Using Nightlight Remote Sensing Imagery and Twitter Data to Study Power Outages

Carolynne Hultquist*
Geoinformatics and Earth Observation Laboratory
Department of Geography and Institute for CyberScience
The Pennsylvania State University
hultquist@psu.edu

Mark Simpson
Department of Geography
Human Factors and GIScience Laboratory
The Pennsylvania State University
marksimpson@psu.edu

Guido Cervone
Geoinformatics and Earth Observation Laboratory
Department of Geography and Institute for CyberScience
The Pennsylvania State University
cervone@psu.edu

Qunying Huang
Department of Geography
University of Wisconsin-Madison
qhuang46@wisc.edu

ABSTRACT
Hurricane Sandy made landfall in one of the most populated areas of the United States, and affected almost 8 million people. The event provides a unique opportunity to study power outages because of the data available and the large impact to a densely populated area. Satellite nightlight imagery of "before" and "after" the landfall of the hurricane is used to quantify the light dimming caused by power outages. Geolocated tweets filtered by keywords provide valuable information on human activity at a high temporal and spatial resolution during the event. Analysis of brightness change in the satellite data and the density of power related tweets points to a spatial relationship that identifies severely impacted areas with human presence. Classification of tweets through text analysis serves to further narrow the information search to find the most relevant and reliable content. Twitter data fused with satellite imagery identifies power outage information at a street-level resolution that is not achievable with satellite imagery alone.

Categories and Subject Descriptors
H.2.8 [**Database Management**]: Database Applications—*data mining, spatial databases and GIS.*

General Terms
Management, Human Factors, Verification.

*Corresponding Author

Conference'15, November 3-6, 2015, Seattle, WA, USA.

Keywords
Disaster management, Remote sensing, Social media, Power outages, Hurricane Sandy.

1. INTRODUCTION
Natural hazards are severe events that pose a threat to the sustainment and survival of our society. Drastic population growth, the emergence of megacities, and high risk facilities such as high dams and nuclear power plants have increased the risk posed by natural hazards at unprecedented levels [15]. A single catastrophic event can claim thousands of lives, cause billions of dollars of damage, trigger a global economic depression, destroy natural landmarks, render a large territory uninhabitable, and destabilize the military and political balance in a region [10]. A number of costly hurricanes have made landfall in populated areas of the United States in recent years. In 2012, Hurricane Sandy hit the Eastern seaboard of the United States in the densely populated areas of New Jersey and New York. Major economic consequences resulted through physical damage from the hurricane, as well as from widespread power outages.

During and after a natural disaster, situational awareness is paramount to assess damage and determine appropriate responses. Access to remote sensing data is critical during disasters and has become the de-facto standard by providing high resolution imagery for damage assessment and the coordination of disaster relief operations [6, 9]. Satellites often can produce high resolution imagery within hours of major events [14, 2]. Satellite remote sensing data is used to provide high spatial resolution imagery for areas of poor accessibility or that are lacking in ground measurements [13]. Traditionally, geospatial information from remote sensing is relied upon for decision-making. However, in the case of a hurricane, it is difficult to use satellite imagery to assess damage due to significant cloud coverage from the storm. In addition, remote sensing only provides physical observations of the Earth and can not directly assess the human condition

Figure 1: Nightlight image taken under normal conditions about a month prior to landfall (2:11 am EDT on 31 August 2012).

