time (CDT) UTC-5.

Table 2
List of hashtags and keywords used for relevant tweets classification

Keywords	Hashtags
usgs, tropical, emergency, high, height, heavy, close, flood, rescue, response, underwater, water, harvey, hurricane, safe, feet, ft, foot, inch, inches, drowned, submerged, overflow, rain, damage, storm, tornado, disaster	#HurricaneHarvey, #Harvey, #Flood, #Floods, #Storm, #Rain, #Damage, #Tropical, #Hurricane, #Water, #Disaster, #Emergency, #Underwater

3.3 Data Analysis

The overall research design is summarized in Fig. 2. The workflow starts with processing and classifying the non-authoritative data for the statistical analysis. VGI data modalities were tested against each other in terms of kernel density and temporal frequency to examine the differences or similarities. In addition, WD from VGI was synthesized to be tested against those from RS and authoritative data. By combining elevation data (i.e. DEM) with the RS image, WD was extracted to be tested against the synthesized WD from VGI. Non-authoritative data (VGI and RS) were first preprocessed to extract WD information. VGI WD for each data modality was compared against each other using repeated measures ANOVA –– for spatial comparison using kernel density –– and using Chi-squared test –– for temporal frequency comparison. In addition, WD from each data modality were used to synthesize the final VGI WD, using kriging, for comparison with authoritative data and RS. Regarding RS data, WD was extracted using the MNDWI and the 1.5 m DEM, and the output was used for the comparison with WD derived from VGI using paired t-test. The final analysis was comparing between VGI WD and authoritative data, including stream gauges and depth grids, using paired t-test.

In order to examine the spatial and temporal differences between VGI modalities (RQs 1 and 2), we started by sorting matching tweets into two groups—those with hyperlinks (multimedia) or those without (text only). Then, the

hyperlink group was categorized into two categories— pictures or videos. Afterwards, WD extraction techniques (described in Sect. 3.3.1) were applied to identify the Harvey-related tweets that contains relevant WD information content, for all VGI modalities, and eliminate those without. After extracting the final WD-relevant tweets, kernel density was derived for each modality for spatial comparison. For temporal analysis, Chi-squared test was used for the WD-relevant tweets on daily basis. A similar approach was used for the crowdsourced pictures and videos to extract WD information and for daily classification. Using the WD information extracted from VGI, spatial interpolation was applied for all the modalities on a daily basis to be used with RS and authoritative data comparisons.

To inspect the differences or similarities between synthesized WD from VGI and RS datasets (RQ3), the required spectral bands from the RS images (i.e. b3 and b11) were atmospherically-corrected and the multiple scenes were mosaicked and clipped to the boundaries of the study area. Once this process is done, the outputs are ready to be used in the WD extraction process, along with the 1.5 m DEM obtained from TNRIS (2017). The detailed steps of RS processing and WD extraction are described in Sect. 3.3.1. Finally, the synthesized WD from RS data was used for comparison with VGI.

Regarding RQ4, relevant WD extracted from VGI was used for comparison with (RQ4a) the USGS stream gauges and (RQ4b) FEMA depth grids. The collected stream gauge data included water level observations in a 15-minute interval during the selected time periods of the tweets. Then, the water records were used to generate a daily average water level. Afterwards, the daily averages were combined with VGI-synthesized WD values for the statistical test on a daily basis. This step is accomplished by using the interpolated WD from VGI (from the steps mentioned above) as a base layer and by using the location of gauges to extract the WD values. On the other hand, the depth grids were first converted from feet to meters, then they were clipped to the boundaries of the study area. Afterwards, the point location of WD-related VGI were used to extract WD values from the depth grids for statistical testing. This process is done based on the dates of the depth grids.

3.3.1 WD Extraction

The author started by reading the tweets matching the keywords in Table 2 and manually extracting the water level mentioned in them, either explicitly or implicitly (Fig. 3a and b, respectively).

For the pictures and videos in multimedia, the water level was extracted visually by inferring the WD from surrounding objects, such as trees, sidewalk curbs, buildings, or cars, etc. A similar approach was followed for the crowdsourced pictures and videos collected by a group of volunteers (HIFLD 2017; NAPSG 2017). An example of water level extraction from multimedia is demonstrated in Fig. 4a, which shows that the water is slightly covering the sidewalk curbs. Therefore, the water level was estimated to be 0.3 m for this picture. Some tweets might include multimedia and text content, combined in one tweet, indicating water level information. An example of such tweets is shown in Fig. 4b. Tweets with such content were treated as both text and multimedia as a water level information source from VGI. Table 3 lists the approximate height of common objects used for water extraction in this study.

Table 3
Example of physical objects with estimated height for multimedia water level extraction.

Object	Approximate height in meters
Traffic marks, e.g. stop sign	2–2.2
Traffic signals	5.5–6
Sidewalk curbs	0.2–0.3
Fire hydrants	0.5–0.7
Street bin	0.8–1
One story of a building	2.5–3

Upon the extraction of WD at geotagged tweets, kriging was used to interpolate a WD surface for each type of VGI (WD_{text}, $WD_{pictures}$, and WD_{videos}) and generate an overall WD of the VGI (WD_{VGI}). Kriging was deployed because of limited overlap between VGI modalities and authoritative data (e.g. stream gauges); this approach showed satisfactory results in an earlier study by Zrinji and Burn (1992). The overall VGI workflow is demonstrated in Figure 5.

To extract water bodies from RS data, MNDWI was used to delineate water features. Since band3 (green) have a higher spatial resolution (10 m) compared to band11 (SWIR) with 20 m, band3 was resampled to match the spatial resolution of band11, then the MNDWI was calculated. The output of the MNDWI was a raster surface with values ranging between − 1 and 1. To determine water pixels from non-water pixels, a threshold (> = 0.10) was selected, based on visual observation, and the results represent the detected water bodies. After extracting and delineating water bodies, the spatial resolution of the raster layer was resampled to match the DEM resolution (1.5 m) to calculate the WD at the pixel level (WD_i). Using the water bodies as zones, the maximum elevation (Max) for each zone was extracted from the DEM using zonal statistics. Then, the elevation value of each pixel from the DEM ($Elevation_i$) within the water body zone was subtracted from the Max ($Max - Elevation_i$) to calculate the WD_i within the water body zone using local statistics. The WD extraction and the overall RS workflow are illustrated in Fig. 6.

Results

4.1 VGI Summary

As mentioned, the VGI sources in this study included both Twitter and crowdsourcing. The total count of tweets for the entire period in the study area was 32,586, and the total Harvey-related tweets count was 9,578, which represents 29.4% of the total tweets. Figure 7 illustrates the summary of Harvey-related and non-related tweets a week after Hurricane Harvey. About 76.5% of Harvey-related tweets occurred between 25th and 29th, which represents the period the hurricane lasted in the study area.

The second step of tweets processing was classifying the tweets to either a text or multimedia (picture or video) type. Figure 8 summarizes the count of Harvey-related tweets for both categories. The majority of tweets were classified as a multimedia type compared to the count of text type.

Among the Harvey-related tweets, the total count of WD information, regardless of data modalities, was 434 observations (about 4.5%). Figure 9 illustrates the summary of tweets with water level information, regardless of the data modality, compared to Harvey-related tweets on a daily basis. The count of tweets with WD information increases during the mid-way of the hurricane, between August 27th and August 29th, and decrease afterwards. The majority of Harvey-related tweets do not include WD information. It's important to note that in some cases, multiple tweets from different users posted at various times shared the exact geographic location. Regarding crowdsourced data, a total of 152 crowdsourced multimedia was shared (1 for text, 17 for videos and 134 for pictures) within Harris County.

After classifying the tweets to text and multimedia (picture or video) and extracting WD from both tweets and crowdsourced data, Fig. 10 summarizes the variation in count of WD information extracted from the three VGI data modalities (text, picture, and video) on a daily basis. Out of the total 586 observations (for both tweets and crowdsourced), pictures had the most WD observations contribution with 350 observations representing 59.6% of the total WD observations, and text shared the least contribution with 30 observations representing 5% of VGI. After manual extraction of WD from crowdsourced data, the total observations with WD extracted was 152 (134 for pictures, 18 for videos, and one text). The geographic distribution of VGI observations is displayed on the map in Fig. 11.

The distribution of VGI observations, in general, is scattered across Harris County and tends to cluster around Downtown Houston (towards the west), within I-610 boundaries, and towards the northwest of Beltway 8 as well (Fig. 11). There were less VGI observations in the east relative to the west. Regarding the geographic distribution of VGI data modalities, pictures and videos had a similar pattern of the overall VGI points distribution (Fig. 12).

4.2 VGI Validation and Comparison

4.2.1 Kernel density differences or similarities among VGI data modalities

A kernel density surface was derived for each of the three VGI data modalities, and each pixel was converted to point to extract the kernel density value from the three layers for further statistical analysis. The spatial distribution of each modality, represented by the kernel density, [by the null hypothesis] is assumed to be not significantly different from each other. A normality test showed that the kernel densities were not normally distributed at $p < 0.001$. The resulting Friedman test ($\chi^2 = 6{,}842{,}186$; $p < 0.001$; $n = 5{,}113{,}697$) revealed a significant difference in spatial distribution across VGI data modalities. Using Kendall's W, the effect size was 0.67. Therefore, the null hypothesis of RQ 1b that there was no significant difference among the three VGI modalities in terms of kernel density was rejected. A non-parametric Wilcoxon post-hoc test showed that there was a significant difference among all pairs of data modalities, including: (1) text and pictures ($Z = -1{,}873.760$; $p < 0.001$), (2) text and videos ($Z = -1{,}704.997$; $p < 0.001$), and (3) pictures and videos ($Z = -1{,}317.366$; $p < 0.001$).

4.2.2 Temporal frequency differences or similarities among VGI data modalities

A chi-square test was conducted to examine the differences in the temporal pattern of the VGI data modalities throughout the study time frame. Table 4 includes the current and expected distribution of observations. It was found that there is a significant difference between the current and the expected temporal distribution of the data

(χ^2 = 88.14; df = 14; $p < 0.001$). Therefore, the null hypothesis of RQ 1c that there was no significant difference among the three VGI modalities in terms of temporal currency was rejected.

Table 4
Observed and expected distribution of VGI observations for the chi-square test.

Observed (Expected) Distribution									
Day	**8–25**	**8–26**	**8–27**	**8–28**	**8–29**	**8–30**	**8–31**	**9 – 1**	**Total**
Picture	1(2.4)	11(21.9)	103(121.2)	47(53.2)	44(51.4)	33(28.1)	93(59.1)	18(14.7)	350
Text	0(0.2)	2(2.8)	13(10.4)	3(5.6)	8(4.4)	1(2.4)	2(5.1)	1(1.2)	30
Video	3(1.4)	22(12.3)	87(71.4)	39(31.3)	34(30.2)	13(16.5)	4(35.8)	4(8.1)	206
Total	4	35	203	89	86	47	99	23	586

4.3 VGI and RS validation

After extracting WD from the RS image (RS WD) and interpolating the VGI observations using kriging (VGI WD), the RS WD raster layer was converted to points. Then, the points were used to extract the interpolated WD values from the VGI WD for comparison (Fig. 13). The VGI WD included observations on August 31st and September 1st, due to the lack of VGI observations on September 1st. Using the interpolated WD surfaces (n = 37,967,601), a paired *t*-test was applied to examine any differences between WD derived from RS and VGI synthesized WD. The results showed that there was a significant difference between VGI and RS WD (t = 264.232; df = 38,067,599; $p < 0.001$). Therefore, the null hypothesis of RQ 2 that there was no significant difference between VGI WD and RS WD was rejected.

4.4 VGI and authoritative data comparison

4.4.1 VGI and USGS stream gauges

The USGS stream gauges scatter across the study area (Fig. 14). The total count of gauges within the county was 55, and eight of the total gauges does not have readings for the entire study period. The daily average water level was calculated for each gauge and was used for statistical analysis. After that, a table with both WD values, VGI and stream gauge, for the entire days was used to perform the normality test of and the paired *t*-test. The analysis did not include August 25th since only four VGI observations were available. The normality test results showed that the differences in WD between the gauges and VGI were normally distributed. Therefore, a paired *t*-test was conducted and the results showed no statistically significantly differences (t = 1.657; df = 278; p = 0.098). Therefore, we failed to reject the null hypothesis of RQ 3a that there was no significant difference in WD between VGI and stream gauges.

4.4.2 VGI and FEMA Modeled Depth Grids

The modeled depth grids from FEMA were available for five days including 27th, 28th, 29th, 30th, and September 1st (Fig. 15). The maximum extent of the inundated areas was during the 30th and September 1st (Fig. 15d and e).

The VGI points were used to extract the WD from FEMA grids for statistical analysis on a daily basis. The normality test of the difference in WD between VGI and the depth grids for the entire days showed a significant difference ($p < 0.001$). The Mann Whitney test showed significant difference (n = 294, $p < 0.001$). Therefore, the null

hypothesis of RQ 3b that there was no significant difference in WD between VGI and FEMA depth grids was rejected.

Discussion

The summary of VGI processing showed that the count of tweets increases up until the mid-way of the event, which can be observed between 27th and 29th of August for both total and relevant tweets with high WD observations, where the maximum was 5 m, 3 m, and 4.8 m during the 27th, 28th, and 29th respectively. Also, there was a weak positive correlation between the average WD of VGI and the gauges on daily basis ($r = 0.23$) (Fig. 16). In addition, 29.4% of all the tweets were classified as relevant, and 76.5% of these relevant tweets occurred during 25th to 29th of August when the hurricane lasted, which indicates that majority of relevant tweets were more available during the time of the disaster relative to pre- and post-disaster.

5.1 Research Questions

Regarding the first and second research questions, there was a significant difference in terms of spatial and temporal distribution of VGI modalities ($p < 0.001$ for both comparisons). This could be associated with the limited availability of text observations (5%) compared with the multimedia (about 95%), which was demonstrated by Li et al. (2017). This finding suggests that disaster studies focusing on text might represent only a small fraction and hence potentially induced sampling bias in their investigation Thus, the limited sample size of text messages could reduce the quality of the assessment outputs and influences the decisions made upon such outputs. When conducting a damage assessment or rapid flood analysis, the user should consider that text modality requires additional text analysis (semantic and sentiment) to extract damage information or flood conditions.

