# Enhancing the temporal resolution of satellite-based flood extent generation using crowdsourced data for disaster monitoring

## George Panteras & Guido Cervone

Taylor & Francis
Taylor & Francis Group

# Enhancing the temporal resolution of satellite-based flood extent generation using crowdsourced data for disaster monitoring

George Panteras[a] and Guido Cervone[a,b]

[a]GeoInformatics and Earth Observation Laboratory, Department of Geography and Institute for CyberScience, The Pennsylvania State University, University Park, PA, USA; [b]Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY, USA

**ABSTRACT**

Remote-sensing satellite data are routinely used during disasters for damage assessment and to coordinate relief operations. Although there is a plethora of satellite sensors able to provide actionable data about an event, their temporal resolution is limited by their revisit time, presence of clouds, and errors in the reception of data. These limitations do not allow for an uninterrupted monitoring, which is crucial during disasters and emergencies. This research presents an approach that leverages the increased temporal resolution of crowdsourced data to partially overcome the limitations of satellite data. The proposed approach focuses on the geostatistical analysis of a combined satellite and Twitter data to help delineate the flood extent on a daily basis. The crowdsourced data are used to augment satellite imagery from Advanced Land Imager instrument on Earth Observating One (EO-1) satellite, Landsat 8, WorldView-2, and WorldView-3. The proposed methodology was applied to estimate the daily flood extents in Charleston, South Carolina, caused by the October 2015 North American storm complex. The results of the proposed methodology indicate that the user-generated data can be utilized adequately to both bridge the temporal gaps in the satellite-based observations and also increase the spatial resolution of the flood extents.

## 1. Introduction

Satellite-based monitoring and assessment of natural hazards and especially of floods have played a key role for the last two decades. Numerous satellite-based systems have the capability of early detection of floods, gathering observations with of high spatial and spectral resolutions. In addition to the widely used electro-optical instruments, which are affected by cloud coverage that occlude observing the land, space-born synthetic aperture radar (SAR) can provide a remotely sensed solution in all weather conditions, day and night. Hence, satellite imagery acquired using both optical and microwave parts of the electromagnetic spectrum was successfully utilized to study

**CONTACT** George Panteras ✉ gxp37@psu.edu 🌐 GeoInformatics and Earth Observation Laboratory, Department of Geography and Institute for CyberScience, The Pennsylvania State University, University Park, PA, USA

flood risk management and mitigation, as well as for real-time emergency response (Van Der Sande, De Jong, and De Roo 2003; Taubenböck et al. 2011; Serpico et al. 2012). During flood events, satellite data are used to generate very accurate flood-extent surfaces and/or maps (Schumann, Di Baldassarre, and Bates 2009; Skakun 2012) which can then be utilized for calibration and validation purposes of hydraulic models (Khan et al. 2011). These flood extents are also crucial for rescuers during an emergency response operation because they provide a comprehensive view that can be used for rapid damage assessment, planning routes, setting priorities, and organize evacuations (Kussul et al. 2012). This becomes most important during time-sensitive operations where the near-real-time delineation and assessment of the flood extent are required very quickly, which preclude the manual mapping through *in situ* observations (Cossu et al. 2009). The satellite-derived flood assessment has been proved to be applicable and effective in all types of flood-related damage assessments, in urban areas (Gamba, DellAcqua, and Dasarathy 2005), in rural areas (Asante et al. 2007), or in general with different land-cover types (Joyce et al. 2009).

Despite the availability of sophisticated satellite sensors, capable of very high spatial and spectral resolution observations, their temporal resolution remains limited. In remote sensing and more specifically in sensors' design, there are always trade-offs between the basic fundamental sensor properties and resolutions: spatial, spectral, radiometric, and the temporal resolutions (Kennedy et al. 2009). The latter is referred to the revisit time of a satellite above the same point on the surface of the Earth. Most remote-sensing satellites are in what is called sun-synchronous orbit, which means that they cross the equator at roughly the same local time. They have a global coverage and a revisit time of about 16 days, which means that they observe the same point of the Earth about every 16 days. However, because satellites do not see a single point but a swath, the temporal coverage is somewhat reduced, depending on the spatial coverage and field of view. Conversely, the geostationary satellites have a very high temporal resolution (as short as 15 min) but with a much lower spatial resolution and coverage (rather than the entire Earth, they observe always the same hemisphere).

In the case of a natural hazard, the situational awareness is very critical, and time sensitivity is essential. Raised by the developments of Web 2.0, ubiquitous computing, and the recent technological advancements in networks and mobile devices, the proliferation of social media has led to the generation of massive amount of geospatial data. As defined by the Homeland Security Act of 2002, situational awareness is information gathered from a variety of sources that, when communicated to emergency managers and decision-makers, can form the basis for incident management decision-making.

Nowadays, massive amount of user-generated geospatial data from real-time data streams is available from social media platforms such as Twitter, Facebook, Instagram, etc., during natural hazards. The temporal resolution of social media is much higher than in the case of remotely sensed data. In fact, the crowdsourced data via social media can be generated and published almost instantaneously after the occurrence of an incident, which makes it very suitable for situational awareness applications (Yin et al. 2012).

The use of social media in disaster response, management as well as in situational awareness started gaining a lot of attention from the Geographic Information Science (GISc) community due to their very high temporal resolution. There is a proliferation

of a number of social media mining methodologies and systems which are leveraging the spatiotemporal properties of such user-generated geospatial data for the response and mitigation of natural hazards as well as for the enhancing of the situational awareness. This kind of data contains rich information about location, attributes, and semantic information (Fan et al. 2014). Social media and particularly Twitter tend to be utilized by ordinary people during the occurrence of a disaster and/or natural hazard, providing this way up-to-date and real-time information which is very valuable for the disaster management agencies. This kind of community-level situation awareness can play a significant role in decision-making for a more effective disaster response (Mukkamala and Beck 2016). Wan et al. (2014) developed a global flood disaster community cyber-infrastructure (CyberFlood), which leveraged cloud computing services and crowdsourcing data collection for the purposes of on-demand, location-based visualization, as well as statistical analysis and graphing functions. Schnebele and Cervone (2013), Goolsby and Cervone (2013), Cervone, Sava et al. (2016) and Cervone, Schnebele et al. (2016) presented a methodology for fusing Volunteered Geographic Information (VGI) data with remotely sensed data and a digital elevation model (DEM) in order to create hazard maps. As they have shown, even a small number of properly located VGI data is adequate to improve the flood assessment.

Eilander et al. (2016) managed to create flood maps for Jakarta in real time by combining DEM with flood depth observations and location references in tweets, demonstrating how useful the near-real-time information of social media for flood disaster management. MacEachren et al. (2011) presented a geovisual analytics approach, called 'Senseplace2,' which is a visual interface that collects and filters geocoded tweet content, organizing and understanding this way spatial, temporal, and thematic aspects of evolving crisis in order to support crisis management and situational awareness. Huang and Xiao (2015) presented a coding schema for separating social media messages into different themes within different disaster stages by utilizing text mining techniques in order to classify the tweets collected during the natural disaster caused by Hurricane Sandy in 2012. Soltani et al. (2016) created an interactive environment, called 'UrbanFlow,' which enables scientists to integrate fine-resolution social media data with authoritative data by suing distributed algorithms in order to gain deeper insight on mobility patterns through complex urban area. Huang and Cervone (2016) addressed the importance of social media and cloud computing for the detection, monitoring, and gaining situational awareness during a natural hazard with unparalleled scale and capacity.

