ABSTRACT
This research proposes a methodology that leverages non-authoritative data to augment flood extent mapping and the evaluation of transportation infrastructure. The novelty of this approach is the application of freely available, non-authoritative data and its integration with established data and methods. Crowdsourced photos and volunteered geographic data are fused together using a geostatistical interpolation to create an estimation of flood damage in New York City following Hurricane Sandy. This damage assessment is utilized to augment an authoritative storm surge map as well as to create a road damage map for the affected region.

Categories and Subject Descriptors
J.2 [Physical Sciences and Engineering]: Earth and atmospheric sciences

General Terms
Management

Keywords
Road Assessment, Remote Sensing, Natural Hazards

1. INTRODUCTION
Accurate and timely flood assessments are critical during all phases of a flood disaster. Knowledge of road conditions and accessibility is especially important for emergency managers, first responders, and residents. Over the past two decades, satellite remote sensing has become the standard technique to determine flood extent. Satellite remote sensing data provide high spatial resolution even for areas of poor accessibility or lacking in ground measurements [18]. However, in some events, particularly hurricanes, high resolution satellite data might be unavailable for days because of cloud cover or orbit revisit time.

Satellite data are often supplemented with additional data, such as digital elevation models (DEM) and river gauge data, to provide a more comprehensive flood assessment [27, 1]. RADAR data, in particular, are a good resource for flood identification because of the capability to distinguish water bodies from other land cover while penetrating through vegetative canopy and cloud cover [12, 24]. Because the application of RADAR data can be difficult due to limited swaths and long revisit times, there are many recent efforts for increasing RADAR’s availability and accessibility. For example, [10] illustrate how a RADAR instrument on an unmanned aerial vehicle (UAV) can be used for flood assessment of targeted areas. [19] propose a multi-sensor approach by combining satellite, aerial, and ground data for a more accurate flood assessment. They test how a RADAR sensor onboard a UAV can provide useful data. Aerial platforms, both manned and unmanned, are particularly suited for coastal monitoring after major catastrophic events because they can fly below the clouds, and thus acquire data in a targeted and timely fashion.

Remote sensing data are also used to catalog damages to the built environment. For the evaluation of transportation infrastructure following Hurricane Katrina, a variety of assessment techniques were utilized including visual, non-destructive, and remote sensing. However, the assessment of transportation infrastructure over such a large area could have been accelerated through the use of high resolution imagery and geospatial analysis [25].

Recent studies have focused on the application of remote sensing data after earthquakes or flooding specifically to assess transportation networks. [2] used multi-sensor, multi-temporal imagery to identify flooded roads. [4] identified infrastructure and road damages after the 2008 Wenchuan earthquake, using pre- and post-disaster very high resolution (VHR) optical imagery (1m or better). The combination of optical satellite imagery with a DEM to assess roads for accessibility after flooding was used to create a model for application in near-real time for emergency managers [6].

The integration of new data sources and methods with traditional approaches can provide additional information regarding on-the-ground conditions. For example, non-authoritative data are data not collected or distributed by
traditional, authoritative emergency management methods or agencies. They are generated, and often distributed, by public citizens and can offer additional insight during and after hazard events. Volunteered geographic information (VGI) is an emerging and quickly growing data source [8]. These data are voluntarily contributed, made available, and contain temporal and spatial information. The sources of VGI vary greatly and include pictures, videos, sounds, text messages, etc. An unprecedented and massive amount of ground data have become available through VGI, often in real-time.

Although non-authoritative data usually carry little scientific merit, it is still possible for them to yield useful information. VGI have been evaluated during disaster and crisis events as a source of situational awareness or as documentation of an event’s progression over time [3, 26]. Volunteered data have also been utilized specifically during flood events. For rapid flood damage estimation, [16] interpolated flood inundation depth from VGI and found estimates to be comparable to interpolated in situ measurements as well as model predictions. [13] estimated flood extent by using VGI and river gauge data to create a DEM as a comparison to the natural topographic surface.

Another source of non-authoritative, volunteered information harnesses the power of group contribution, or the ‘wisdom of crowds’ [22]. Crowdsourcing, a process where a task is undertaken by a large group of people rather than by a single individual or expert, can result in successful problem solving [11]. Examples of successful crowdsourcing include Wikipedia and Open Street Map, where information is voluntarily contributed and the public manages content and errors. [9] found the use of crowdsourcing during disasters to provide valuable information, although data quality was a concern.

Because of uncertainty in non-authoritative data, they have yet to be regularly, systematically applied during large scale disasters [5, 23]. But, despite their non-scientific nature, they can offer new and additional information which harnesses the power of ‘citizens as sensors’ and ‘wisdom of crowds’ to fill in gaps in the data infrastructure [22, 8, 21].