The results of Wilcoxon post-hoc test indicated that people tend to share multimedia content regarding their conditions and situation rather than describing it through text messaging. Since picture and video modalities had no significant differences in terms of frequency ($p = 0.557$) and the robust availability of multimedia during disasters, it should be taken under consideration to include VGI multimedia in big data studies because of its large sampling availability and its better context, compared with text. Moreover, the large availability of multimedia could provide a broader insight for the user in terms of rapid assessment of flood damage or risk with more spatial and temporal coverage. On the other hand, extracting flood-related information from multimedia would require image interpretation and digital processing, which can be more time-consuming and labor-intensive and present practical barriers to acute needs of emergency response.

Regarding the kernel density, the results of Friedman test showed that all VGI data modalities have varying spatial extent. This could be related to the limited sample size and distribution of text observations, where it was closer to downtown Houston and limited towards to the fringe of Harris County. On the other hand, pictures and videos scattered across the study area with more coverage in space and time than text (Fig. 17). In addition, WD visualized in multimedia provides better spatial context because they could be captured at locations with varying distances from the actual flooded area. For example, a multimedia could be taken on the third floor of a building showing a flood at the parking lot in front of the building, while other multimedia could show WD up to the sidewalk across the street in a small neighborhood with a short distance from the actual location where the user was standing to take the picture or the video. Without tedious description of the spatial context, however, text information is often assumed to indicate WD at the location of geotagged tweet. Also, zooming could be a factor influencing the distance of measured WD point from the actual multimedia location. Although pictures and videos

had a visually similar distribution, the post-hoc test showed a significant difference between both data modalities. It could be related with the difference in sample size of pictures (350) compared with videos (206) and the broader spatial distribution of pictures where it may influence the kernel density outputs, compared with the videos (Fig. 18).

Another reason to explain the spatial variation of VGI could be associated with the spatial pattern of the digital divide, where population subgroups may have varying access to digital information and communication technology (ICT) (Riggins and Dewan 2005). Based on the digital divide index (DDI) calculated by Gallardo (2018), the DDI score is composed of two main components including infrastructure/adoption and socioeconomic characteristics. The DDI score ranged between 0–100 where high DDI score indicates high existence of a digital divide and vice versa. Overall, there was a weak to low negative correlation coefficient between the DDI and the count of VGI points at each tract in the study area with − 0.09, -0.20, and − 0.16 for text, pictures, and videos respectively. Therefore, digital divide did not show significant influence on the spatial distribution of VGI. Additionally, most VGI observations occurred in areas with DDI less than 50 (moderate to low digital divide), and DDI between 45 and 50 exhibited the most frequency of VGI observations (Fig. 19). This finding supports the correlation results above where digital divide did not show significant impact on the VGI data.

The temporal distribution of the data also varied significantly according to the chi-square test (Table 3). Once again, the count of text observations might play a role in this variation. Furthermore, there is a possibility that users in some areas might have experienced technical issues with sharing tweets in real-time, such as power outage or temporarily limited internet access, that prevented them from sharing their situation during the event. According to power outage reports between August 27th and September 1st, about 71.2% of VGI observations were in areas with up to 5% of power outage, while 22.8% of the observations were in areas between 5% and 20% of power outage, which might suggest an inverse relationship between the count of tweets and the percentage of power outage in the study area (Fig. 20). A Friedman test was conducted, and the results revealed a significant difference between the count of VGI observations and the class of power outage with $p < 0.001$, indicating that power outage might be related to the VGI frequency across data modalities.

The overall findings of the VGI data modalities analysis suggests that the limited count of text observations influence its contribution in flood analysis when used as a single source of WD input. In a previous study, limited counts of text observations along with uneven spatial coverage could possibly result in underestimated flood extent (Li et al. 2017). However, it is recommended to integrate VGI with traditional flood mapping methods, despite the limitation of using it by its own (Rosser, Leibovici, and Jackson 2017).

To examine the quality of WD derived from VGI and RS (i.e. the second research question), the results showed a significant difference between the two WD datasets. It is important to note that the VGI surface used was interpolated, which could have “smoothed” the two datasets and narrowed any significant differences. For example, the interpolation surface has varying WD values ranging between 0.4–1.8 m, while VGI had a maximum WD of 3 m. Because the only post-disaster RS image available was on September 1st, VGI observations on that day were sampled and aggregated for interpolation, and locations where multiple tweets coincide (e.g. 3m along with varying WDs) were averaged for interpolation. One complication to interpret this finding, however, is the relatively coarse spatial resolution of the RS data being used (20 m after resampling). It is possible to take advantage of pan-sharpening methods to enhance the detection of water pixels more accurately at fine details (Du et al. 2016). However, cloud cover limits the availability of the usable RS image and the ability to detect water leading to possible underestimation of water body (Schnebele et al. 2014; Li et al. 2017).

The third research question examined the validation of WD from VGI against authoritative data sources of USGS stream gauges and FEMA depth grids. Using the stream gauges as references, the analysis showed that there was no significant difference in WD between VGI and the gauges. In light of water information available at fine temporal resolution for both datasets, this finding of indifferences between the data sources is indicative of the quality of VGI. In addition, the close proximity between the geographic locations of stream gauge and VGI points could also attribute to such agreement.

The second validation was against FEMA depth grids. The analysis showed a significant difference between the two datasets. Despite this finding, one uncertainty in this study was the manual extraction method from VGI, where WD may be subjected to the geographic reference (e.g. height of curbside, hydrant) being used (Table 4). On the other hand, the modeled depth grids from FEMA simulated at a given time would have higher internal consistency. In addition, the spatial variation between FEMA depth grid and VGI might also influenced the comparison between both datasets. The depth grids from flood simulation were mainly modeled and calibrated at stream gauges, whereas majority of the VGI observations were closer to the urban landscapes and more are found within residential areas (Fig. 15). As mentioned, the low spatial accuracy of some VGI points could jeopardize the examination of any significant differences between the two datasets. Nevertheless, the findings of no significant difference between VGI and USGS stream gauge data but its use in modeled WD grid suggest a myriad effect of different forces in the propagation of uncertainties in flood modeling.

5.2 Limitations

Most of the matching tweets did not include WD information –– only 4.5% of all Harvey-related tweets. It could be associated with the selected hashtags and keywords that might filter out tweets that could be counted as relevant, or with the way users might share their experience with the hurricane by using words might not be explicitly related to the event. One of the reasons explaining the large count of relevant tweets is the participation and engagement many Harvey-related hashtags were promoted by individuals and different governmental agencies as a tool for information dissemination regarding the event (Smith et al. 2015). Nevertheless, these promotion efforts did not target the solicitation of WD information from the public. With reference to established protocols deploying specific hashtags collect structured VGI useful for disaster response (Starbird and Stamberger 2010), it is possible to improve the quantity and quality of relevant tweets.

Other reasons for the limited count of tweets with relevant WD information could be attributed to the overwhelming use of a popular hashtag (e.g. #HurricaneHarvey) on irrelevant content, missing or broken link, or the overuse of a redundant post (e.g. retweet) by multiple users (Fig. 21). These factors influence the time and effort required to collect, preprocess, classify, and extract relevant information regarding the disaster event from VGI sources. A fifth possible source of limited water observations count is overlapping of the geotagged tweets. For example, 44 tweets with water observations from different users shared the same latitude and longitude location on different days. This could be related to the attachment of the same location by the users when posting the tweet. Multiple users might prefer to tag the generic name of a place (e.g. Houston) in a tweet, which would assign the same coordinates for all the tweets with such tags (Steiger et al. 2015).

In overall, the study limitations were associated with the preprocessing and extraction of WD from the VGI and RS data availability and analysis. The manual extraction from relevant VGI is time-consuming and might be difficult when no obvious physical marks are available, or when they are distant from the position of the user when he/she took the picture or the video. Moreover, the retweet of an image by multiple users makes it difficult to find the

duplicate posts with a large amount of data to be observed and may lead to overestimating the count of relevant tweets. In addition, part of the relevant tweets, according to the hashtags and keywords, were sharing non-relevant information (e.g. family pictures or food pictures), and a group of tweets shared the same geographic location, which influences the analysis. Besides, some multimedia had poor quality due to insufficient lighting to observe the level of water.

The availability of RS data with respect to reasonable cloud coverage during or right after the event was limited. In a storm or hurricane, the weather condition suitable for remote sensing is often limited. In addition, the spatial resolution (20 m) influenced the results of delineated water bodies. The processing, extraction, and interpolation of VGI to WD was time consuming and computationally intensive for a county or regional project.

Conclusion

This study focuses on assessing the quality of VGI by comparing each VGI data modality (text, picture, and video) against each other and to traditional sources of flood data, including remote sensing and authoritative data sources. The VGI data modalities showed significant differences in terms of spatial, and temporal properties. This finding is significant because existing literature typically focuses only on text data and overlooked its implications and limitations. By comparing the WD derived from VGI with RS images, this study shows that limitation of temporal coverage and the existence of clouds in RS data are possible reasons for the disagreement between both datasets. For future studies, using radar data might increase the probability of leveraging the RS data sources and overcome the cloud problem in the scenes. Finally, the validation of VGI-derived WD against authoritative data showed insignificant difference with the USGS gauges, indicative to the quality of VGI.

This study provides empirical evidence that VGI in all modalities can provide WD as an input for flood monitoring and modeling. In addition of stream gauge data, the potential inclusion of VGI may increase the number of observations spatially and temporally and could be used to calibrate and leverage the outputs of physical flood models, e.g. HEC-RAS. The addition of VGI can also be useful for risk assessment and emergency response to floods. Specifically, the availability of WD in urban areas, where the coverage of stream gauges is limited, could increase the quality of the flood modeling outputs. However, the modeled depth grids from FEMA did not agree with the synthesized VGI. For future analysis, WD extraction from multimedia should be assessed and calibrated to increase the accuracy of the interpolated surfaces to be compared with the modeled depth grids. Future research should consider the following recommendations:

1. Utilization of pictures and videos in VGI as a supplementary source to other data in flood analysis.
2. Measuring and improving the spatial and temporal accuracy of the geotagged tweets for better spatial and temporal analysis.
3. Future research can consider comparing multiple spatial interpolation methods and suggest suitable technique(s) for a given set VGI sample for flood analysis.
4. Conduct damage assessment from VGI and compare it with authoritative damage assessment data for rapid damage assessment analysis.
5. Inclusion of different socioeconomic status to enhance and leverage the DDI used in this study for better understanding of the impact of the digital divide on sharing relevant information during disasters at multiple locations.

6. For rapid flood analysis, developing a new method to automate WD extraction from VGI using pattern recognition, change detection, and semantic and sentiment analysis should be taken under consideration in the future.

Declarations

Conflicts of interest The authors declare that they have no conflicts of interest to this work.

Funding No funds, grants, or other support was received.

References

1. Abbott M, Bathurst J, Cunge J et al (1986) An introduction to the European Hydrological System—Systeme Hydrologique Europeen,"SHE", 2: Structure of a physically-based, distributed modelling system. J Hydrol 87:61–77
2. Ahearn EA (2004) Regression equations for estimating flood flows for the 2-, 10-, 25-, 50-, 100-, and 500-year recurrence intervals in Connecticut
3. Amarnath G (2014) An algorithm for rapid flood inundation mapping from optical data using a reflectance differencing technique. J Flood Risk Manag 7:239–250. https://doi.org/10.1111/jfr3.12045
4. Arnaud P, Lavabre J (2002) Coupled rainfall model and discharge model for flood frequency estimation. Water Resour Res 38:1–11. https://doi.org/10.1029/2001WR000474
5. Brunner GW (2002) Hec-ras (river analysis system). ASCE, pp 3782–3787
6. Casas A, Benito G, Thorndycraft V, Rico M (2006) The topographic data source of digital terrain models as a key element in the accuracy of hydraulic flood modelling. Earth Surf Process Landf 31:444–456. https://doi.org/10.1002/esp.1278
7. Cervone G, Sava E, Huang Q et al (2015) Using Twitter for tasking remote-sensing data collection and damage assessment: 2013 Boulder flood case study. Int J Remote Sens 37:100–124. https://doi.org/10.1080/01431161.2015.1117684
8. Cigler BA (2017) US Floods: The Necessity of Mitigation. State Local Gov Rev 0160323X17731890
9. de Albuquerque JP, Herfort B, Brenning A, Zipf A (2015) A geographic approach for combining social media and authoritative data towards identifying useful information for disaster management. Int J Geogr Inf Sci 29:667–689. https://doi.org/10.1080/13658816.2014.996567
10. Deng Q, Liu Y, Zhang H et al (2016) A new crowdsourcing model to assess disaster using microblog data in typhoon Haiyan. Nat Hazards 84:1241–1256
11. Dilley M, Chen RS, Deichmann U et al (2005) Natural disaster hotspots: a global risk analysis. World Bank Publications
12. Du Y, Zhang Y, Ling F et al (2016) Water bodies' mapping from Sentinel-2 imagery with modified normalized difference water index at 10-m spatial resolution produced by sharpening the SWIR band. Remote Sens 8:354
13. Environmental Systems Research Institute (ESRI) (2018) Hurricane_Harvey_Power_Outages_by_Zip_Code. http://www.arcgis.com/home/item.html?id=6d251bb5e48e492c90ebac123011ffdb. Accessed 22 April 2018
14. Federal Emergency Management Agency (FEMA) (2015) Business Infographics. https://www.fema.gov/media-library/assets/documents/108451. Accessed 17 February 2018