In the present study, an approach is presented to leverage the increased temporal resolution of crowdsourced data to partially overcome the limited temporal resolution of satellite data, bridging the gap during flood monitoring, and providing a seamless situational awareness during incidents. The proposed approach focuses on the geostatistical analysis of Twitter data to help delineate the flood extent on a daily basis. The crowdsourced data are used to augment satellite imagery from Advanced Land Imager (ALI) instrument on Earth Observating One (EO-1) satellite, Landsat 8, WorldView-2, and WorldView-3 by fusing them together to complement the satellite observations. The proposed methodology was applied to estimate the daily flood extents in Charleston, South Carolina (SC), caused by hurricane Joaquin on October 2015.

## 2. Data

Multi-sourced crowdsourced and authoritative data relative to the October 2015 North American storm complex were collected and utilized for this study. More specifically, data were collected for the most impacted area which occurred in SC, between 27 September 2015 and 18 October 2015. The rainfall reached its maximum on 3 October 2015, and data were collected approximately $\pm 2$ weeks in order to ensure sufficient coverage before and after the peak of the event and to compare to the normal conditions, the maximum extent of the flood, and the receding of the waters. Specific details about the data collected are given in the following subsections, while a more detailed description of the study area is given in Section 3.

Figure 1 shows the temporal resolutions for the multi-source data set along with the gage heights of the three major rivers of Charleston, SC, which was one of the most heavily impacted areas. The data are provided by the U.S. Geological Survey Surface-Water Data for the Nation. Areas that were affected the most by the flooding event were near smaller creeks and streams, particularly the tributaries to the three major rivers crossing the city of Charleston: the Ashley, Cooper, and Wando rivers. As a result, thousands of homes and businesses were severely impacted, and many roads and bridges were damaged, causing serious travel disruptions lasting for many days.

### 2.1. Crowdsourced data

Twitter was selected as the key crowdsourced data source due to its very high temporal resolution as well as comprehensive geographical coverage. When a flood occurs in a remote area where Internet access is unavailable and the road network is limited or impassable, the spatial distribution of the tweets may be sparse and under-representative
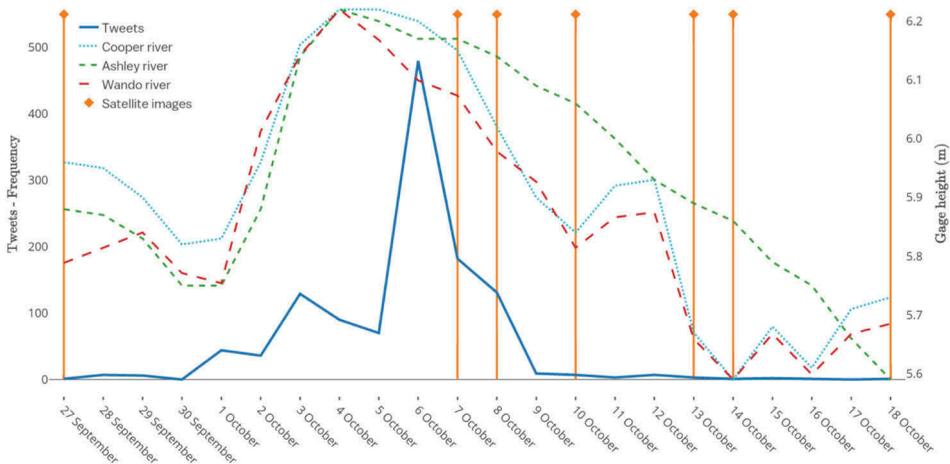


**Figure 1.** Temporal resolutions and time stamps of the multi-source data sets collected. The blue continuous line represents the daily number of tweets related to the flood. The remaining three lines show the rivers' gages height during the entire event. Satellite imagery is shown by the diamond terminated lines as their timestamp corresponds to a single day each; therefore, they are scaleless in the plot.

of the event. In this case, Twitter data were readily available and reliable. The study area selected, a densely populated area with good road infrastructure, enables a realistic analysis of the flood event because both of the sufficient amount of tweets as well as of their dense geographical distribution.

A total of 2393 geotagged tweets relevant to the study area were collected over the period between 27 September 2015 and 18 October 2015, using the R twitter library. Prior to the event, an automatic Twitter ingesting process was started by the authors and their collaborators to store all tweets with geographical coordinates originating from the USA in a MongoDB server. Therefore, the selection of the geotagged tweets consisted in a spatio-temporal query to the MongoDB server, which returned all messages occurring within the area of interest (AOI) and in the time window specified. A further selection was performed by specifying a more focused AOI relative only to the city of Charleston and a set of relevant hashtags, reducing the amount of relevant tweets to 1210, about half of the original data. The analysis presented in this article is based on this number of tweets. The hashtags used for filtering the tweets were selected to specifically identify the flooding event (e.g. as #hurricane, #joaquin, #flood).

## 2.2. *Remote-sensing data*

The satellite images were downloaded from the United States Geological Survey (USGS) Hazards Data Distribution System (HDDS), an event-based platform that provides access to remotely sensed imagery and other geospatial data sets relative to worldwide disasters. This study is based on data listed in the HDDS under Hurricane Joaquin, the main cause for the floods. Specifically, for the AOI of Charleston, SC, satellite data from four different sensors of varying spatial, spectral, and temporal resolution were available (Table 1). Figure 2 shows a satellite image used in the study along with the 1210 tweets available. The Twitter messages are also shown in Figure 2 and are grouped in four specific consecutive phases to illustrate their geographic distribution during the key phases of the flood progression.

Although the collected remote-sensing data provide a good assessment for the flood event, their temporal coverage is limited. First, based on Table 1, the highest possible temporal resolution which can be achieved from satellite data alone is at most one day. In contrast, social media feeds have a rate which ranges in hours or minutes. In addition, based on Figure 2, not all of the satellite data have full coverage of the AOI. For instance, in the case of Landsat 8, the extent of the data has dimensions of $170\,km \times 183\,km$. However, in the case of WorldView-3, the extent size is $66.5\,km \times 112\,km$ (5 strips) in 'mono' acquisition mode, and $26.6\,km \times 112\,km$ (2 pairs) in 'stereo' mode. The decrease in resolution is expected because of the trade-off between spatial and spectral resolutions previously discussed.