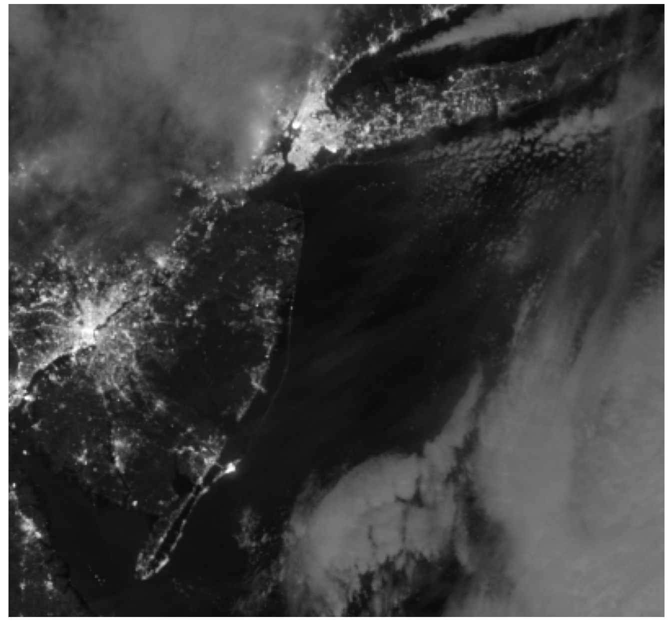

Figure 2: Nightlight image taken two days after Hurricane Sandy made landfall, during a short break in cloud cover (2:52 am EDT on 1 November 2012).

in that environment. Integration of other sources of spatially rich crisis information on human activity and interests could benefit emergency management.

Social media streams are a novel source of data being considered to provide actionable information during emergencies [12, 3]. There is not widespread practitioner acceptance of social media as a source of information for disaster management. Some relief organizations, such as the American Red Cross, use social media to supplement authoritative information and official news reporting about events, as well as, respond with information to those who need help [5]. The American Red Cross uses social media as an indicator that a certain event is occurring at a certain place based on keywords. The premise of many academic studies is that geolocated social media provides spatially relevant information.

However, social media data is generally aggregated and rarely validated with another source. Even so, recent studies purport there is valuable information in open data when properly analyzed [5]. The integration of multiple forms of data provides a collective check on the reliability of the spatial information. Emergency responders commonly have access to Geographical Information Systems (GIS) that enable them to fuse spatial data from multiple sources, and analyze spatial patterns that may be present. Geolocated social media can provide a GIS with temporally and spatially rich information on human activity and behavior that can be added to existing datasets. Social media can be monitored to identify spatial patterns of concern such as human activity in evacuation zones and areas heavily damaged. Twitter data has been used to supplement or replace imagery when it is not available [12]. Social media for natural disasters is increasingly used as a source of information from a population, especially in urban areas [3].

This study demonstrates the ability for spatial analysis in GIS by the fusion of social media with nightlight remote sensing imagery to detect power outages at a street-level resolution. By comparing geolocated tweets of power outage keywords to nightlight imagery of power outages we can form a methodology for future social media usage when imagery is not available. Identification of areas of severe impact can be further validated by text analysis classification and reviewing tweet content. The evaluation shows that real-time social media event data during natural disasters can supplement existing geospatial technologies with validated information.

2. CASE STUDY

In 2012, Hurricane Sandy had a devastating effect on the East coast of the United States causing a loss of lives, properties, and damage to the environment. The storm claimed a total 117 people in the United States: 53 in New York state, 34 in New Jersey, 12 in Pennsylvania, and 18 elsewhere in the United States [4]. Economic costs were estimated in the billions; immediately following Sandy, New Jersey estimated related costs at $36.8 billion and New York state with estimates amounting to $41.9 billion. Environmental concerns of severe shoreline erosion and inundation of toxic floodwaters have been observed [4].

Social media collection began on 26 October 2012, the day some states declared a state of emergency as the category 1 hurricane approached with winds of 80 mph [4]. The following day New York City ordered evacuation of low-lying areas and closed public schools. Hurricane Sandy approached the Eastern seaboard, with hurricane force winds reaching 175 miles from the eye and cloud coverage