15. FEMA (2017a) Historic Disaster Response to Hurricane Harvey in Texas. https://www.fema.gov/news-release/2017/09/22/historic-disaster-response-hurricane-harvey-texas. Accessed 3 December 2017
16. FEMA (2017b) National Disasters. https://data.femadata.com/NationalDisasters/HurricaneHarvey/. Accessed 10 April 2018
17. FEMA (2018) Significant Flood Events. https://www.fema.gov/significant-flood-events. Accessed 18 February 2018
18. Feyisa GL, Meilby H, Fensholt R, Proud SR (2014) Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery. Remote Sens Environ 140:23–35
19. Floodlist (2016) Houston Floods – President Declares Major Disaster. http://floodlist.com/america/usa/houston-floods-president-declares-major-disaster. Accessed 3 Dec 2017
20. Fohringer J, Dransch D, Kreibich H, Schröter K (2015) Social media as an information source for rapid flood inundation mapping. Nat Hazards Earth Syst Sci 15:2725–2738. https://doi.org/10.5194/nhess-15-2725-2015
21. Fortune (2017) Hurricane Harvey Damages Could Cost up to $180 Billion. http://fortune.com/2017/09/03/hurricane-harvey-damages-cost/ . Accessed 23 February 2018
22. Gallardo R (2017) 2015 Digital Divide Index - Mississippi State University, Extension Service Intelligent Community Institute. http://ici.msucares.com/sites/ici.msucares.com/files/2015_ddi.pdf. Accessed 22 April 2018
23. Goodchild MF (2007) Citizens as sensors: the world of volunteered geography. GeoJournal 69:211–221
24. Gregg WW, Casey NW (2004) Global and regional evaluation of the SeaWiFS chlorophyll data set. Remote Sens Environ 93:463–479
25. Guan X, Chen C (2014) Using social media data to understand and assess disasters. Nat Hazards 74:837–850. https://doi.org/10.1007/s11069-014-1217-1
26. Guinot V, Gourbesville P (2003) Calibration of physically based models: back to basics? J Hydroinformatics 5:233–244
27. Hese S, Heyer T (2016) Earth observation data based rapid flood-extent modelling for tsunami-devastated coastal areas. Int J Appl Earth Obs Geoinformation 46:63–83. https://doi.org/10.1016/j.jag.2015.11.005
28. Homeland Infrastructure Foundation-Level Data (HIFLD) (2017) Hurricane Harvey Video Resources. https://respond-harvey-geoplatform.opendata.arcgis.com/datasets/hurricane-harvey-video-resources?geometry=-106.583%2C27.242%2C-85.676%2C30.607. Accessed 4 Dec 2017
29. Howe J (2006) The rise of crowdsourcing. Wired Mag 14:1–4
30. Hwang SN, Lee S-W (2017) Does Environmental Risk Affect Human Migration Behavior? J Geogr Inf Syst 9:493–504
31. Karaska MA, Huguenin RL, Beacham JL et al (2004) AVIRIS measurements of chlorophyll, suspended minerals, dissolved organic carbon, and turbidity in the Neuse River, North Carolina. Photogramm Eng Remote Sens 70:125–133
32. Kia MB, Pirasteh S, Pradhan B et al (2012) An artificial neural network model for flood simulation using GIS: Johor River Basin, Malaysia. Environ Earth Sci 67:251–264
33. Larkin TJ, Bomar GW (1983) Climatic atlas of Texas. Texas Department of Water Resources
34. Li Z, Wang C, Emrich CT, Guo D (2017) A novel approach to leveraging social media for rapid flood mapping: a case study of the 2015 South Carolina floods. Cartogr Geogr Inf Sci 1–14.

https://doi.org/10.1080/15230406.2016.1271356
35. Mandel B, Culotta A, Boulahanis J et al (2012) A demographic analysis of online sentiment during hurricane irene. Association for Computational Linguistics, pp 27–36
36. Marsh WM, Grossa J Jr (2005) Environmental geography: science, land use, and earth systems. John Wiley and Sons, New York
37. Mississippi State University (2018) DDI Digital Divide Index. http://ici.msucares.com/resources/ddi. Accessed 22 April 2018
38. Montz BE, Tobin GA, Hagelman RR, Tobin GA (2017) Natural hazards: explanation and integration, Second Edition. The Guilford Press, New York
39. National Alliance for Public Safety GIS (NAPSG) (2017) Hurricane Harvey Photos. https://napsg.maps.arcgis.com/apps/StoryMapCrowdsource/index.html?appid=b6ef838e4d26489e8f62102639dc3d91. Accessed 4 Dec 2017
40. Poser K, Dransch D (2010) Volunteered geographic information for disaster management with application to rapid flood damage estimation. Geomatica 64:89–98
41. Pradhan B (2010) Flood susceptible mapping and risk area delineation using logistic regression, GIS and remote sensing. J Spat Hydrol 9
42. Riggins FJ, Dewan S (2005) The digital divide: Current and future research directions. J Assoc Inf Syst 6:13
43. Rosser JF, Leibovici DG, Jackson MJ (2017) Rapid flood inundation mapping using social media, remote sensing and topographic data. Nat Hazards 87:103–120
44. Schnebele E, Cervone G, Waters N (2014) Road assessment after flood events using non-authoritative data. Nat Hazards Earth Syst Sci 14:1007–1015. https://doi.org/10.5194/nhess-14-1007-2014
45. Shalunts G, Backfried G, Prinz P (2014) Sentiment analysis of German social media data for natural disasters
46. 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. J Flood Risk Manag
47. Starbird K, Stamberger J (2010) Tweak the tweet: Leveraging microblogging proliferation with a prescriptive syntax to support citizen reporting. In: Proceedings of the 7th International ISCRAM Conference. ISCRAM Seattle, WA, pp 1–5
48. Steiger E, Westerholt R, Resch B, Zipf A (2015) Twitter as an indicator for whereabouts of people? Correlating Twitter with UK census data. Comput Environ Urban Syst 54:255–265
49. Strömberg D (2007) Natural disasters, economic development, and humanitarian aid. J Econ Perspect 21:199–222
50. Texas Natural Resources Information System (TNRIS) (2017) Data Search and Download. https://tnris.org/data-download/#!/quad/Rose%20Hill. Accessed 4 Dec 2017
51. U.S. Census Bureau (2017) Quick Facts - Houston city, Texas. https://www.census.gov/quickfacts/fact/table/houstoncitytexas/PST045216. Accessed 3 Dec 2017
52. U.S. Geological Survey (USGS) (2016) Flood inundation mapping (FIM) program. https://water.usgs.gov/osw/flood_inundation/. Accessed 11 Oct 2017
53. USGS (2018a) EarthExplorer. https://earthexplorer.usgs.gov/. Accessed 21 February 2018
54. USGS (2018b) National Water Information System. https://nwis.waterdata.usgs.gov/nwis/uv. Accessed 10 April 2018

55. Zrinji Z, Burn DH (1992) Application of Kriging to surface water level estimation. Can J Civ Eng 19:181–185

Figures

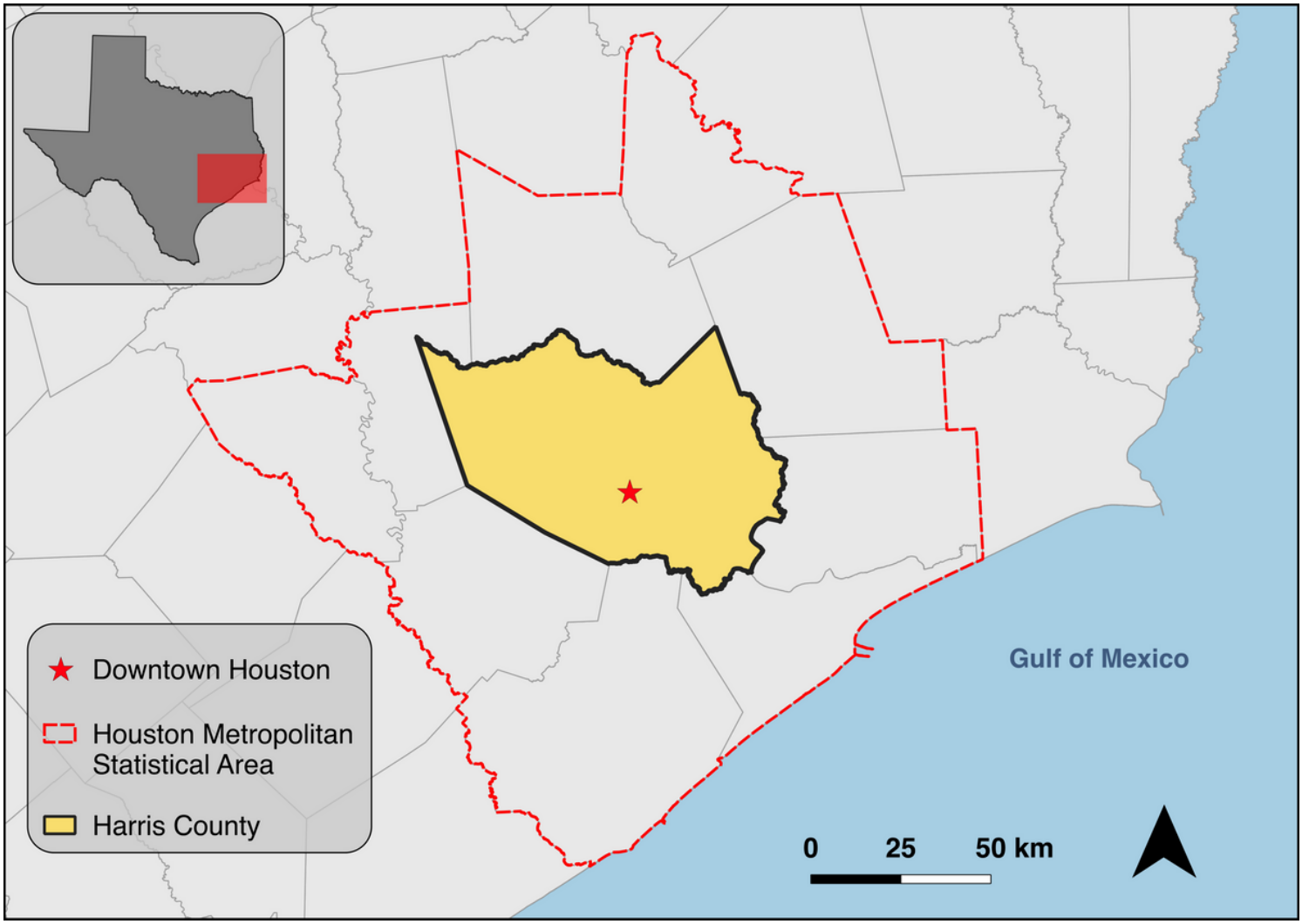

Figure 1

Study area. (data source: Texas Natural Resources Information System (TNRIS) 2017 and US Census Bureau 2012)

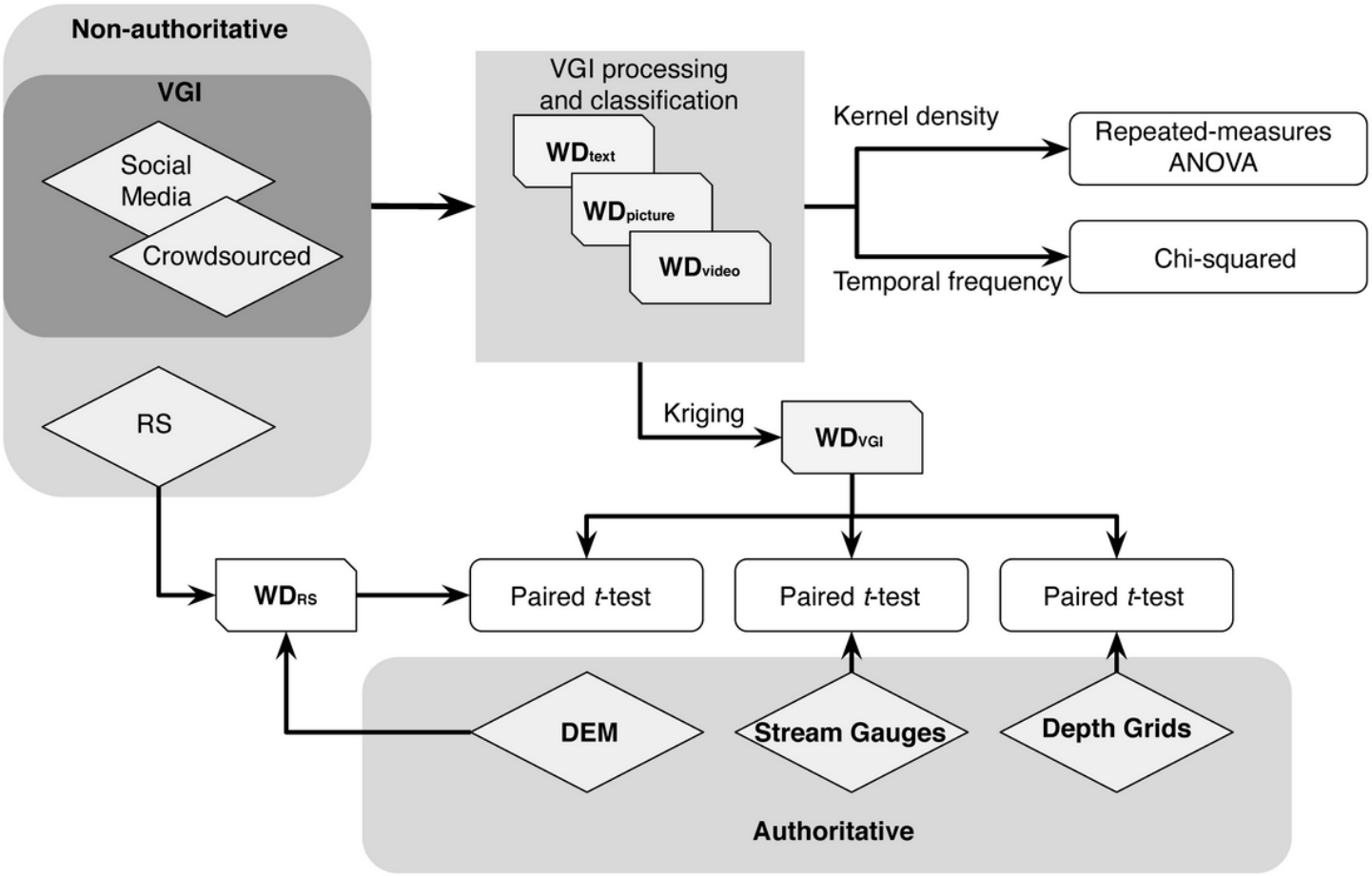

Figure 2

The overall research design.