**Table 1.** Sensor properties of the collected satellite imagery.

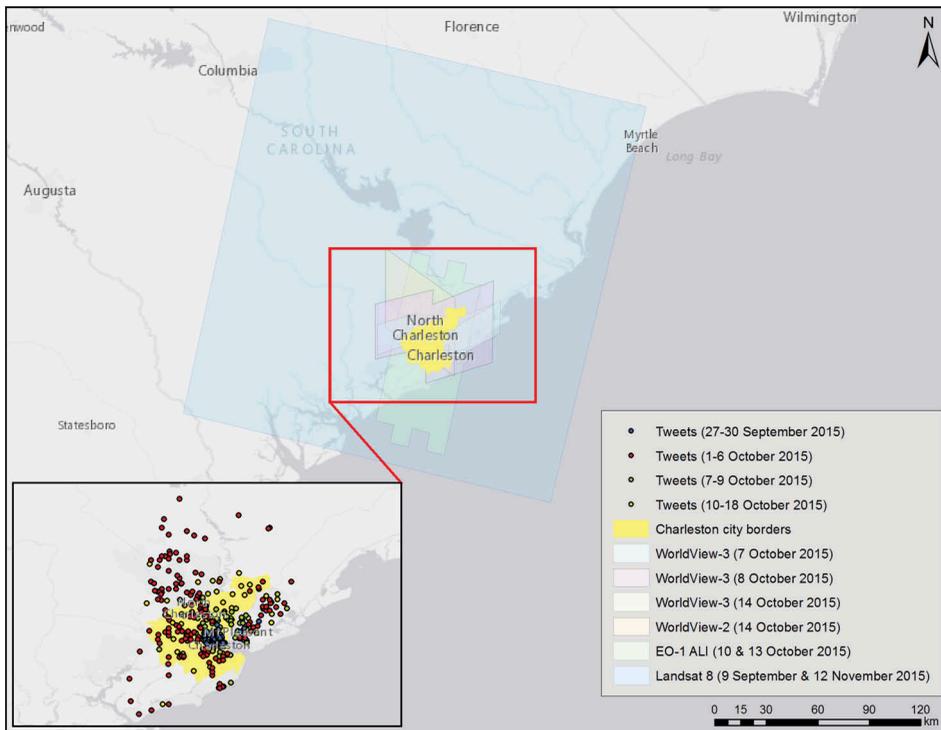| Sensor | Spatial (pan) | Spectral | Temporal | Acquisition date | | |
|---|---|---|---|---|---|---|
| Landsat 8 | 15 m | 11 bands | 16 days | 9 September 2015 | 12 November 2015 | – |
| EO-1 ALI | 10 m | 10 bands | 16 days | 10 October 2015 | 13 October 2015 | – |
| WorldView-2 | 0.46 m | 8 bands | 1.1 days | 14 October 2015 | – | – |
| WorldView-3 | 0.31 m | 29 bands | <1 day | 7 October 2015 | 8 October 2015 | 14 October 2015 |

**Figure 2.** Spatial extent of the satellite imagery acquired for the Hurricane Joaquin as well as the Twitter feeds over the study area of Charleston, SC.

Therefore, although satellite data are available for this particular date, the spatial coverage is not sufficient to assess the entire AOI. The lack of coverage prevents a complete and seamless flood assessment. In addition, another significant factor that is preventing a seamless flood monitoring is the cloud coverage present in the image and that it cannot be avoided with this type of data. For example, the WorldView-3 scene, which was acquired on 8 October had 56% cloud coverage, which limits the analysis when the clouds occlude the flooded areas.

## 3. Methodology

### 3.1. *Case study – AOI*

The October 2015 North American storm complex, a high precipitation event, was the cause of the historic flooding occurred in SC. The heavy rainfall was the result of an upper atmospheric low-pressure system that funnelled tropical moisture from Hurricane Joaquin into the State (Musser et al. 2016). During the night between 3 and 4 October, the Charleston International Airport recorded a 24-hour rainfall of 290 mm, which was the pick and considered a 500-year event level. The flood event culminated on 4 October, where as an outcome of the over-flooded rivers, many roads, bridges, vehicles, and homes were washed away. The weather complex in SC was responsible for 19 deaths and a cost of damages estimated at $12 billion (USD). Out of the many counties

that were affected, Charleston was selected to be the study area due to the fact that three major rivers run through the city and also is the second biggest city in SC.

## 3.2. *Satellite-based flood detection*

Among the various methods for extracting water pixels from satellite imagery, the two most established are the single-band and the multi-band method, depending on the number of the bands utilized. In the first category, the water detection is based on applying a threshold in a single band related to water features. In the second category, the two main methods are supervised classification based on spectral signatures and band-ratio based on two multispectral bands. The method that was chosen for this study was the latter one of which the most well-known index is the normalized difference water index (NDWI) as was proposed by McFeeters (1996). Specifically, due to the fact that the NDWI tends to mix extracted water with the surrounding built-up noise, since is not very robust in suppressing the signal from the built-up land, a modification of this index was utilized instead. Xu (2006) proposed a modification of the NDWI which is based on the shortwave infrared (SWIR) radiation instead of the near-infrared (NIR) one, called modified NDWI (MNDWI) (see also Li et al. 2013). This index is able to discriminate water from non-water features in higher accuracy and it is very suitable for built-up environments as in the case of a city such as Charlottesville. The MNDWI can be expressed as follows:

$$MNDWI = \frac{\rho_{Green} - \rho_{SWIR}}{\rho_{Green} + \rho_{SWIR}} \tag{1}$$

The computation of MNDWI was implemented via the statistical package R. Before generating MNDWI, an essential preprocessing step consists in converting the at-sensor radiance, which refers to raw quantized calibrated pixel values, to the top-of-atmosphere (TOA) reflectance. This is necessary because TOA removes artefacts caused by the cosine effect at different solar zenith angles, solar irradiance arising from spectral band differences, and variation in the Earth–Sun distance between different acquisition dates (Chander et al. 2009). The TOA can be expressed as follows:

$$\rho_\lambda = \frac{\pi L_\lambda d^2}{(ESUN)_\lambda + \cos \vartheta_s} \tag{2}$$

where $\rho_\lambda$ is the TOA reflectance of wavelength $\lambda$ (unitless), $d$ is the Earth–Sun distance (astronomical units), $ESUN_\lambda$ is the mean exoatmospheric solar irradiance (W m$^{-2}$ μm$^{-1}$), $\vartheta_s$ is the solar zenith angle (°), and $L_\lambda$ is the spectral radiance at wavelength $\lambda$ at the sensor's aperture (W m$^{-2}$ sr$^{-1}$ μm$^{-1}$). After calculating the reflectance, the computation of MNDWI is based on the definition of the Green and SWIR bands for each satellite. The wavelengths used might vary among satellites, since the spectral bands of each satellite usually cover a slightly different part of the electromagnetic spectrum. The selected properties of MNDWI for each satellite sensor with respect to the selected bands and their corresponding wavelengths are summarized in Table 2.