This paper utilizes crowdsourced aerial remote sensing data along with volunteered geographic data for flood damage assessment and the identification of road damages in the New York City area following Hurricane Sandy. Hurricane Sandy was a major storm which impacted a large portion of the US East coast in October 2012 with damages and recovery costs estimated to be between $50-$60 billion.\(^1\)

2. DATA

2.1 Non-authoritative data

2.1.1 Volunteered geographic data

Geolocated videos which documented flooding and damages were collected from a Hurricane Sandy Google Earth website where YouTube videos supplied by Storyful could be accessed.\(^2\)

2.1.2 Crowdsourced data

\(^1\)http://www.washingtonpost.com
\(^2\)https://www.storyful.com

Figure 1: Crowsourced assessments for the Civil Air Patrol data. Damage assessment: red=high, yellow=medium, green=none.

2.2 Authoritative data

The FEMA Modeling Task Force (MOTF) created storm surge maps for the US East Coast following Hurricane Sandy from field-verified high water marks and storm surge sensor data. FEMA employed these data along with a digital elevation model (DEM) to create a surge boundary for each state.

A FEMA MOTF shapefile was downloaded from FEMA’s GeoPlatform website and imported into ArcGIS 10 for analysis.\(^5\) The shapefile utilized for this research was the finalized version (dated February 14, 2013) for New York City with a 1 meter horizontal resolution (Figure 2a).

2.3 Road layer

A 2012 TIGER/line\(^\circ\) shapefile of road networks for the New York City area was downloaded from the US Census Bureau.\(^6\) The layer was georeferenced to New York State Plane coordinates in ArcGIS 10.

3. METHODOLOGY

\(^3\)http://google.org/crisismap/sandy-2012
\(^4\)http://mapmill.org
\(^5\)http://fema.maps.arcgis.com
\(^6\)http://www.census.gov
Storm surge created by FEMA MOTF for New York City.

Locations of Civil Air Patrol photos and geolocated videos documenting flooding.

Damage assessment generated from non-authoritative data within FEMA surge boundary.

Road damage assessment based on analysis of non-authoritative data.

Figure 2: Storm surge extent generated by FEMA and the locations of Civil Air Patrol photos and geolocated videos (a and b). Flood damage assessment generated from non-authoritative data and the subsequent classification of potential road damages (c and d).
3.1 Overview

This work is based on the fusion of non-authoritative data and its integration with traditional authoritative sources. Figure 3 illustrates the proposed methodology where non-authoritative data are combined to create a damage assessment and road damage map.

Although in this paper specific crowdsourced data (Civil Air Patrol photos) and volunteered data (YouTube videos) are utilized, this methodology can be extended to other sources with data pre-processing being source dependent. In this work, the classification of remote sensing data was accomplished by crowdsourcing, while the volume of volunteered data allowed for manual inspection and interpretation.

3.2 Non-authoritative damage assessment

We integrate non-authoritative data by interpolating to create a damage assessment surface. The geostatistical technique of kriging creates an interpolated surface from the spatial arrangement and variance of the nearby measured values [20]. Kriging creates a variogram to estimate spatial autocorrelation between observed values \( Z(x_i) \) at points \( x_1, \ldots, x_n \). The variogram determines a weight \( w_i \) at each point \( x_i \), and the value at a new position \( x_0 \) is interpolated as

\[
\hat{Z}(x_0) = \sum_{i=1}^{n} w_i Z(x_i).
\]

3.3 Integration with authoritative data

After a damage assessment surface is created from non-authoritative data, it is integrated with available authoritative information. For this research, authoritative data in form of a storm surge map created by FEMA MOTF is utilized as a comparison of flood extent as well as to illustrate how non-authoritative data can provide a range of damage estimations enhancing traditional storm surge products.

3.4 Generation of road damage map

Road damage is determined using the damage assessment surface created from the fusion of non-authoritative sources. Utilizing ArcGIS 10 software, a road network is layered over the damage assessment surface. Roads are classified based on the underlying damage assessment surface.

4. RESULTS

4.1 Damage assessment and authoritative data

Civil Air Patrol damage assessments for the area from 33N to 26N latitude and 90W to 84W longitude were downloaded directly from MapMill. The photographs were collected by the Civil Air Patrol between October 31-November 11, 2013 (within days of Hurricane Sandy impacting the New York City area). The photos were aggregated into a 500m grid structure. The value for each grid point is a function of the number of images present in each grid and their average crowdsourced damage assessment. As a result, each grid has a value from 1 to 10, with 1 representing no damage and 10 severe damage/flooding.

The videos were provided with geolocated information, and were visually assessed by the author. The small number of videos (n=15) did not require any crowdsourcing or automated assessment. Furthermore, it is shown in [17] that even a small number of properly located VGI data can help improve flood assessment. Each video point was assigned a value of 10 (severe damage/flooding).