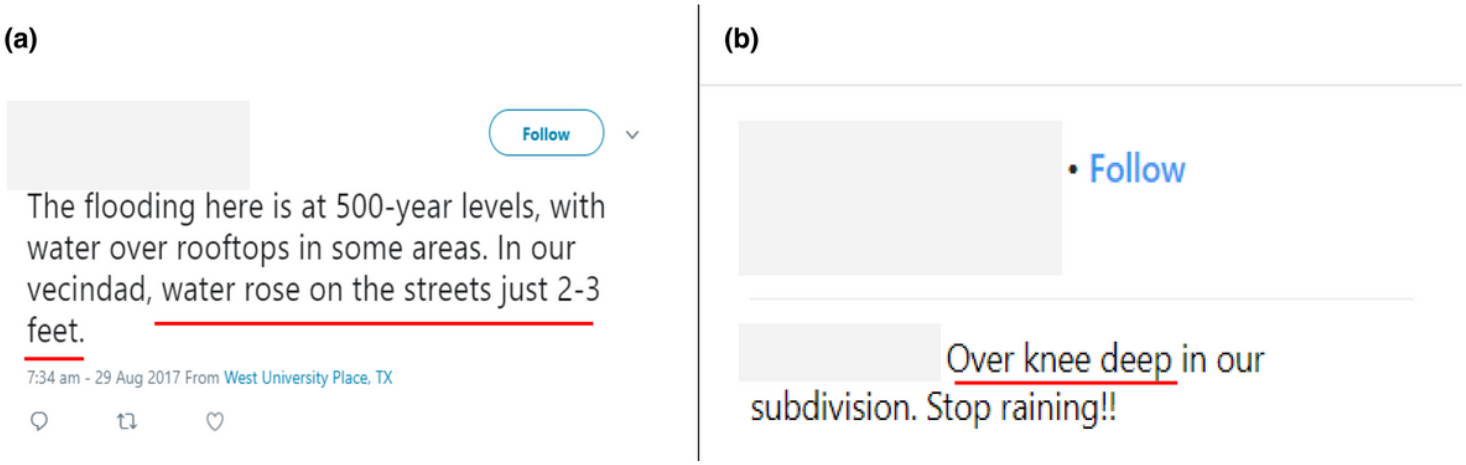

Figure 3

An example of explicit (a) and implicit (b) indication of water level indicated in the text of a tweet.

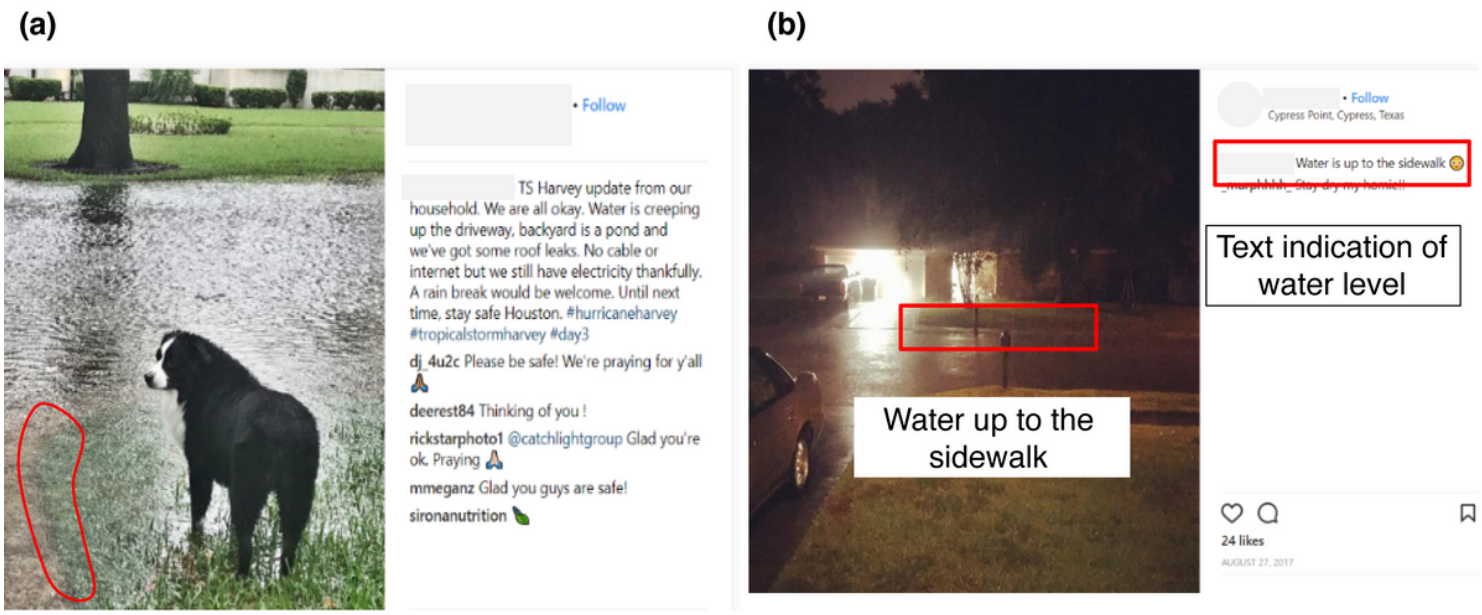

Figure 4

An example of extracting water level from a picture using the surrounding physical objects (a). The red line shows the edge between the road and the sidewalk. While (b) shows an example of water level information included in two VGI data modalities (text and picture).

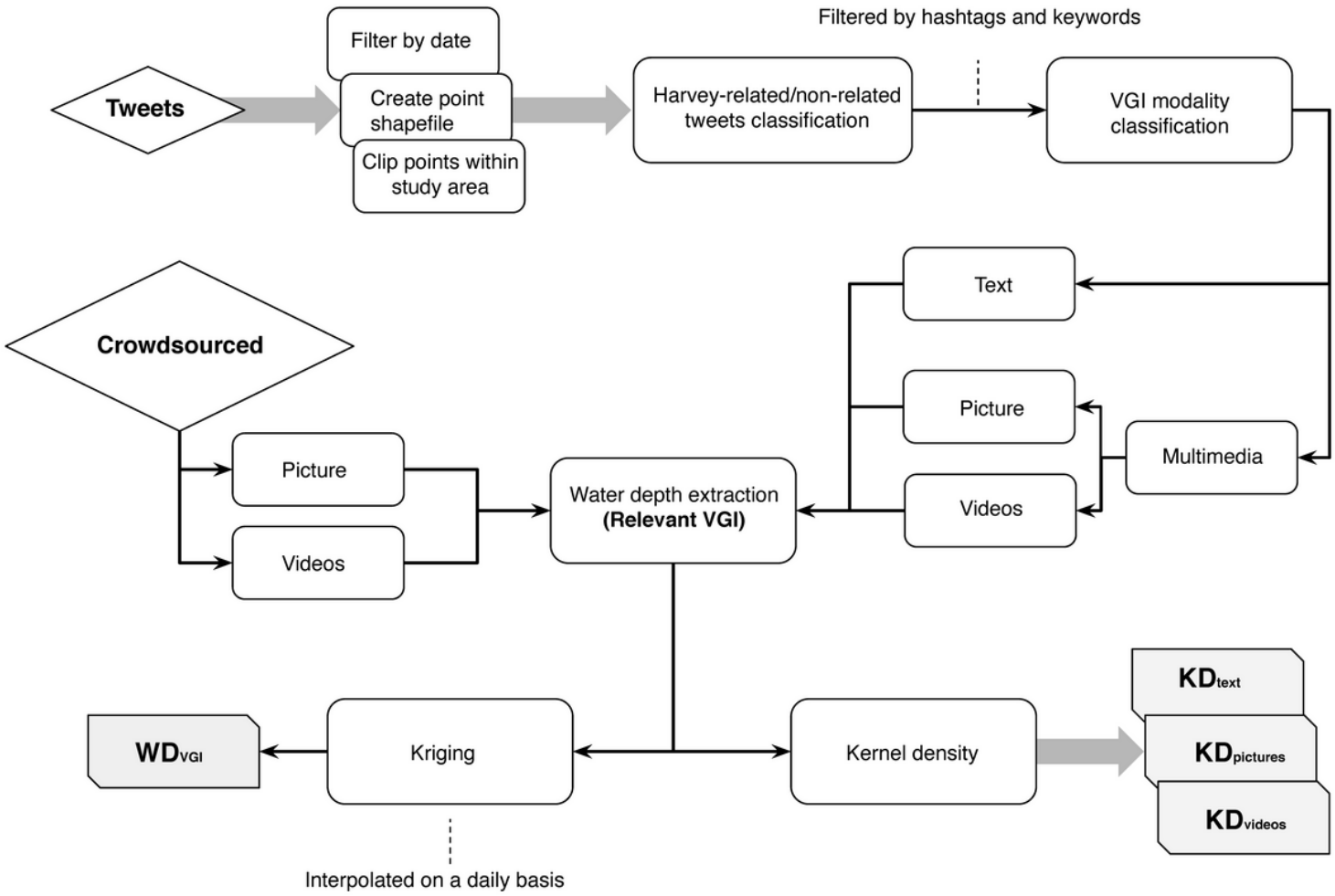

Figure 5

Twitter and crowdsourced data processing and WD extraction.

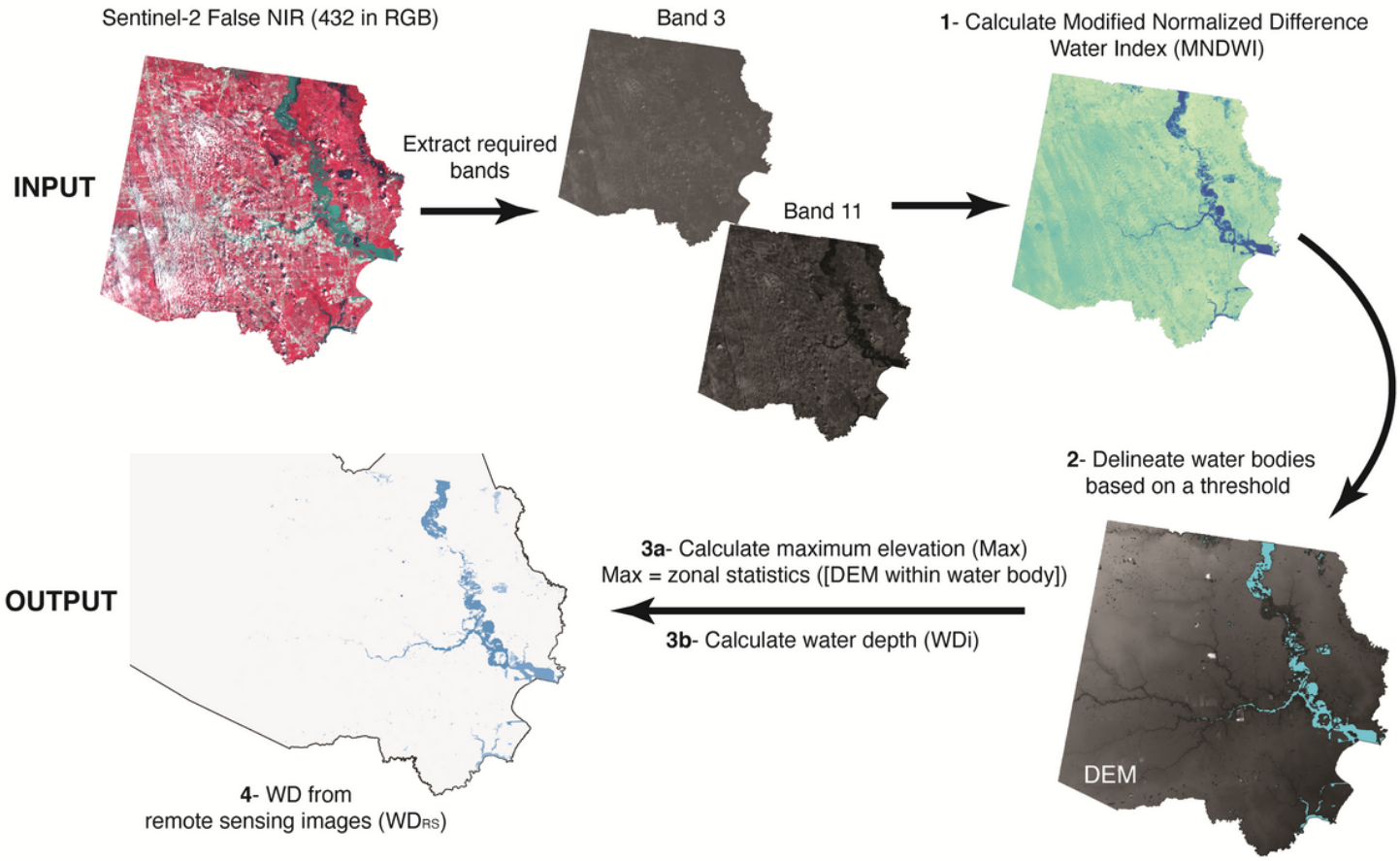

Figure 6

RS processing and WD extraction. Refer to the text for further details.