**Table 2.** Bands used to compute MNDWI for each satellite sensor.

| Sensor | SWIR | | | Green | | |
|---|---|---|---|---|---|---|
| | Band | $\lambda$ (μm) | GSD (m) | Band | $\lambda$ (μm) | GSD (m) |
| Landsat 8 | SWIR-1 | 1.57–1.65 | 30 | Green | 0.53–0.59 | 30 |
| EO-1 ALI | MS-5 | 1.55–1.75 | 30 | MS-2 | 0.52–0.60 | 30 |
| WorldView-2 | NIR-1 | 0.86–1.04 | 1.84 | Green | 0.51–0.58 | 1.84 |
| WorldView-3 | SWIR-2 | 1.55–1.59 | 3.70 | Green | 0.51–0.58 | 1.24 |

### 3.3. *Crowdsourced-based flood detection*

In the case of flood assessment using crowdsourced data, the goal is primarily to analyse and visualize the statistical significance of the areas that are more prone to get flooded. Among the most well-known geostatistical techniques for this purpose, the hotspot detection found to be the most suitable solution, providing with both descriptive spatial statistics and complete visual capabilities. There are three major methods for the detection of hotspots, the Getis-Ord ($G_i$*) statistic, the Kernel Density Estimation (KDE), and the spatial autocorrelation, global Morans I statistic (Kuo, Zeng, and Lord 2011). The Getis-Ord $G_i$*-statistic (Ord and Getis 1995), which results on the identification of the statistically significant spatial clusters for both high cell values ('hotspots') and low cell values ('cold spots') in a heat map, was selected over the other two methodologies. One of the main advantages of the $G_i$* statistic is that enables the user to test the statistical significance of the results based on the calculation of the z-scores (Burt, Barber, and Rigby 2009). In addition, it is very useful because it acts as an indicator for local autocorrelation, meaning that it captures the spatial autocorrelation as it varies locally over the study area and calculates a statistic for each datapoint (Haining 2003). This is achieved by the evaluation of the level to which each point is bordered to other point of similarly high or low values within a specified geographical distance and/or neighbourhood (Peeters et al. 2015). The $G_i$* statistic can be described by the following equation.

$$Gi^* = \frac{\Sigma_j w_{ij}(d) x_j}{\Sigma_j x_j} \qquad (3)$$

where $G_i$* is the local G-statistic for a point ($i$) within a distance ($d$), $x_j$ is the attribute value of each neighbour, and $w_{ij}$ are the spatial weights for the target-neighbour $i$ and $j$ pair. Concerning the spatial weights, they are the $n$ by $n$ elements ($n$ is the number of observations) of the spatial weight matrix W.

Based on Equation (3), the assignment of appropriate weights can impact significantly the result of the hotspot analysis, and therefore, it is critical to define optimal weights based on the particular characteristics of the phenomenon and study area. In the present study, the crowdsourced data are generated by users which act as sensors. As in any remote-sensing technology, the position of the sensor with respect to the AOI is essential. For instance, one of the fundamental properties of a sensor, the flight altitude, plays significant role in the spatial resolution. Similarly, it can be argued that in the case of crowdsourcing during a flood, the feeds from users that are located closer to the river have higher significance.

For the elevation information, the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) global digital elevation model (GDEM) version 2 was

acquired from the USGS EarthExplorer website. The GDEM v2 has resolution of 1 arc-second (approximately 30 m at the equator) grid and referenced to the 1984 World Geodetic System (WGS84)/1996 Earth Gravitational Model (EGM96) geoid. The overall vertical accuracy is of around 17 m and the horizontal spatial resolution in the order of 75 m. Based on the elevation data, the maximum height for the study area was found to be 28 m. Six weights were created which correspond to six hypsometric zones ever 5 m each.

In order to visualize the detected hotspots in a heat map, the Kernel Interpolation with barriers (KIB) was chosen. This interpolation is a variant of a first-order local polynomial interpolation that improves traditional kernel estimation methods by accounting for barriers within the study area (Fan and Gijbels 1996). KIB model uses the shortest distance between points so that points on the sides of the specified nontransparent (absolute) barrier are connected by a series of straight lines. This is very critical for applications as the present one where the natural disaster is flooding, especially with the presence of the river. Since the main interest is the monitoring of the flood extent, the focus of the geostatistical analysis should be concentrated in the areas next to the river. For this purpose, the barriers utilized for the KIB interpolation are the river borders. Specifically, the river extent that was utilized was the one extracted from Landsat 8 at 9 September 2015. The resulted heat map is using five classes according to the z-scores thresholds: 90% significant ($z$-score $\geq$ 1.64), 95% significant ($z$-score $\geq$ 1.96), 99% significant ($z$-score $\geq$ 2.57), 99.9% significant ($z$-score $\geq$ 3.29), and all the non-significant cells were grouped in a fifth class.

## 4. Results

### 4.1. *Flooded area assessment based on remote sensing*

The application of MNDWI resulted in the creation of rasters where the brighter pixel values correspond to flooded water. In order to delineate the flooded areas, as part of the classification post-processing, the next step was to apply an appropriate threshold in order to accurately extract only the flooded surfaces. The critical aspect is to maximize the variance between the water surface and other background features such as vegetation, soil, etc., in order to minimize the probability of misclassification. In Figure 3, the results of computation of MNDWI are illustrated for every satellite image of the data set. As it can be seen, in the black and white NDWI images, the with pixels (higher values) resemble water, while the dark pixels (lower values) resemble non-water features.

Figure 4 shows the result of applying the threshold. The values were set across the interval 0.25–0.31, selecting manually the optimal value for the images per each satellite sensor. After the delineation of the flooded areas in the MNDWI of all the satellite imagery, this pixels were converted into polygons (see Figure 4(d)), as needed for the next part of the analysis.