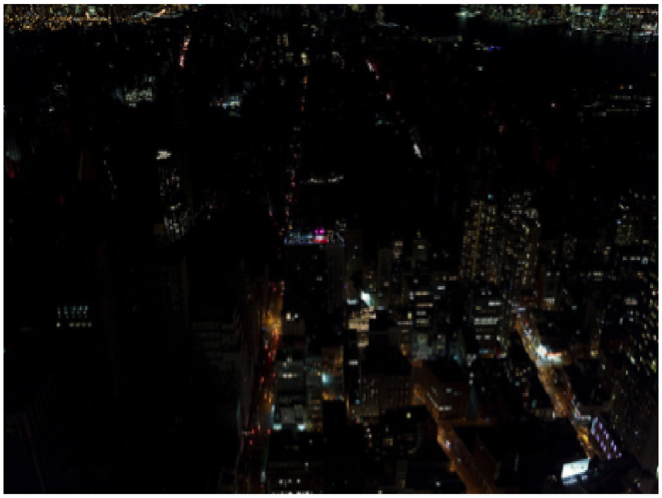

Figure 3: Image associated with a tweet taken from the Empire State Building looking south shows the entire lower end of Manhattan without power.

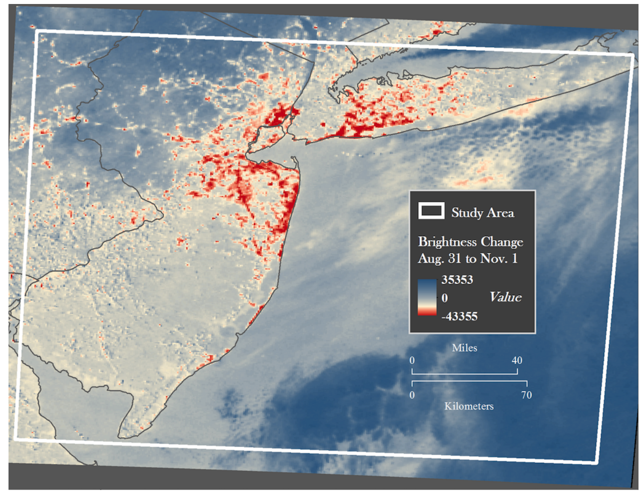

Figure 4: A change detection analysis between the "before" and "after" satellite images shows several areas of reduced brightness due to the power outages.

of a thousand miles. On 30 October 2012, there were electrical outages of 7.9 million businesses and households in 15 states and the District of Colombia [4]. The storm caused major infrastructural damage with high winds and flooding which shutdown the New York Stock Exchange, the subway system, and local airports. Power outages due to the landfall of Hurricane Sandy doubled the rate of Internet network outages in the United States and the recovery of networks took four days before reaching normal levels [7]. Recovery of infrastructure was relatively quick as only 600,000 people were without electricity on 7 November 2012.

3. DATA

3.1 Remote Sensing

Two low-light images from the Visible Infrared Imaging Radiometer Suite (VIIRS) on NASA/NOAA's Suomi National Polar-orbiting Partnership (NPP) satellite were obtained, the first about one month prior under normal conditions (Figure 1) and the latter taken two days after landfall during a short break in cloud cover (Figure 2). The "day-night band" captures wavelengths from green to near-infrared and measures the intensity of night light emissions [11].

Nightlight imagery is a reliable source of information on light emissions only if clear imagery is available. In addition, nightlight imagery is available at a low temporal resolution due to the uniqueness of the satellite collection and its orbit. Power outage maps can only be created by change detection methods if there is previously collected imagery of normal conditions. Changes in nightlight levels can occur for many reasons, thus it is important to have temporally comparable imagery to limit extraneous effects such as seasonal variation.

In the case of Hurricane Sandy, ideal nightlight imagery is available for the study area due to a chance break in the otherwise heavy post-hurricane cloud cover. However, in other cases, such as the Indian Blackout of 2012 [1], clear nightlight imagery may not be available during the power outage. It should be noted that even this otherwise largely pristine imagery was taken days after the maximum extent of the power outages, on 1 November 2012. Lack of imagery at critical times demonstrates the need to augment gaps in remote sensing data with other sources of spatial information.