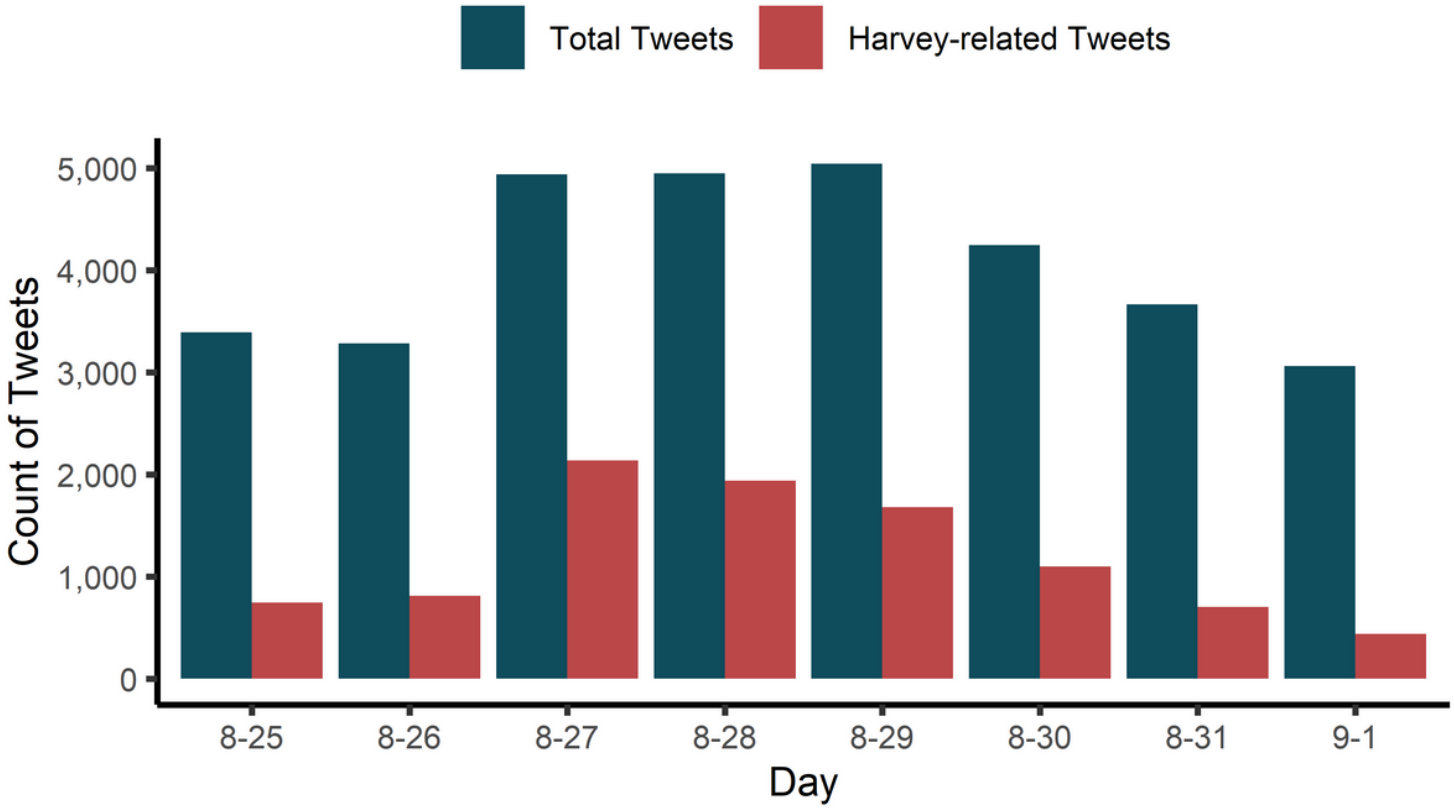

Figure 7

Summary of Harvey-related and non-related tweets in Harris County.

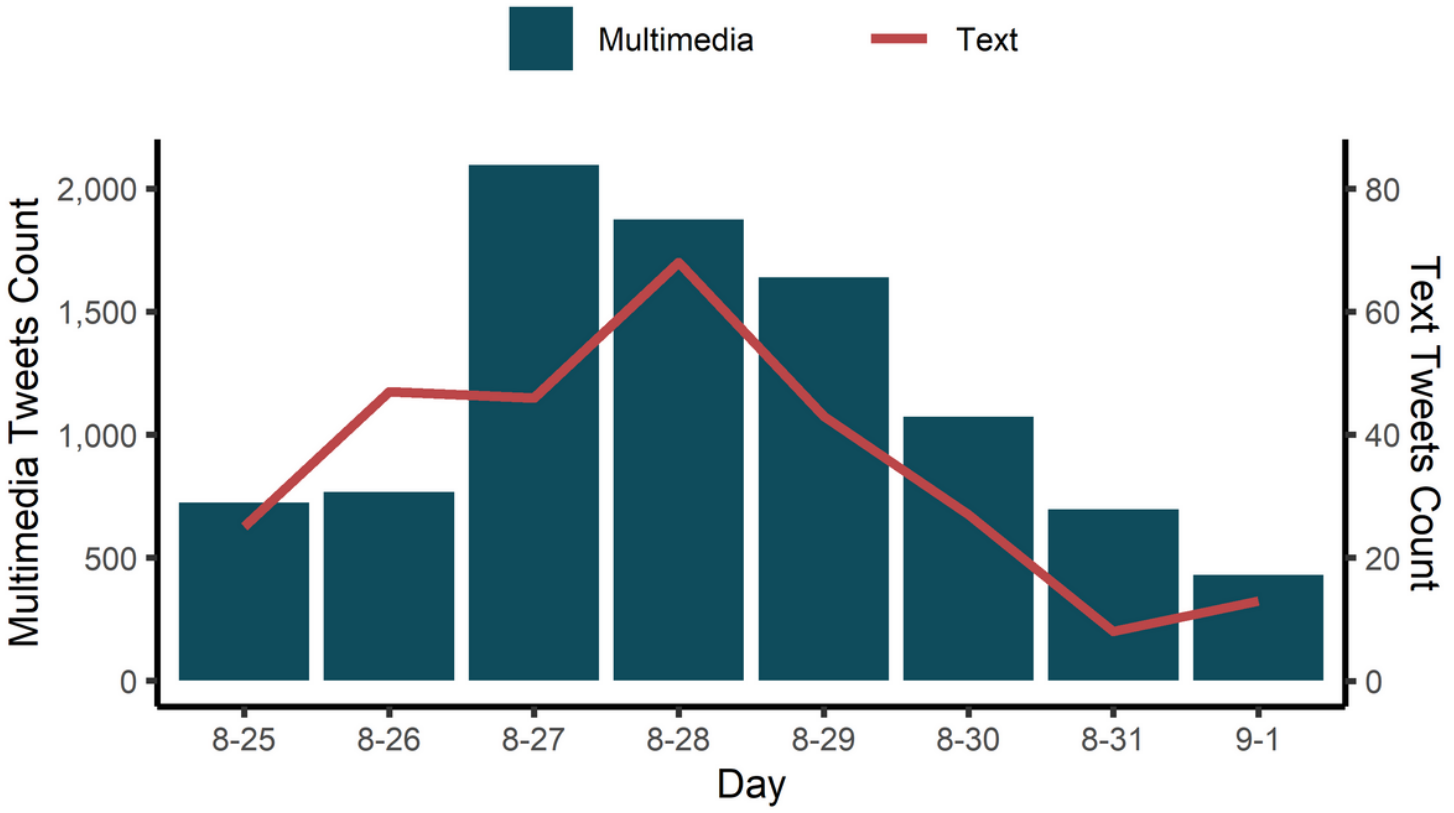

Figure 8

Summary of Harvey-related tweets for text and multimedia.

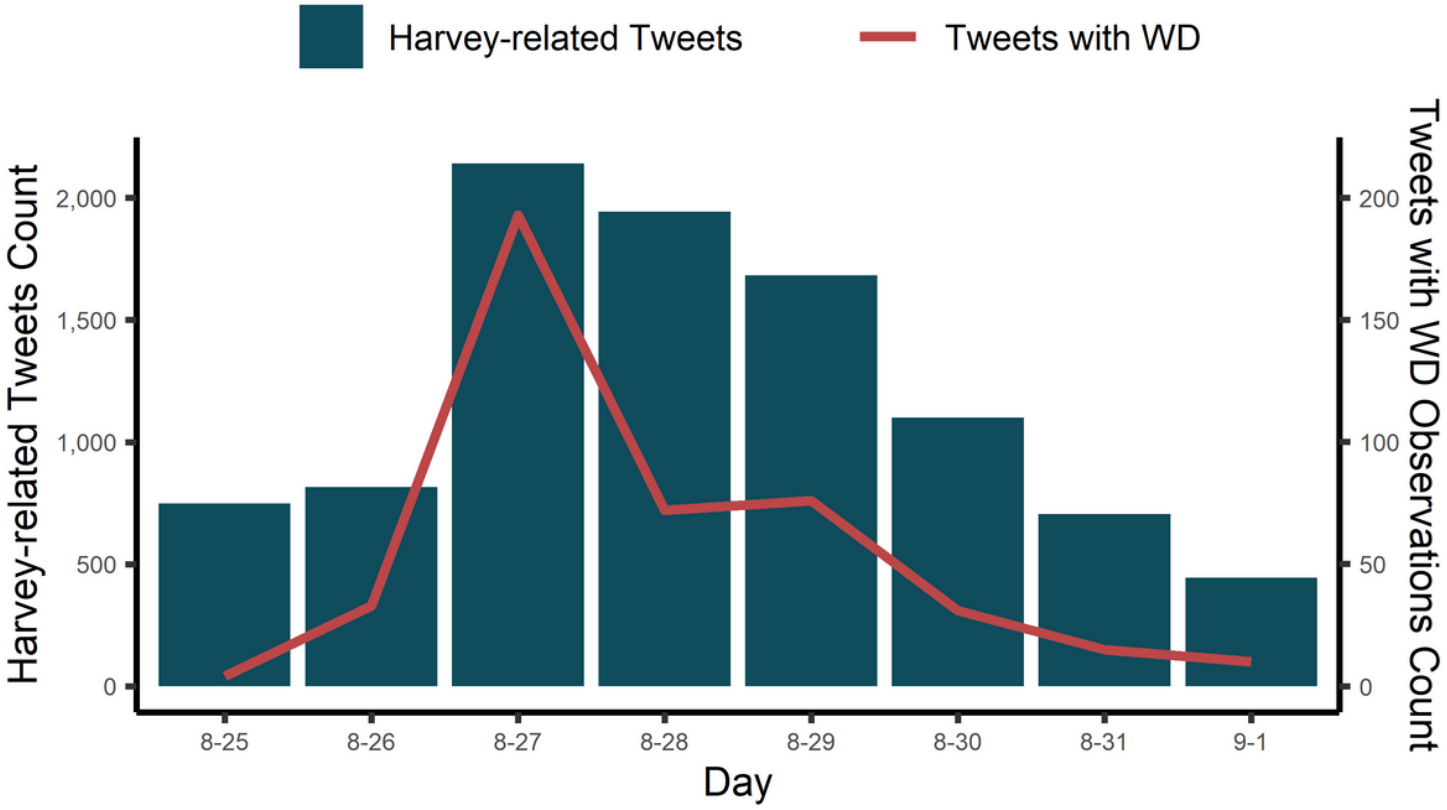

Figure 9

Summary of tweets with WD information (regardless of data modalities) compared to Harvey-related tweets on a daily basis.

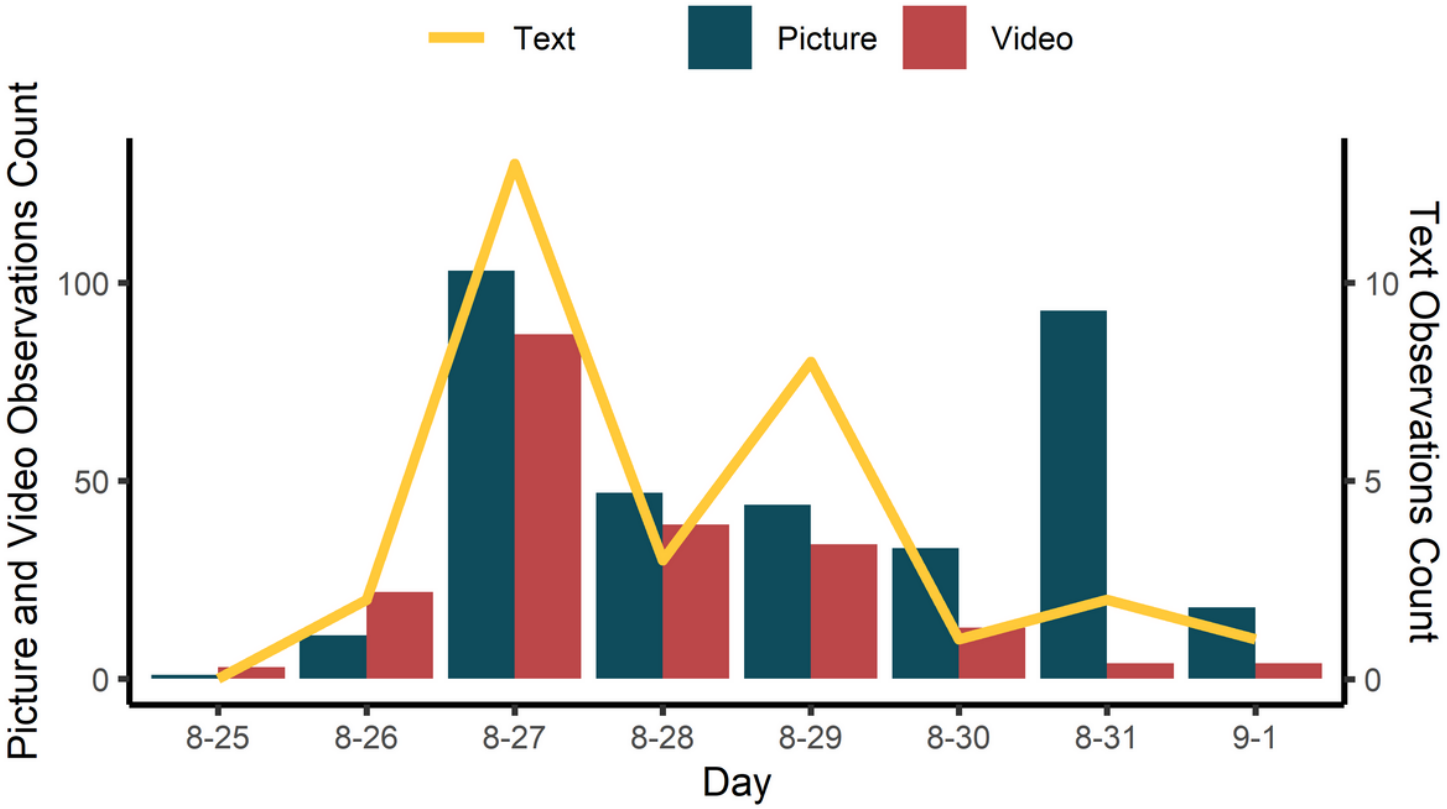

Figure 10

Summary of the variation in count of WD information extracted from the three VGI data modalities (text, picture, and video) from both tweets and crowdsourced data on a daily basis.