### 4.2. *Flooded area assessment based on crowdsourcing*

In order to monitor the flood progression, the crowdsourced data were divided in four consecutive phases, allowing this way a spatio-temporal monitoring of the incident,
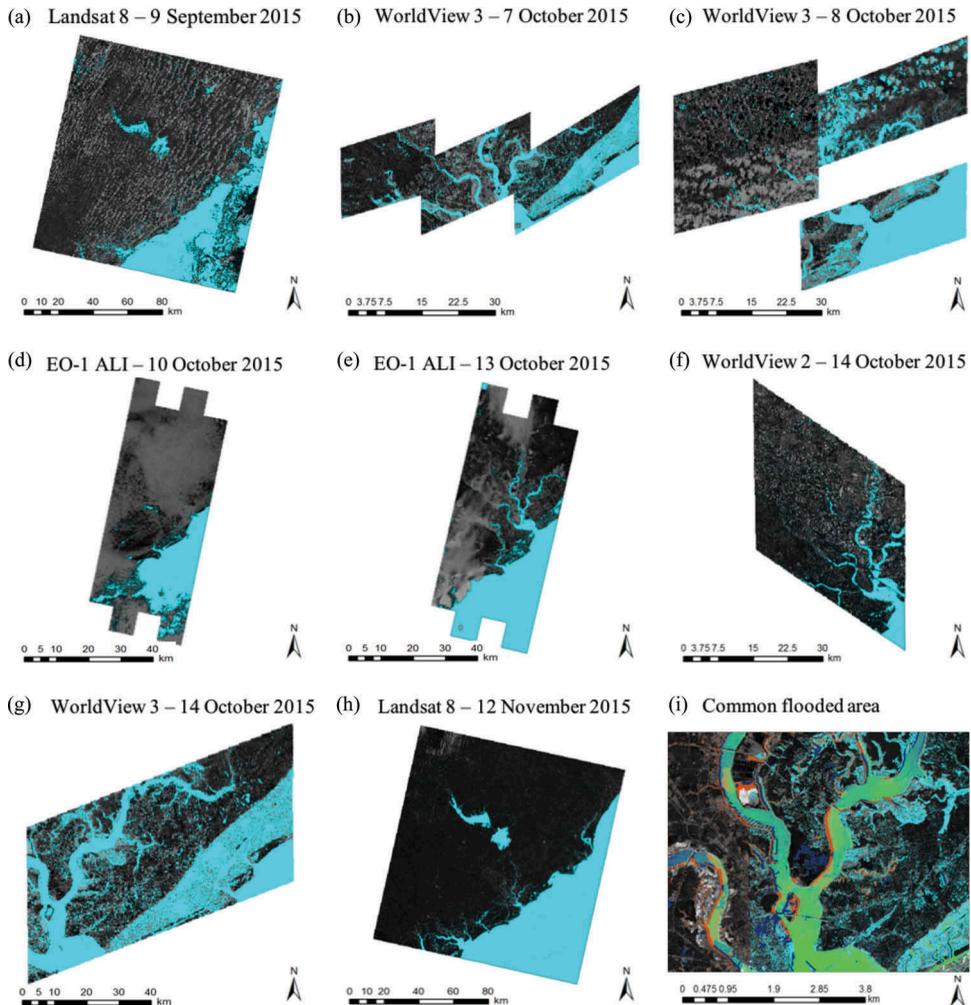
(a)　Landsat 8 – 9 September 2015　　(b)　WorldView 3 – 7 October 2015　　(c)　WorldView 3 – 8 October 2015

(d)　EO-1 ALI – 10 October 2015　　(e)　EO-1 ALI – 13 October 2015　　(f)　WorldView 2 – 14 October 2015

(g)　WorldView 3 – 14 October 2015　　(h)　Landsat 8 – 12 November 2015　　(i)　Common flooded area

**Figure 3.** Detected water in every satellite image based on the MNDWI. (a) Landsat 8 (9 September 2015), (b) WorldView-3 (7 October 2015), (c) WorldView-3 (8 October 2015) (d) ALI (10 October 2015), (e) ALI (13 October 2015), (f) WorldView-2 (14 October 2015), (g) WorldView-3 (14 October 2015), (h) Landsat 8 (9 September 2015), and (i) common area.

based on the stages of emergency of a natural hazard. Specifically, the four defined phases were pre-crisis (27–30 September), peak (1–6 October), receding (7–9 October), and recovery (10–18 October). The following figures (Figures 5–8) are illustrated the results of the hotspot analysis and the KIB interpolation for each of the event phases.

## 5. Conclusions

The results of the proposed methodology show that the user-generated data can be utilized to both bridge the temporal gaps in the satellite-based observations and also increase the spatial resolution of the flood extents. Results show that by fusing satellite and crowdsourced data, it is possible to increase the temporal resolutions of the
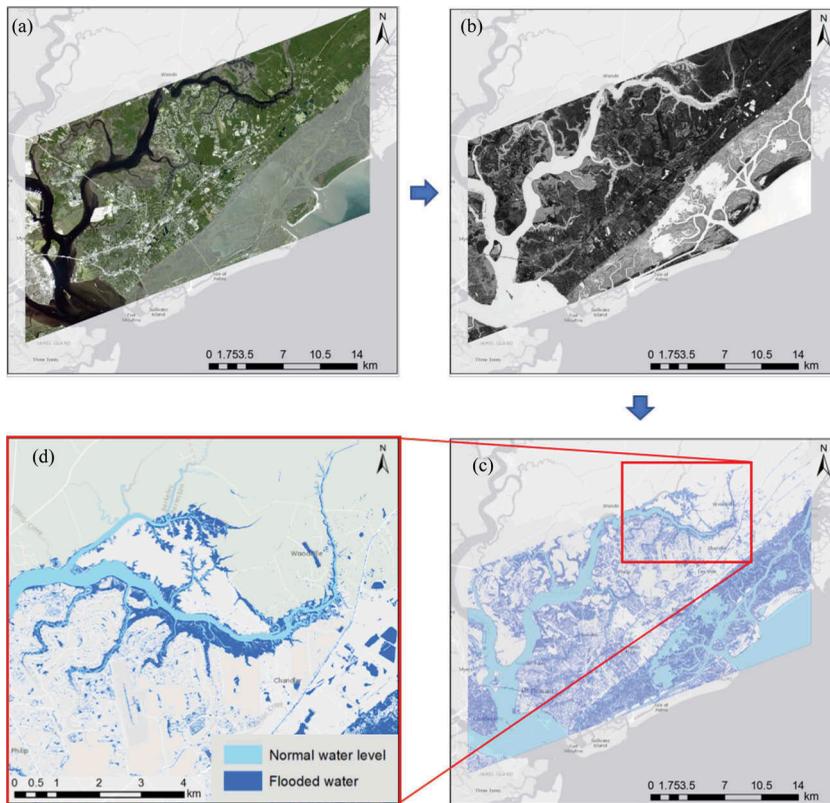
**Figure 4.** (a) WorldView-3 scene, true colour band combination (RGB = 4,3,2), acquisition date: 14 October 2015, (b) MNDWI raster, (c) flood map after applying appropriate threshold, and (d) zoomed in area of the flood map.

answers, as well as their accuracy. This advantage is highest during the peak period (1 October–10 October 2015), when no satellite data are available. In the case of Twitter data, the coverage is continuous for the entire period of the natural hazard and especially during the peak phase when the social media feeds are largest. Despite the fact that WorldView-3 satellite imagery is available shortly after the peak of the floods, during the receding phase, the cloud coverage prevents the generation of a precise flood extent map. Hence, in these cases, crowdsourced data are most critical to complement the satellite data where coverage gaps are caused by the clouds.