3.2 Social Media

Real-time collection of social media is the most reliable method for gaining a sufficient quantity of streaming tweets during an event. This study uses two Twitter datasets that provide the most dense data for their spatial extents. The first dataset contains about 70,000 geolocated tweets collected by the Pennsylvania State University during the event from 26 October 2012 to 12 November 2012 for an area that encompasses New York City and most of New Jersey with keywords such as "hurricane", "Sandy", or "storm". Geolocated tweets related to power outages were identified by mining the text of the tweets for keywords such as "power", "blackout", "electric", "lights", and "outage". Power outage related tweets can indicate where the communication of power outages is the most dense given the quantity of tweets in an area.

The second Twitter dataset has about 35,000 geotagged tweets that contain disaster relevant information collected by the University of Wisconsin-Madison for the same temporal period, but only focused on the extent of New York City. The geolocated Twitter data were classified into different categories (or themes) during various disaster phases using machine learning techniques [8]. Specifically, 17 sub-categories were created in the four major emergency management categories of preparedness, impact, response, and recovery. Additionally, the "other" category is defined to describe tweets that include the predefined disaster relevant hashtags and keywords in the text but do not contribute to situational awareness. The category of "utilities" includes

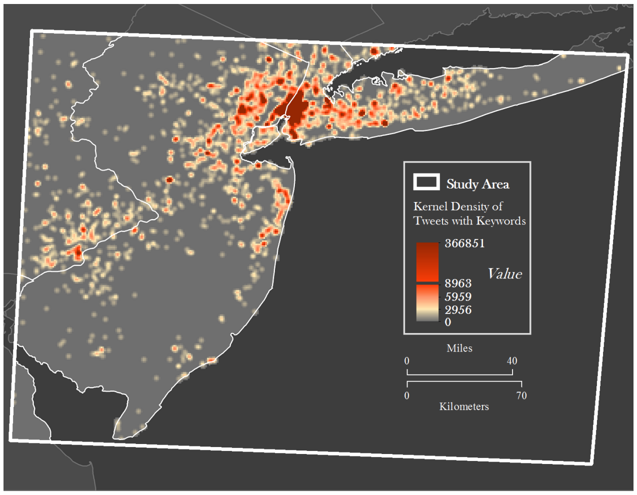

Figure 5: Kernel density raster of tweets in study area using power outage keywords.

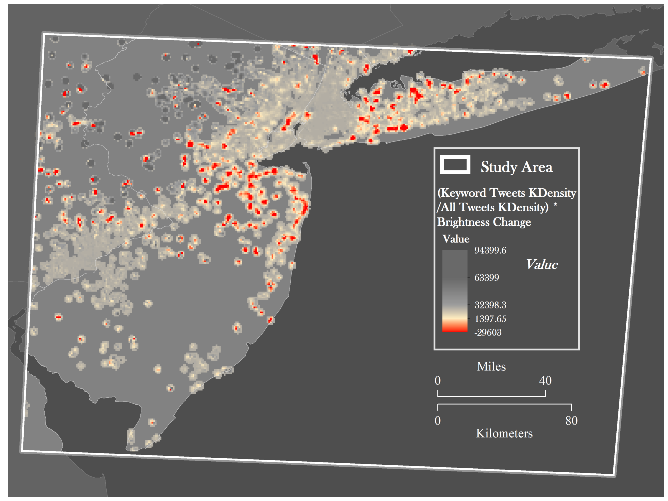

Figure 6: Highlights areas of overlap from high proportion of power-tagged tweets and remotely-sensed power outages.

situational awareness tweets that report information about power outages, in addition to, the failure or disaster impact on Internet, cable, Wi-Fi, and heating. All of these topics included in the "utilities" category are directly related to the impact of power outage which broadens the selection of relevant tweets in New York City.