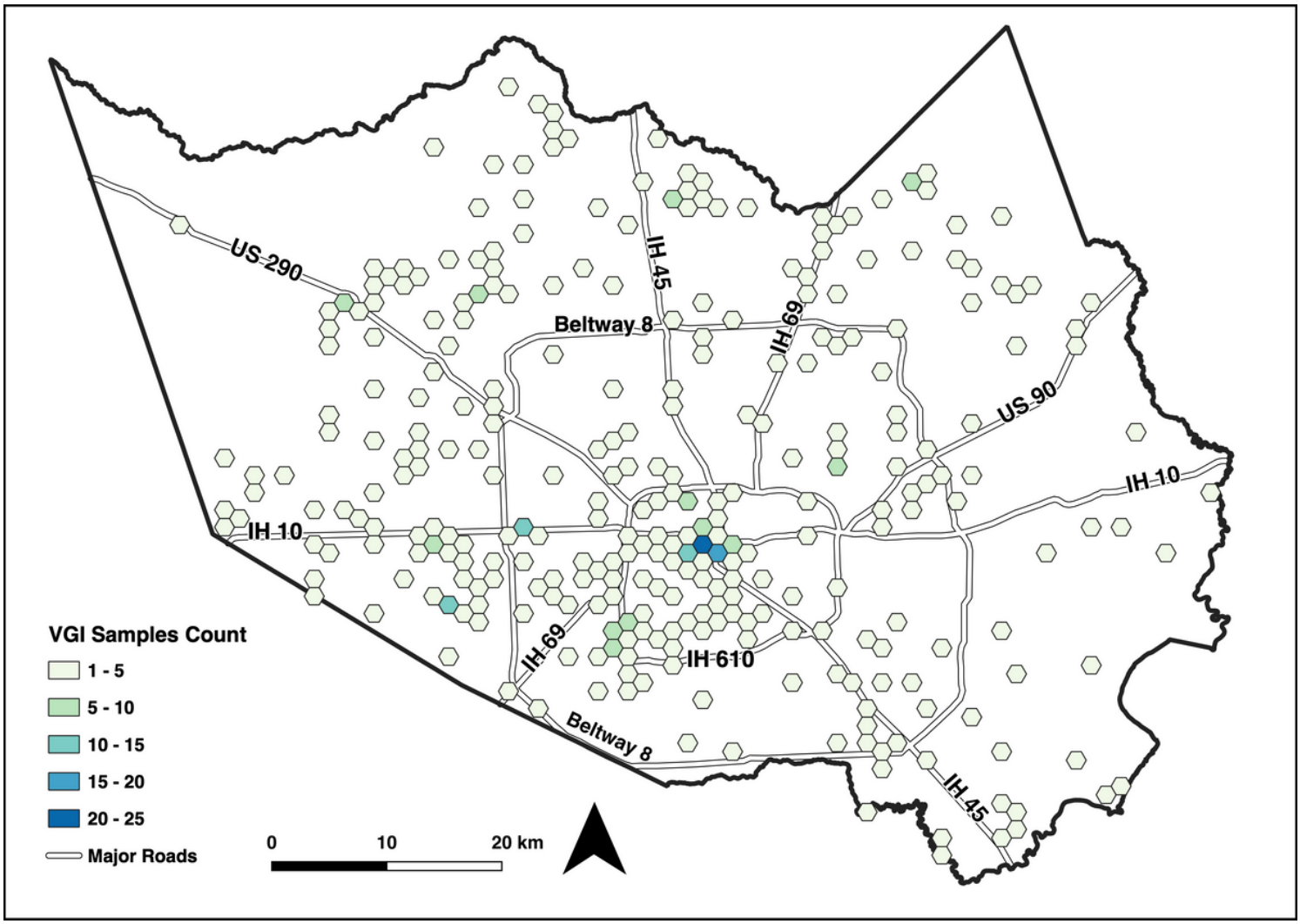

Figure 11

Geographic distribution of VGI observations in Harris County using hex bins of 1.5 by 1.5 km2.

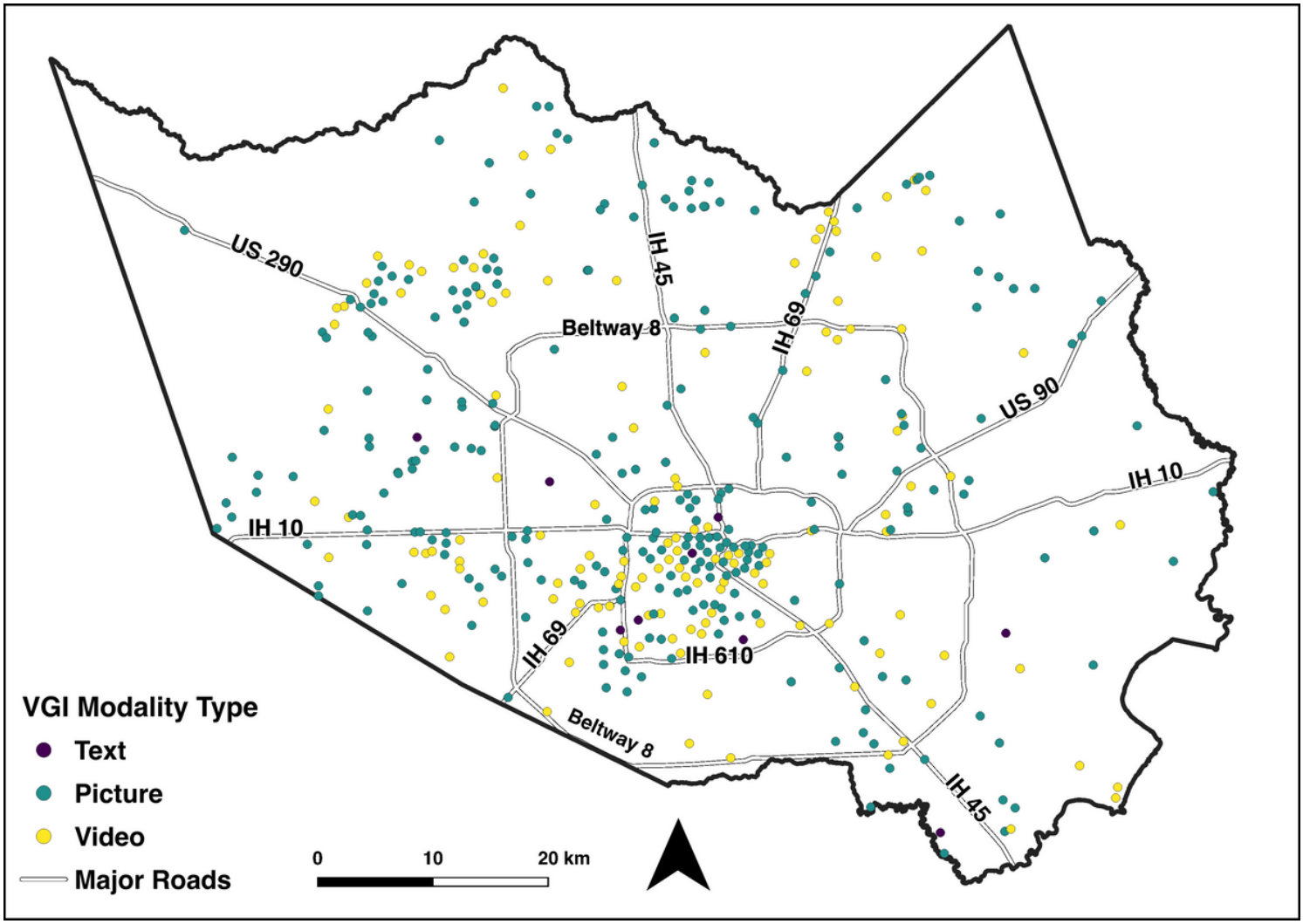

Figure 12

Geographic distribution of VGI data modalities in Harris County.

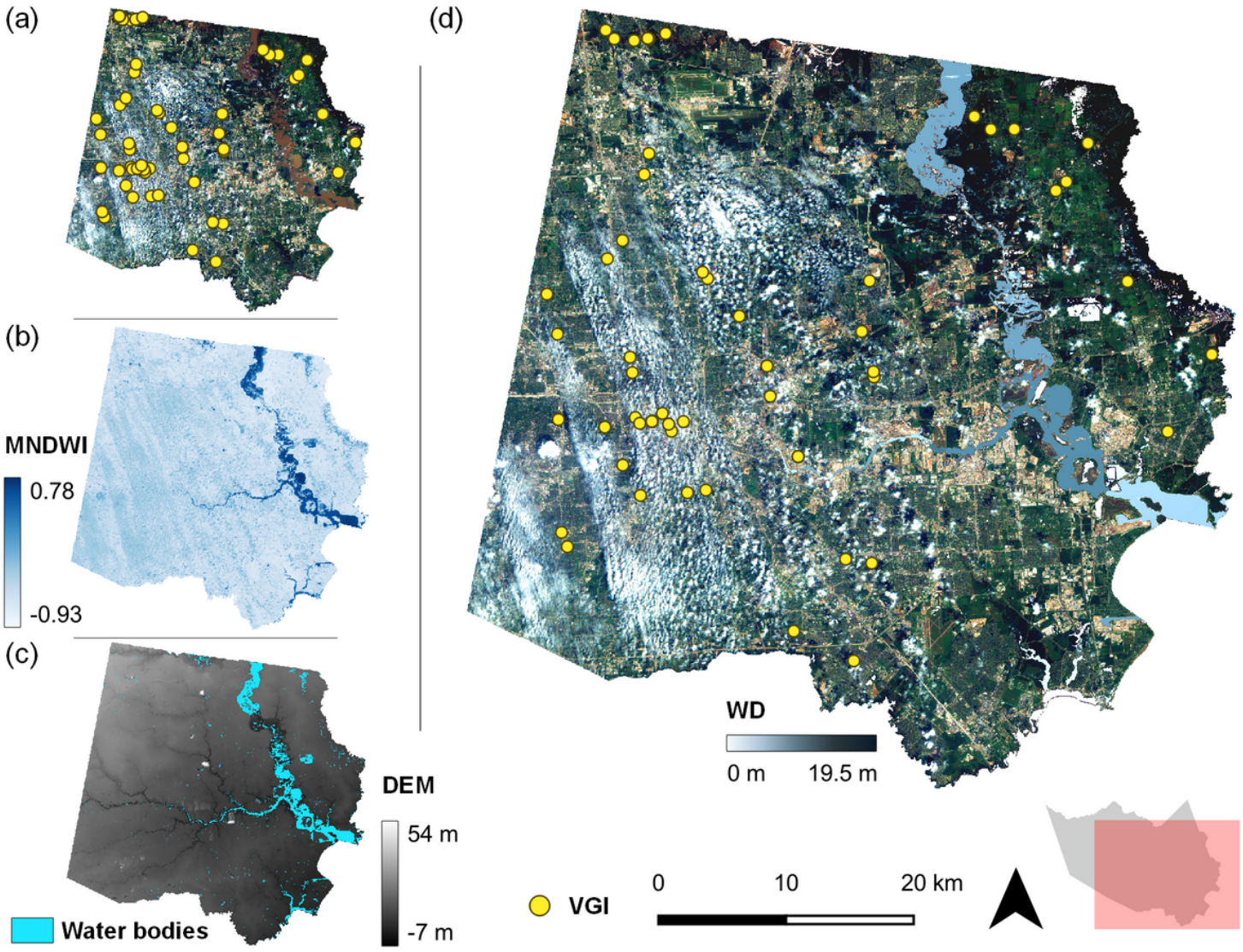

Figure 13

WD extraction from RS images. (a) Sentinel-2 true color image, (b) The output of the MNDWI, (c) water bodies delineated after applying the threshold (>= 0.10) and (d) The final WD derived from RS image.