The proposed geostatistical analysis of the tweets shows an indirect flood assessment by providing the probability that specific areas are likely to become flooded. The results show that users' contribution is increased during the peak and the receding phases which are the most critical for the emergency response and mitigation. This is particularly true if the entire Twitter stream is analysed and not only the geolocated messages. The subset of tweets used in the analysis only represents a subset of the stream, namely the geographical location of where they are originated from. The main problem is that the number of geolocated tweets is only a small percentage (1–2%) of the total number of messages, and thus, they can skew the analysis towards specific geographical areas.
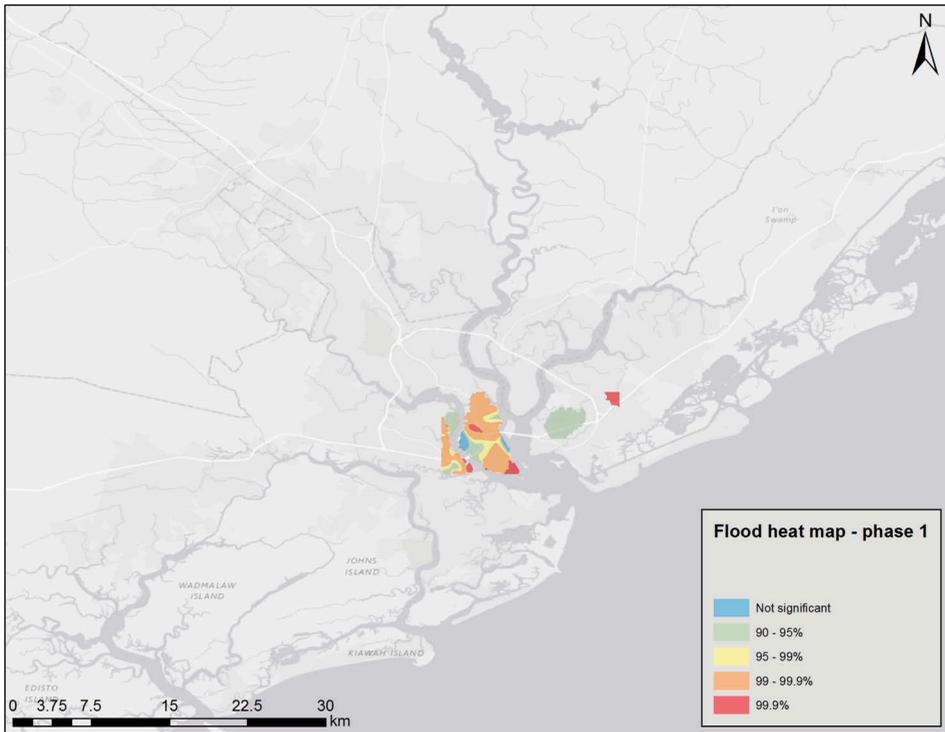
**Figure 5.** Heat map based on Twitter feeds during the pre-crisis phase crisis (27 September–30 September 2015), illustrating the probability of the surface that is flooded.
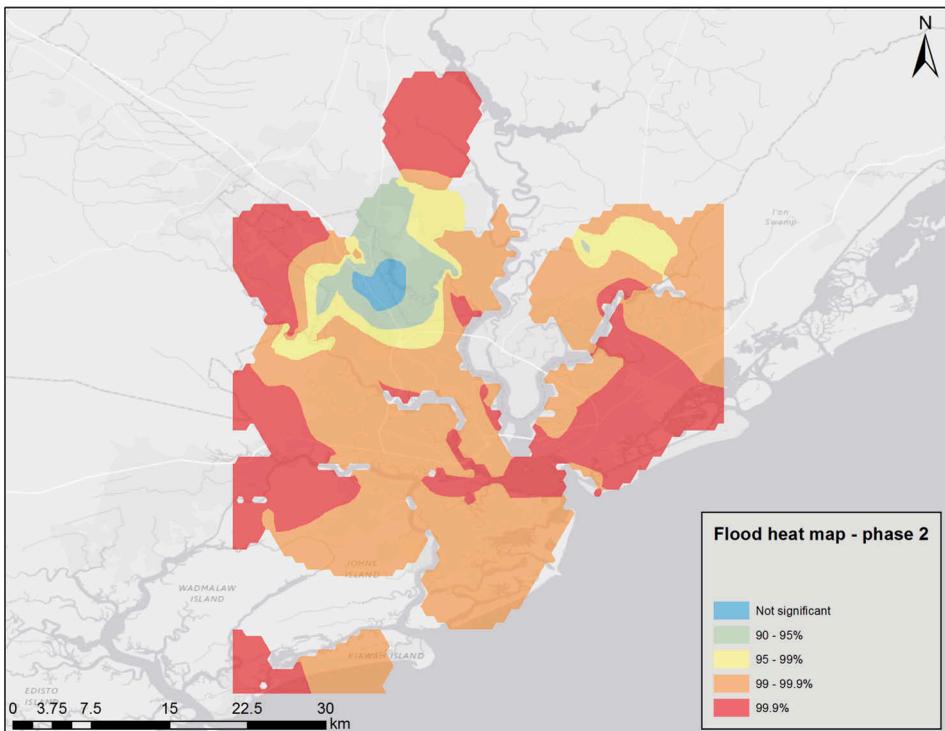