The first dataset is used to study the overall region of New York and New Jersey in order to identify regions of spatial correlation between the Twitter dataset and satellite imagery. The second dataset is used to correlate the data at a street-level resolution in lower Manhattan and check the classification of topic categories for "utilities" that refer to power outages. This second dataset includes several tweets with links to photos which are tagged with power outage related keywords. Figure 3 shows a photo taken from the Empire State Building on 30 October 2012 directed towards the south which gives a view of the entire lower end of Manhattan without power. The main topic of this particular Manhattan geotagged tweet was accurately classified as "impact" that is crucial for disaster response with its subtopic as "utilities".

4. RESULTS

Nightlight imagery for the study area from pre- and post- Hurricane Sandy enables an analysis of change in light emissions occurring due to the event. Areas of reduced brightness due to the storm are calculated through change detection to produce the brightness difference raster (Figure 4). The areas showing the deepest red color denote raster cells that most significantly decreased in brightness after the effects of Hurricane Sandy. These areas are known to have significant populations that suffered power outages as a result of the storm. The analysis of geolocated tweets show that there are more tweets using power related key terms within the study area than outside the extent. The communication on social media about power outages and the condition of utilities is more relevant to the population that is directly involved in the incident.

In order to use GIS to compare the spatial aggregation of the vector points of tweets to the raster-based satellite imagery, two kernel density rasters were created from the tweets with the same pixel size and extent as the imagery; one of power-related tweets (Figure 5) and one of all tweets in the study area. This essentially converts the density of geolocated tweets into the same geographic data type as the imagery for comparison purposes. Another kernel density raster of tweets was made by dividing the quantity of power-related tweets raster by a raster of the quantity all tweets, yielding a raster that shows the proportion of power-related tweets over the area. The proportional representation identifies raster areas that do not only have large populations using Twitter to discuss the impact of Hurricane Sandy, but also have tweets communicating information specifically regarding power outages.

The proportional density raster of power outage related tweet is used to adjust for social media population when comparing the spatial distribution of Twitter to other datasets. Figure 6 shows a spatial comparison calculated by multiplying the pixel values of the brightness difference raster and the proportion of power outage tweets kernel density raster. This highlights areas with high density of power-related tweets and a drop-off in imagery observed brightness (indicative of a power outage). The spatial overlap in the comparison of the brightness change detection imagery and the kernel density of power related tweets suggests that an aggregation of the tweets accurately identified areas of power outage. Specifically, the southern tip of Manhattan was identified as a region of human activity with dense power related keyword tweets and one of the most extreme decreases in brightness.

The identified areas of power outages are further verified by classifying tweets for specific regions using text analysis. The goal of this analysis is to identify at a street-level resolution the areas in which both satellite data and tweets indicate power outages in an urban environment. For this