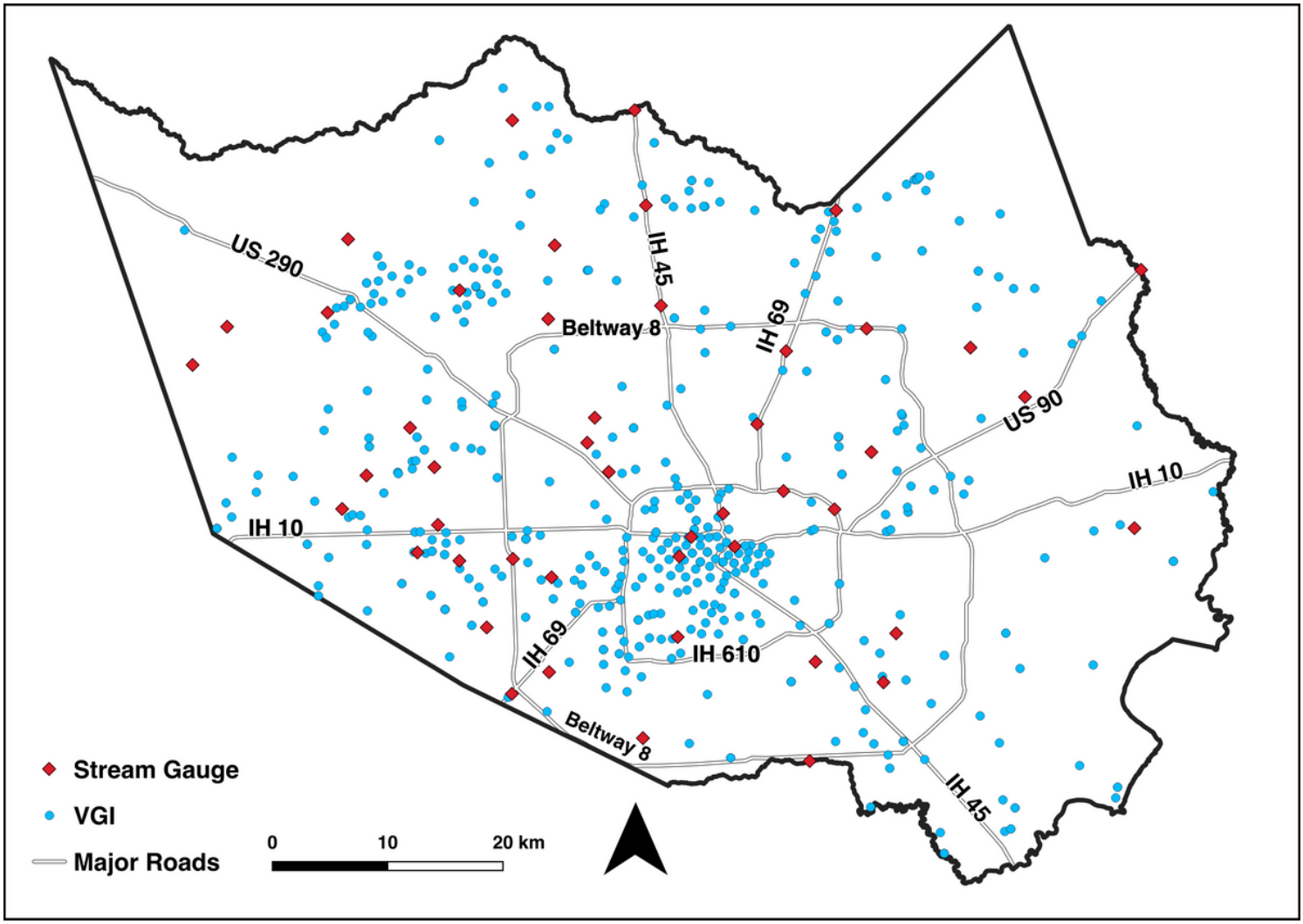

Figure 14

Geographic distribution of USGS stream gauges in Harris County.

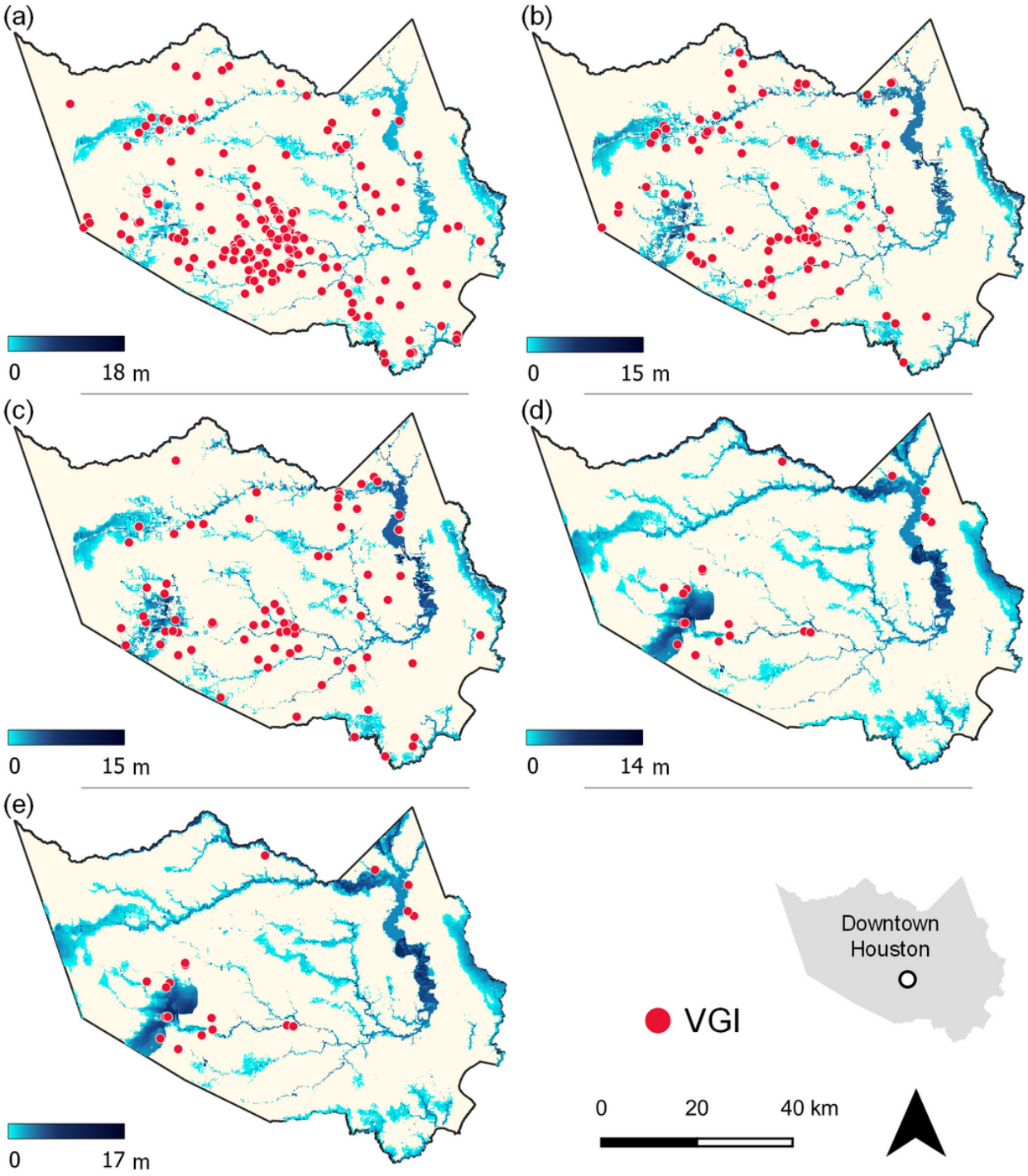

Figure 15

Modeled depth grids in meters from FEMA. (a) August 27th, (b) August 28th, (c) August 29th, (d) August 30th and (e) September 1st.

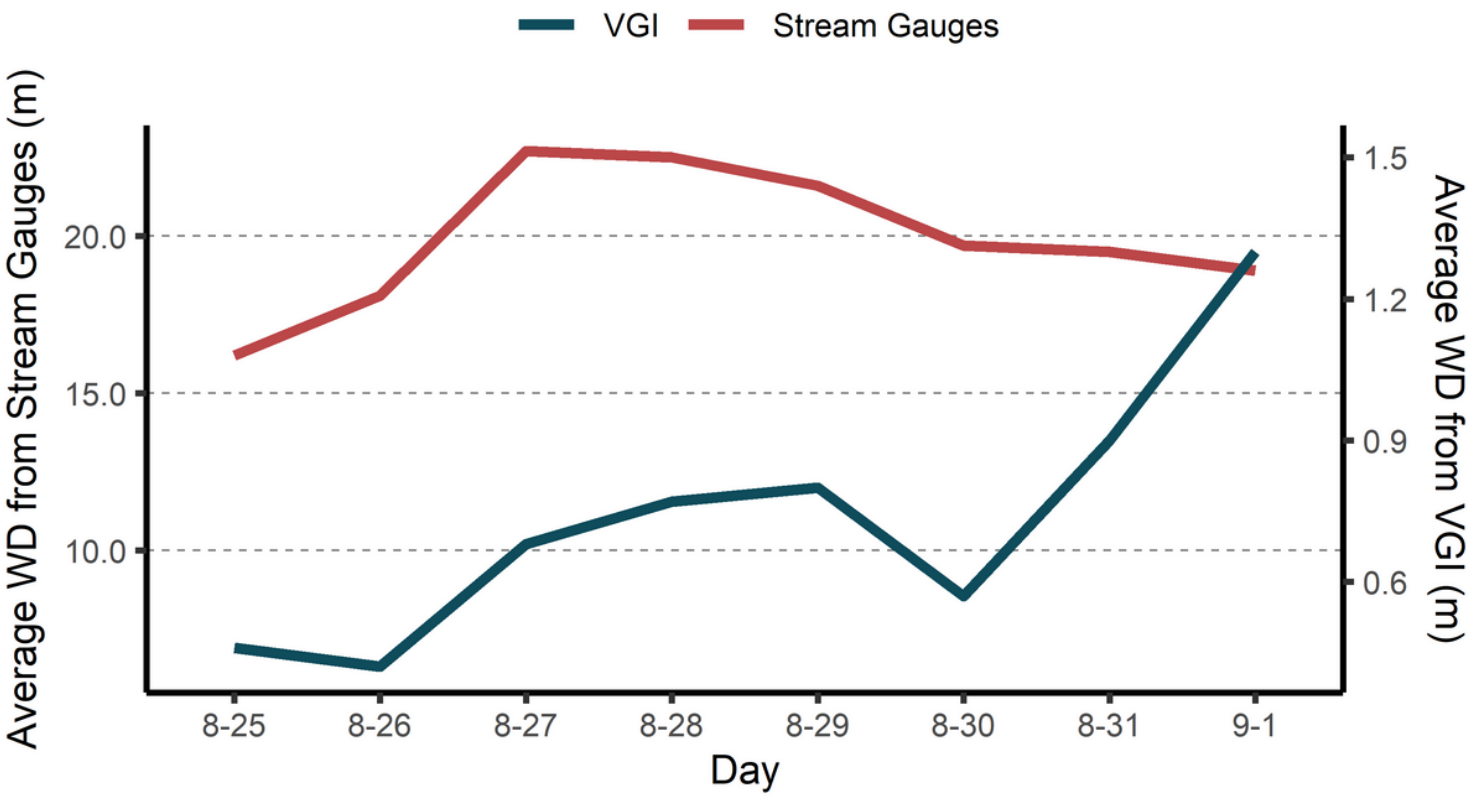

Figure 16

Daily comparison between average WD from VGI and average WD from stream gauges. The units are in meters.

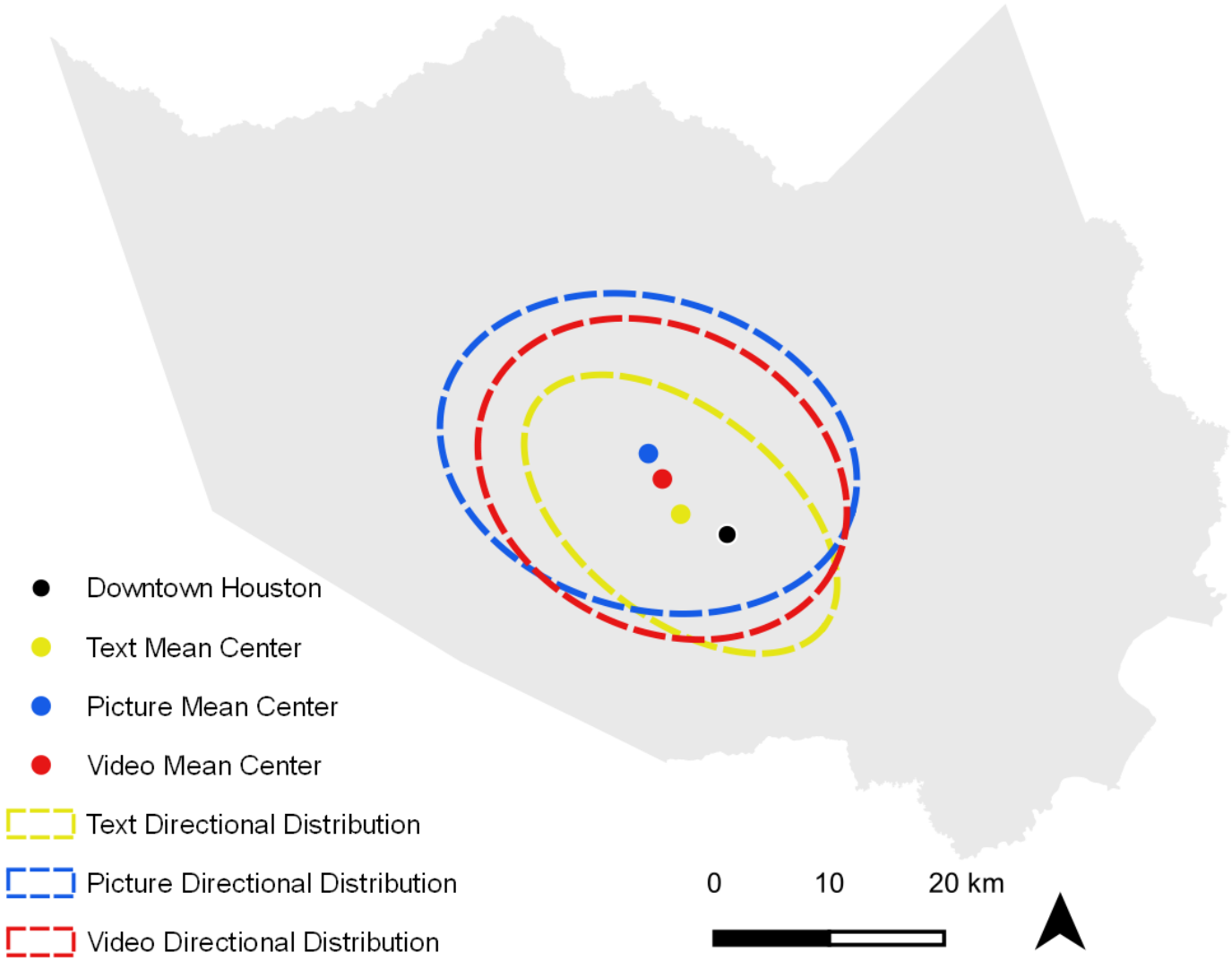

Figure 17

Mean center and directional distribution of VGI data modalities in Harris County.