**Figure 6.** Heat map based on Twitter feeds during the peak phase (1 October–6 October 2015), illustrating the probability of the surface that is flooded.
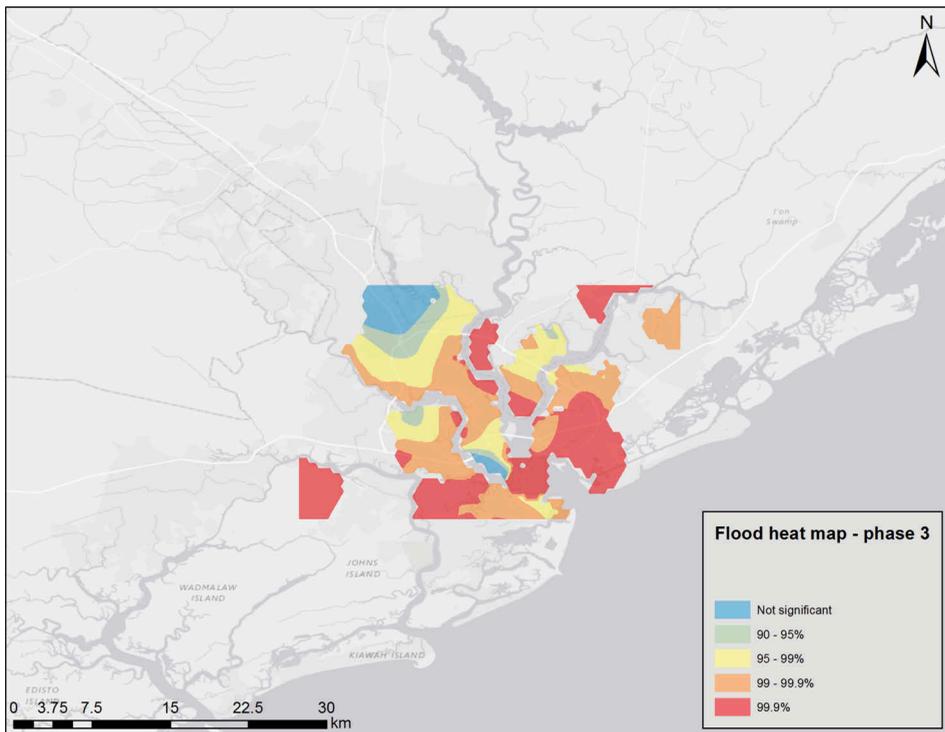
**Figure 7.** Heat map based on Twitter feeds during the receding phase (7 October–9 October 2015), illustrating the probability of the surface that is flooded.
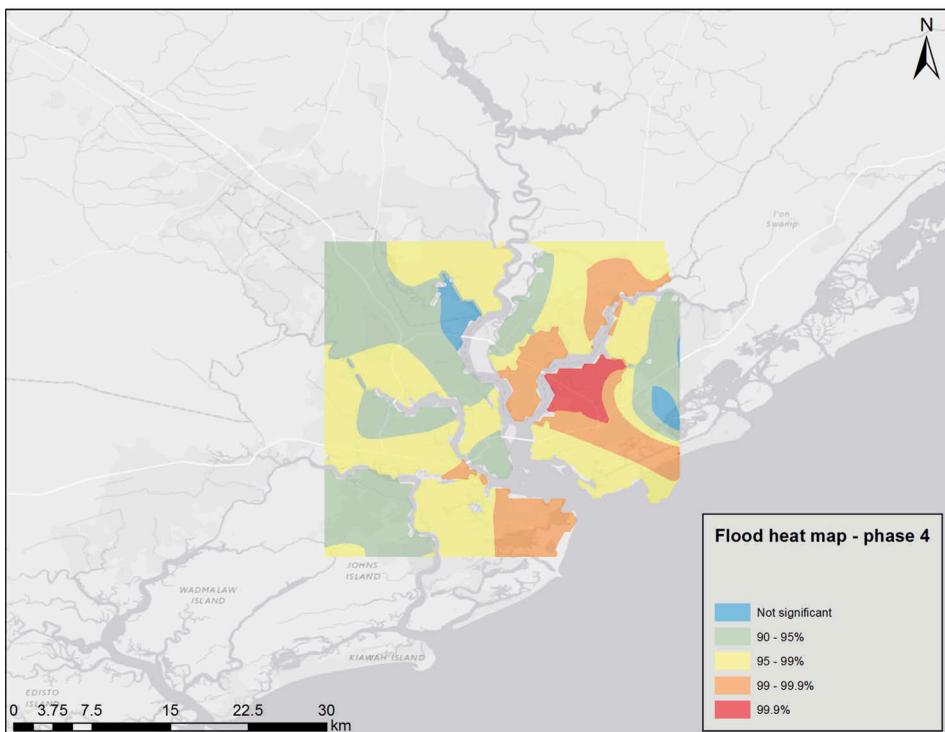


**Figure 8.** Heat map based on Twitter feeds during the recovery phase (10 October–18 October 2015), illustrating the probability of the surface that is flooded.

Despite the advantages and promising results show in this article, there are some draw-backs that need to be considered when fusing these two data sets. In the case of major hurricanes or other extreme events, mobile communication can be disrupted, thus limiting the scope of the methodology. It has been shown that in the case of Hurricane Sandy, which caused massive blackouts, social media use peaked during the event, but almost ceased in some areas due to the prolonged absence of power and network connections.

An additional consideration is that tweets might be uploaded by some users at a later time, thus decreasing the temporal resolution of the data. While this is true, the shear volume of the data is expected to continue providing a good coverage, despite the asynchronous availability of a portion of the data. The literacy and propensity to the use of technology in the population can also influence the timely availability of data. Some studies showed that there is a high correlation between socio-economic status and activity on social media.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

## References

Asante, K. O., R. D. Macuacua, G. A. Artan, R. W. Lietzow, and J. P. Verdin. 2007. "Developing a Flood Monitoring System from Remotely Sensed Data for the Limpopo Basin." *IEEE Transactions on Geoscience and Remote Sensing* 45 (6): 1709–1714. doi:10.1109/TGRS.2006.883147.

Burt, J. E., G. M. Barber, and D. L. Rigby. 2009. *Elementary Statistics for Geographers*. Guilford Press.

Cervone, G., E. Sava, Q. Huang, E. Schnebele, J. Harrison, and N. Waters. 2016. "Using Twitter for Tasking Remote-Sensing Data Collection and Damage Assessment: 2013 Boulder Flood Case Study." *International Journal of Remote Sensing* 37 (1): 100–124. doi:10.1080/01431161.2015.1117684.

Cervone, G., E. Schnebele, N. Waters, M. Moccaldi, and R. Sicignano. 2016. "Using Social Media and Satellite Data for Damage Assessment in Urban Areas during Emergencies." In *Seeing Cities Through Big Data*, edited by P. Thakuriah et al., 443–457. Springer.

Chander, G., B. L. Markham, and D. L. Helder. 2009. "Summary of Current Radiometric Calibration Coefficients for Landsat Mss, Tm, Etm+, and Eo-1 Ali Sensors." *Remote Sensing of Environment* 113 (5): 893–903. doi:10.1016/j.rse.2009.01.007.

Cossu, R., E. Schoepfer, P. Bally, and L. Fusco. 2009. "Near Real-Time Sar-Based Processing to Support Flood Monitoring." *Journal of Real-Time Image Processing* 4 (3): 205–218. doi:10.1007/s11554-009-0114-4.

Eilander, D., P. Trambauer, J. Wagemaker, and A. van Loenen. 2016. "Harvesting Social Media for Generation of Near Real-Time Flood Maps." *Procedia Engineering* 154: 176–183. doi:10.1016/j.proeng.2016.07.441.

Fan, H., A. Zipf, Q. Fu, and P. Neis. 2014. "Quality Assessment for Building Footprints Data on Openstreetmap." *International Journal of Geographical Information Science* 28 (4): 700–719. doi:10.1080/13658816.2013.867495.

Fan, J., and I. Gijbels. 1996. *Local Polynomial Modelling and Its Applications: Monographs on Statistics and Applied Probability*. 66 vols. CRC Press.

Gamba, P., F. DellAcqua, and B. V. Dasarathy. 2005. "Urban Remote Sensing Using Multiple Data Sets: Past, Present, and Future." *Information Fusion* 6 (4): 319–326. doi:10.1016/j.inffus.2005.02.007.

Goolsby, R., and G. Cervone. 2013. "Using Social Media to Fill the Gaps in Observations during Emergencies." In *Innovation*, edited by L. Schuette and C. Hughes, 11 vols., 19–22. Office of Naval Research. Winter.

Haining, R. P. 2003. *Spatial Data Analysis: Theory and Practice*. Cambridge University Press.

Huang, Q., and G. Cervone. 2016. "Usage of Social Media and Cloud Computing during Natural Hazards." In *Cloud Computing in Ocean and Atmospheric Sciences*, edited by T. C. Vance, N. Merati, C. Yang and M. Yuan, 297–324. Academic Press.

Huang, Q., and Y. Xiao. 2015. "Geographic Situational Awareness: Mining Tweets for Disaster Preparedness, Emergency Response, Impact, and Recovery." *ISPRS International Journal of Geo-Information* 4 (3): 1549–1568. doi:10.3390/ijgi4031549.