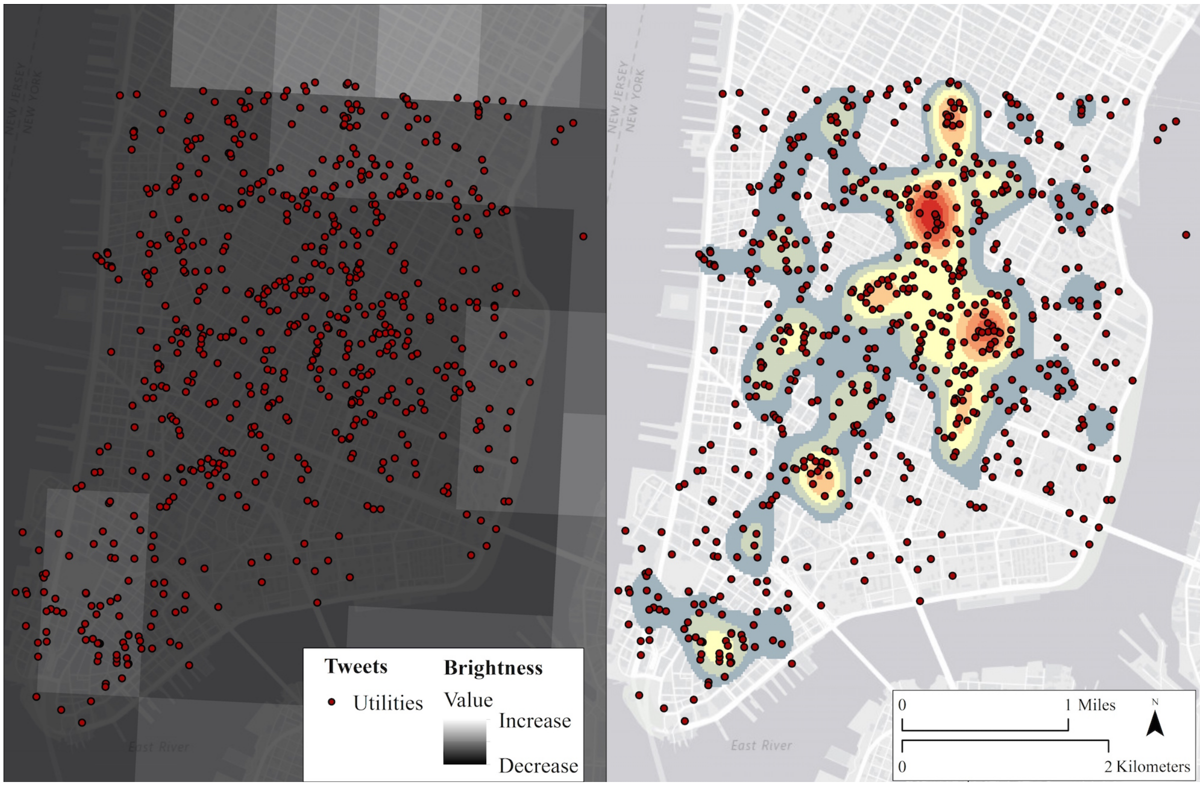

Figure 7: Power outage observed by imagery and tweets in Lower Manhattan: (left) Geotagged tweets classified as "utilities" overlaid on the nightlight brightness change raster and (right) Kernel density of "utilities" with several hotspots detected in areas of decreased brightness.

analysis, the second Twitter dataset is used due to its high density of tweets in New York City. A spatial analysis was performed only for tweets geotagged in lower Manhattan, and for which the logistic regression assigned a topic of "impact" and a subtopic of "utilities" based on the content of the tweet.

Tweets classified as "utilities" are specifically of interest due to the power outage related information they contain. Figure 7 (left) shows the distribution of tweets classified as "utilities" overlaid on nightlight imagery. The darkest raster cells observed are those showing the greatest decrease in brightness within the nightlight imagery change detection with transparency applied for locational awareness. Figure 7 (right) emphasizes the spatial distribution of the density of utility tweets with a scale of low value blue to high value red. The high density "hotspots" of the "utilities" tweets is observed to be within areas of the greatest decrease of brightness in the nightlight imagery.

The nightlight satellite imagery identified the entire lower tip of Manhattan as being dimmed compared to the pre-event imagery. The Twitter data, interpolated using a Gaussian kernel estimator, also identified the same region during the event. However, the spatial resolution of the low light satellite data is too coarse to perform a street-level resolution analysis. Regardless, the imagery corresponding spatially to the classification gives moderate certainty that the information communicated through the geolocated tweet is relevant to the location and the event. Text analysis accurately classified tweets which can serve to narrow the information search to find the most relevant and reliable information. Spatially dense social media data provides street-level coverage to determine power outages in an urban area.

5. CONCLUSIONS

This paper presents an application of spatial interpolation of Twitter data and nightlight satellite imagery for the identification of areas with power outage during a natural hazard. The fusion of these two datasets can lead to street-level identification of power outages, and it is especially effective in areas with high population density where social media data are available at a large volume. Augmenting traditional sources of data with real-time information on human activity could improve situational awareness. In fact, geolocated tweets filtered by keywords can be used to identify damaged areas which have been fully evacuated, versus areas where there is still significant human presence.