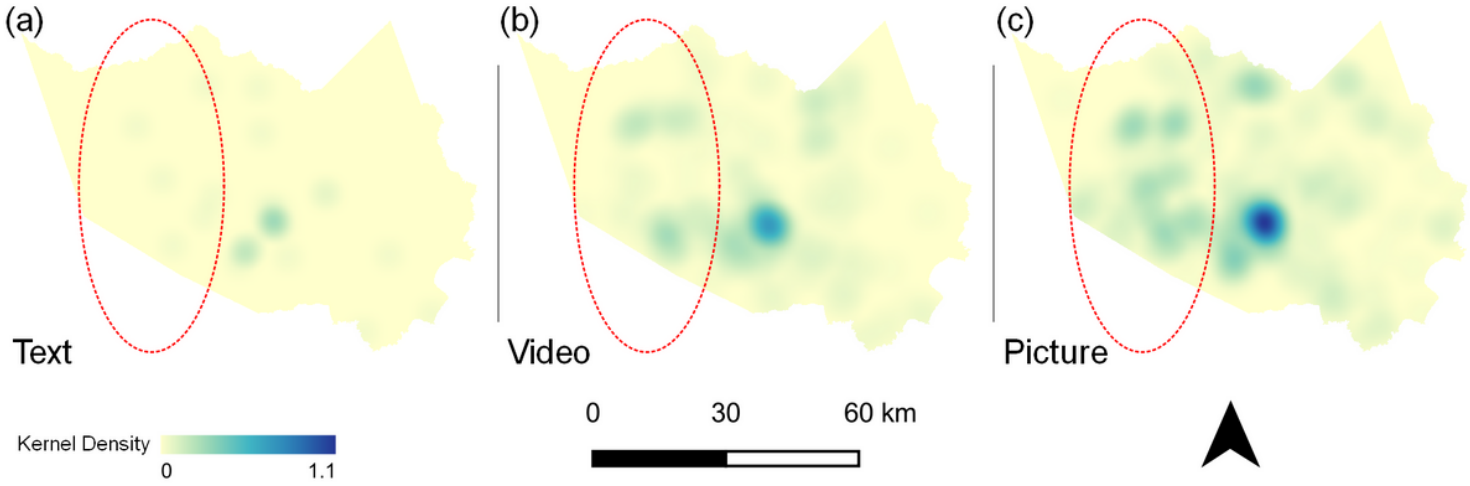

Figure 18

Kernel density of VGI across data modalities. Due to the variation in sample size and distribution, the western areas of Harris county, highlighted in red-dotted line, showed less text modality density (a), compared with the video modality (b) and picture modality (c).

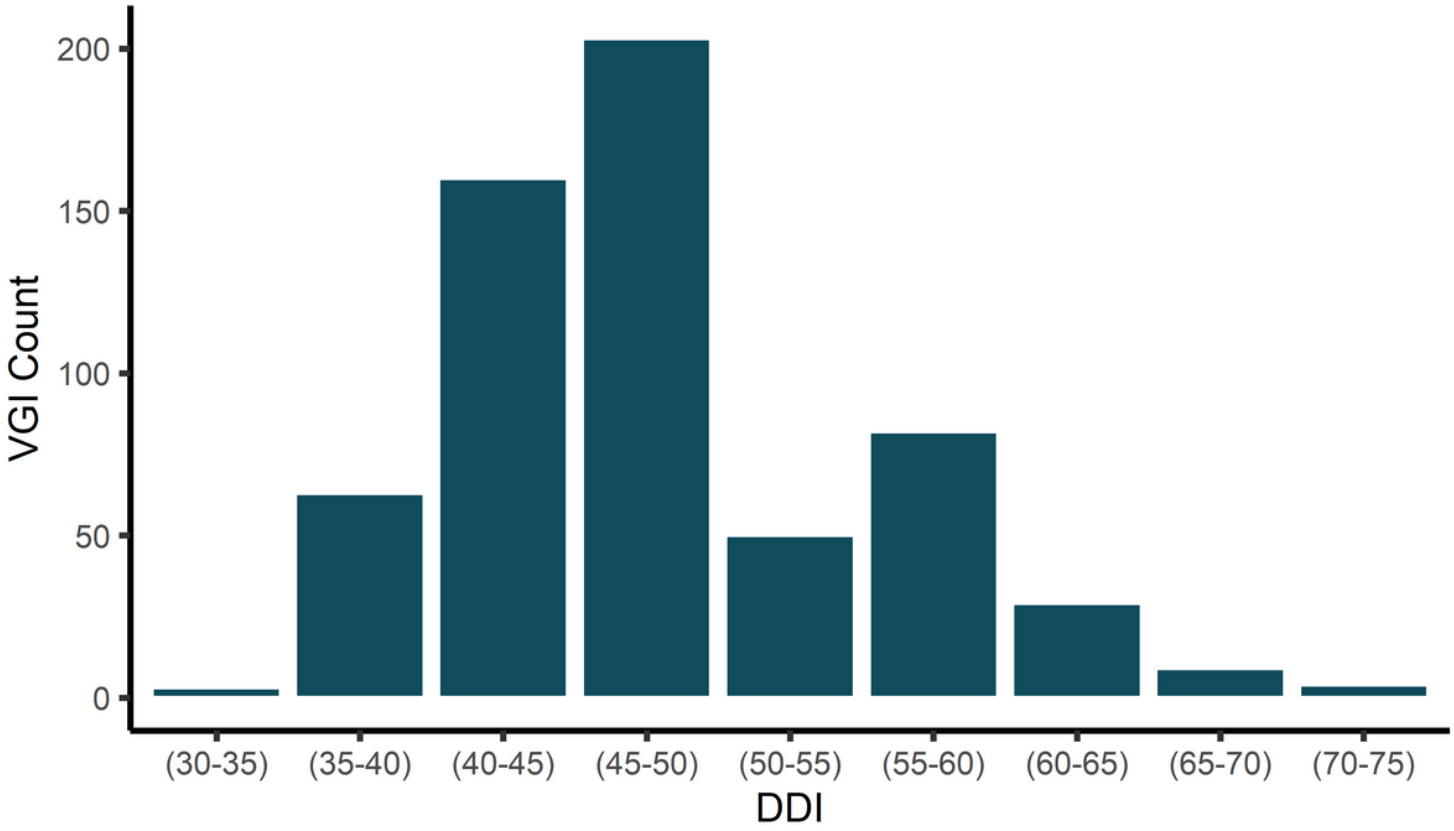

Figure 19

Bar chart of VGI count in DDI bins. Majority of VGI observations were at areas with moderate to lower digital divide. The DDI data were obtained from Mississippi State University (2018).

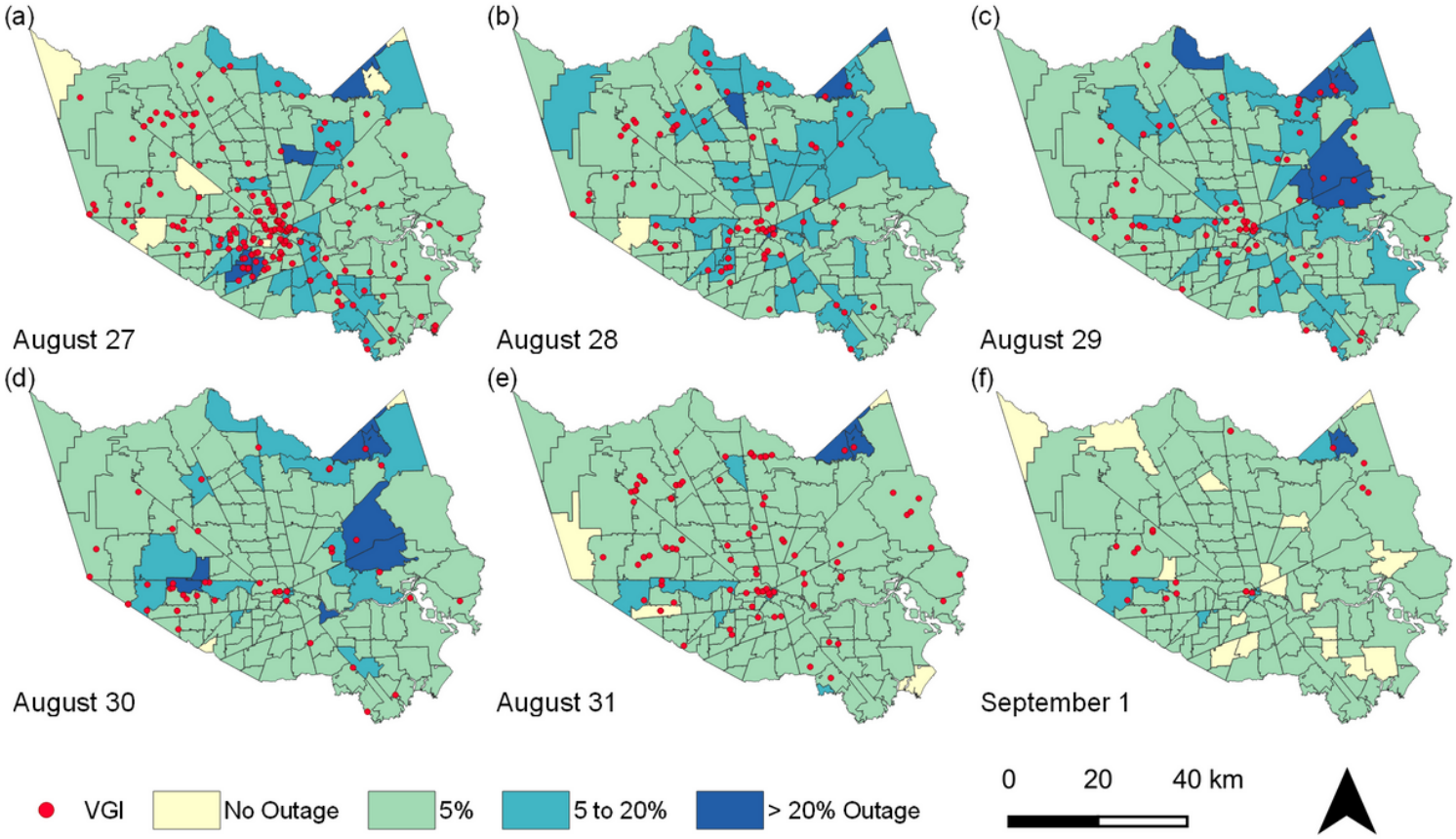

Figure 20

Power outages between August 27th (a) and September 1st (f) in Harris County. (data source: ESRI 2018).

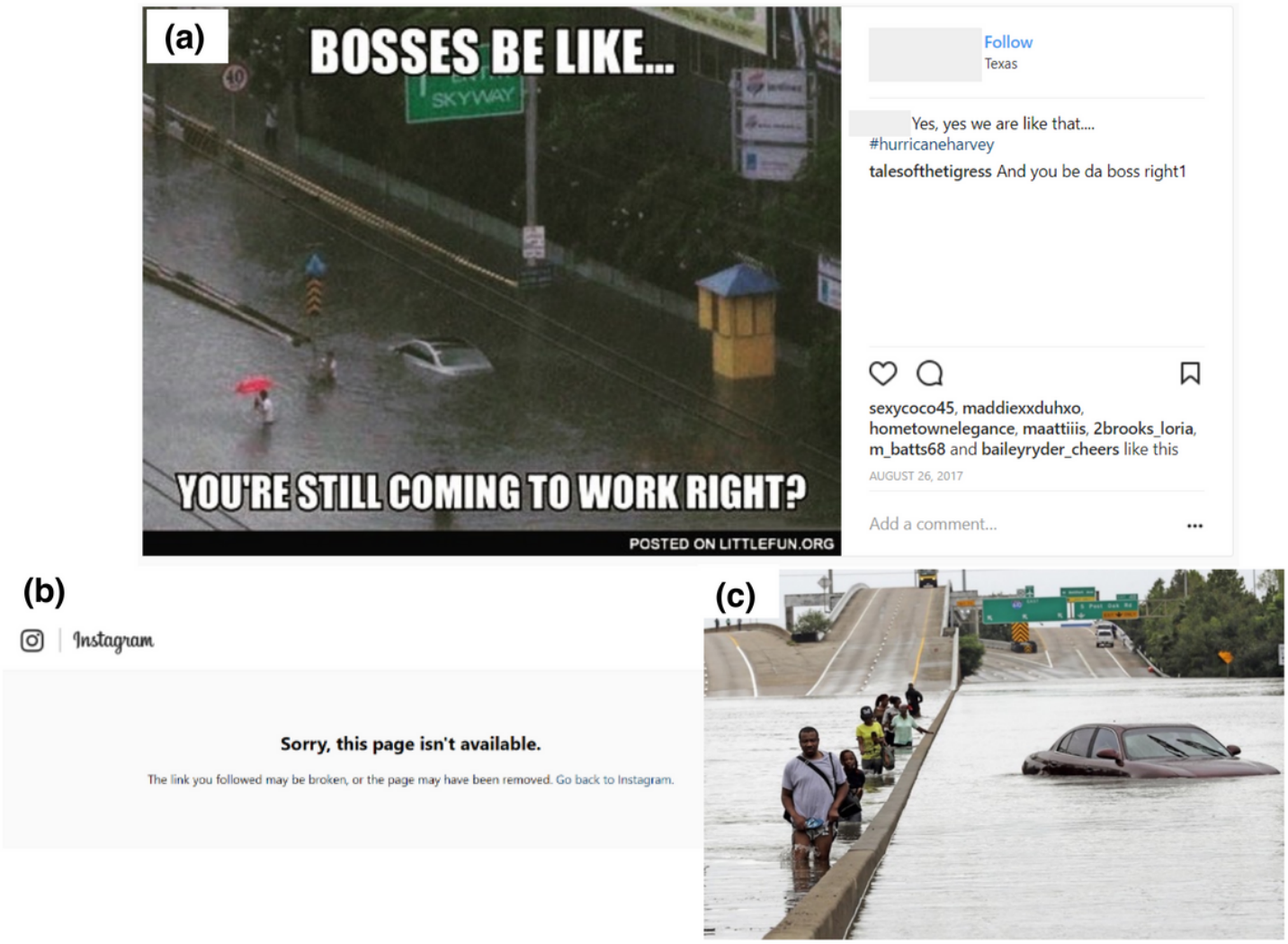

Figure 21

An example of a tweet with relevant hashtag but (a) irrelevant content, (b) missing or broken links or (c) repeated usage in multiple crowdsourced points.