The Civil Air Patrol and YouTube data were fused together using a kriging interpolation as described in section 3.2, resulting in a damage assessment surface generated solely from non-authoritative data. Kriging was selected as the interpolation method because it allows for spatial correlation between values (ie. locations/severity of flooding) to be considered and is often used with Earth science data [15, 14]. Ordinary kriging generated a strong interpolation model. Cross-validation statistics yielded a standardized mean prediction error of 0.0008 and a standardized root-mean-squared prediction error of 0.9967. Figure 2c illustrates the damage assessment within the boundaries of the FEMA surge extent. A histogram (Figure 4) shows the ranges in these damage assessment values. The peak in medium/severe damage values (7-8) illustrates how non-authoritative data can provide damage information not conveyed in the FEMA map.
providing flood information not conveyed in the Civil Air Patrol photos. As illustrated in (Figure 2b), the locations of the videos (green triangles) did not coincide with locations of photos rated as medium/severe damage (larger orange circles, values 7-10). Reasons for this disparity may include flooding captured on video had receded before the Civil Air Patrol flights or were captured at night, or flooding may have occurred in areas which were not in a flight path or were unable to be seen from aerial platforms (i.e. flooding in tunnels, under overpasses). By using multiple data sources, flood or damage details not captured by one source can be provided by another.

Figure 5: Example of YouTube video documenting flooding.

Figure 6: Designated areas ranging from medium to severely damaged (medium=7,8 severe=9,10) based on non-authoritative data.

A comparison of flood surface area between the two maps was also conducted. The storm surge area on the FEMA map is approximately 121 km$^2$. Using the higher rated areas of damage (regions with values from 7-10) from the non-authoritative assessment yielded an approximate surface area of flooding and damages of 157 km$^2$ (Figure 6). Using only the areas classified as medium-severely damaged, the surface area generated from non-authoritative sources is within 23% of FEMA’s surge extent for New York City.

Overall, there is a very good agreement between the flood extent from FEMA and the assessment generated with the proposed methodology. Sources of error in non-authoritative data, such as incorrect information (false positive/negative) or improper geolocation needed to be considered. Incorrect information can be mitigated by including visually verified photos/videos and the application of multiple sources. Crowdsourcing, in particular, can increase accuracy and enhance information reliability compared to single source observations [7]. Geolocation errors can be reduced with automation.

Sparse data or data skewed in favor of densely populated or landmark areas makes the use of non-authoritative data sources especially challenging. Increasing data volume and integrating authoritative data into the methodology can yield increased confidence and include underrepresented areas. Table 1 compares and summarizes some features of each type of data. Although non-authoritative data can provide timely, local information, they are often viewed with uncertainty. Conversely, the verification and authentication of authoritative data can be slower to ascertain and collect but yield trusted results.

<table>
<thead>
<tr>
<th>Benefits</th>
<th>Non-authoritative Data</th>
<th>Authoritative Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>volume</td>
<td>real-time</td>
<td>verified</td>
</tr>
<tr>
<td>citizens as sensors</td>
<td>authenticated</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Comparison between non-authoritative and authoritative data.

4.2 Road damage map

In Figure 2c, the damage assessment is limited to the FEMA generated surge extent for the sake of comparison. The fusion of the non-authoritative data predicted flooding and damages outside the FEMA surge boundary, so the full damage assessment was utilized for the road classification. A road network from the TIGER/line® shapefile was layered over the damage assessment surface. Road damages were then classified based on the underlying damage assessment (Figure 2d).

By using the damage assessment surface along with a high resolution road network layer, roads which may have severe damage can be identified at the street level. This allows authorities to prioritize site inspections, task additional aerial data collection, or identify routes which may compromised.

5. Conclusions

The application and integration of non-authoritative data offers opportunities to augment traditional data and methods for flood extent mapping and damage assessment. Although questions of reliability and validity are of concern when utilizing non-authoritative data, especially during natural disasters, these data can be employed along with traditional authoritative data and methods to enhance our knowledge of ground conditions. Although not considered ground truth, the fusion of multiple non-authoritative data sources helps fill in gaps in the spatial and temporal coverage of an event. In addition, the ability to identify potential
areas of road damage or inaccessibility from flooding can optimize response initiatives by identifying areas of severe damage.

6. ACKNOWLEDGMENTS

Work performed under this project has been partially supported by US DOT’s Research and Innovative Technology Administration (RITA) award # RITARS-12-H-GMU (GMU #202717). DISCLAIMER: The views, opinions, findings and conclusions reflected in this presentation are the responsibility of the authors only and do not represent the official policy or position of the USDOT/RITA, or any State or other entity.

7. REFERENCES