Joyce, K. E., S. E. Belliss, S. V. Samsonov, S. J. McNeill, and P. J. Glassey. 2009. "A Review of the Status of Satellite Remote Sensing and Image Processing Techniques for Mapping Natural Hazards and Disasters." *Progress in Physical Geography* 33: 183–207. doi:10.1177/0309133309339563.

Kennedy, R. E., P. A. Townsend, J. E. Gross, W. B. Cohen, P. Bolstad, Y. Wang, and P. Adams. 2009. "Remote Sensing Change Detection Tools for Natural Resource Managers: Understanding Concepts and Tradeoffs in the Design of Landscape Monitoring Projects." *Remote Sensing of Environment* 113 (7): 1382–1396. doi:10.1016/j.rse.2008.07.018.

Khan, S. I., Y. Hong, J. Wang, K. K. Yilmaz, J. J. Gourley, R. F. Adler, G. R. Brakenridge, F. Policelli, S. Habib, and D. Irwin. 2011. "Satellite Remote Sensing and Hydrologic Modeling for Flood Inundation Mapping in Lake Victoria Basin: Implications for Hydrologic Prediction in Ungauged Basins." *IEEE Transactions on Geoscience and Remote Sensing* 49 (1): 85–95. doi:10.1109/TGRS.2010.2057513.

Kuo, P.-F., X. Zeng, and D. Lord. 2011. "Guidelines for Choosing Hot-Spot Analysis Tools Based on Data Characteristics, Network Restrictions, and Time Distributions." In *Proceedings of the 91 Annual Meeting of the Transportation Research Board*, 22–26.

Kussul, N. N., A. Y. Shelestov, S. V. Skakun, G. Li, and O. M. Kussul. 2012. "The Wide Area Grid Testbed for Flood Monitoring Using Earth Observation Data." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 5 (6): 1746–1751. doi:10.1109/JSTARS.2012.2201447.

Li, W., Z. Du, F. Ling, D. Zhou, H. Wang, Y. Gui, B. Sun, and X. Zhang. 2013. "A Comparison of Land Surface Water Mapping Using the Normalized Difference Water Index from Tm, Etm+ and Ali." *Remote Sensing* 5 (11): 5530–5549. doi:10.3390/rs5115530.

MacEachren, A. M., A. Jaiswal, A. C. Robinson, S. Pezanowski, A. Savelyev, P. Mitra, X. Zhang, and J. Blanford. 2011. "Senseplace2: Geotwitter Analytics Support for Situational Awareness." In *Visual Analytics Science and Technology (VAST), 2011 IEEE Conference On*, 181–190. IEEE.

McFeeters, S. K. 1996. "The Use of the Normalized Difference Water Index (Ndwi) in the Delineation of Open Water Features." *International Journal of Remote Sensing* 17 (7): 1425–1432. doi:10.1080/01431169608948714.

Mukkamala, A., and R. Beck. 2016. "Enhancing Disaster Management through Social Media Analytics to Develop Situation Awareness What Can Be Learned from Twitter Messages about Hurricane Sandy?." In *Proceedings of the 2016 Pacific Asia Conference on Information Systems (PACIS)*, 1 vols, 11.

Musser, J. W., K. M. Watson, J. A. Painter, and A. J. Gotvald. 2016. *Flood-Inundation Maps of Selected Areas Affected by the Flood of October 2015 in Central and Coastal South Carolina*. Technical report, US Geological Survey.

Ord, J. K., and A. Getis. 1995. "Local Spatial Autocorrelation Statistics: Distributional Issues and an Application." *Geographical Analysis* 27 (4): 286–306. doi:10.1111/j.1538-4632.1995.tb00912.x.

Peeters, A., M. Zude, J. Käthner, M. Ünlü, R. Kanber, A. Hetzroni, R. Gebbers, and A. Ben-Gal. 2015. "Getis–Ord's Hot-And Cold-Spot Statistics as a Basis for Multivariate Spatial Clustering of

Orchard Tree Data." *Computers and Electronics in Agriculture* 111: 140–150. doi:10.1016/j.compag.2014.12.011.

Schnebele, E., and G. Cervone. 2013. "Improving Remote Sensing Flood Assessment Using Volunteered Geographical Data." *Natural Hazards Earth System Science* 13: 669–677. doi:10.5194/nhess-13-669-2013.

Schumann, G., G. Di Baldassarre, and P. D. Bates. 2009. "The Utility of Spaceborne Radar to Render Flood Inundation Maps Based on Multialgorithm Ensembles." *IEEE Transactions on Geoscience and Remote Sensing* 47 (8): 2801–2807. doi:10.1109/TGRS.2009.2017937.

Serpico, S. B., S. Dellepiane, G. Boni, G. Moser, E. Angiati, and R. Rudari. 2012. "Information Extraction from Remote Sensing Images for Flood Monitoring and Damage Evaluation." *Proceedings of the IEEE* 100 (10): 2946–2970. doi:10.1109/JPROC.2012.2198030.

Skakun, S. 2012. "A Neural Network Approach to Flood Mapping Using Satellite Imagery." *Computing and Informatics* 29 (6): 1013–1024.

Soltani, K., A. Soliman, A. Padmanabhan, and S. Wang. 2016. "Urbanflow: Large-Scale Framework to Integrate Social Media and Authoritative Landuse Maps." In *Proceedings of the XSEDE16 Conference on Diversity, Big Data, and Science at Scale*, 2. ACM.

Taubenböck, H., M. Wurm, M. Netzband, H. Zwenzner, A. Roth, A. Rahman, and S. Dech. 2011. "Flood Risks in Urbanized Areas–Multi-Sensoral Approaches Using Remotely Sensed Data for Risk Assessment." *Natural Hazards and Earth System Sciences (NHESS)* 11: 431–444. doi:10.5194/nhess-11-431-2011.

Van der Sande, C., S. De Jong, and A. De Roo. 2003. "A Segmentation and Classification Approach of Ikonos-2 Imagery for Land Cover Mapping to Assist Flood Risk and Flood Damage Assessment." *International Journal of Applied Earth Observation and Geoinformation* 4 (3): 217–229. doi:10.1016/S0303-2434(03)00003-5.

Wan, Z., Y. Hong, S. Khan, J. Gourley, Z. Flamig, D. Kirschbaum, and G. Tang. 2014. "A Cloud-Based Global Flood Disaster Community Cyber-Infrastructure: Development and Demonstration." *Environmental Modelling & Software 58* 58: 86–94. doi:10.1016/j.envsoft.2014.04.007.

Xu, H. 2006. "Modification of Normalised Difference Water Index (Ndwi) to Enhance Open Water Features in Remotely Sensed Imagery." *International Journal of Remote Sensing* 27 (14): 3025–3033. doi:10.1080/01431160600589179.

Yin, J., A. Lampert, M. Cameron, B. Robinson, and R. Power. 2012. "Using Social Media to Enhance Emergency Situation Awareness." *IEEE Intelligent Systems* 27 (6): 52–59. doi:10.1109/MIS.2012.6.