Fusion of spatial data can provide greater reliability for location specific information for use during emergencies. Social media is increasingly being monitored and analyzed by emergency management without consideration of spatial relationships within and between datasets. A GIS methodology can be adopted by emergency management to identify the spatial relationships of observations through change detection, raster comparison, and kernel density.

Spatial processes built into GIS can be used to quickly convert geolocated social media points to a raster grid that aligns to imagery in order to identify areas of interest.

Social media streams are available at high temporal resolution but with sparse spatial resolution unless aggregated, while remote sensing data are available with high spatial resolution and coverage dependent on atmospheric conditions. Nightlight imagery can be used to indicate power outages, but because of the coarse pixel resolution it can only provide indicators for regions. Nightlight imagery alone is not able to identify street-level resolution of power outages and Twitter alone provides an overwhelming volume of information. By merging imagery and social media analysis the relevant information is filtered to the identified spatial areas with greater reliability.

6. REFERENCES

[1] M. Carlowicz. Moonlit Clouds Obscure Indian Blackout. In *NASA Earth Observatory*. Aug. 2012.

[2] G. Cervone and G. Manca. Damage assessment of the 2011 Japanese tsunami using high-resolution satellite data. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 46(3):200–203, 2011.

[3] G. Cervone, E. Schnebele, N. Waters, J. Harrison, M. Moccaldi, and R. Sicignano. Using social media to task data collection and augment observations in urban areas during emergencies: 2013 boulder floods case study. In *Proceedings of Big Data for Urban Informatics Conference (BDUIC)*, Chicago, IL, August 11-12 2014.

[4] CNN. Hurricane Sandy Fast Facts, July 2013.

[5] J. Crowley. Connecting Grassroots and Government for Disaster Response. Technical report, Woodrow Wilson International Center for Scholars Commons Lab, Washington D.C., 2013.

[6] S. L. Cutter. Giscience, disasters, and emergency management. *Transactions in GIS*, 7(4):439–446, 2003.

[7] J. Heidemann, L. Quan, and Y. Pradkin. A Preliminary Analysis of Network Outages During Hurricane Sandy. Technical Report November, USC/ISI Technical Report ISI-TR-685, 2012.

[8] Q. Huang and Y. Xiao. Geographic situational awareness: mining tweets for disaster preparedness, emergency response, impact, and recovery. *ISPRS International Journal of Geo-Information*, 4(3):1549–1568, 2015.

[9] K. E. Joyce, S. E. Belliss, S. V. Samsonov, S. J. McNeill, and P. J. Glassey. A review of the status of satellite remote sensing and image processing techniques for mapping natural hazards and disasters. *Progress in Physical Geography*, 33(2):183–207, 2009.

[10] V. Keilis-Borok. Earthquake prediction: State-of-the-art and emerging possibilities. *Annual Review of Earth and Planetary Science*, 30:1–33, 2002.

[11] S. D. Miller, W. Straka, S. P. Mills, C. D. Elvidge, T. F. Lee, J. Solbrig, A. Walther, A. K. Heidinger, and S. C. Weiss. Illuminating the Capabilities of the Suomi National Polar-Orbiting Partnership (NPP) Visible Infrared Imaging Radiometer Suite (VIIRS) Day/Night Band. *Remote Sensing*, (5):6717–6766, 2013.

[12] E. Schnebele, G. Cervone, and N. Waters. Road assessment after flood events using non-authoritative data. *Natural Hazards and Earth System Science*, 14(4):1007–1015, 2014.

[13] L. Smith. Satellite remote sensing of river inundation area, stage, and discharge: A review. *Hydrological Processes*, 11(10):1427–1439, 1997.

[14] T. Stryker and B. Jones. Disaster response and the international charter program. *Photogrammetric Engineering and Remote Sensing*, 75(12):1342–1344, 2009.

[15] C. Tate and T. Frazier. A GIS Methodology to Assess Exposure of Coastal Infrastructure to Storm Surge & Sea-Level Rise: A Case Study of Sarasota County, Florida. *Journal of Geography & Natural Disasters*, 1:2167–0587, 2013.