THE PICTURE PILE TOOL FOR RAPID IMAGE ASSESSMENT: A DEMONSTRATION USING HURRICANE MATTHEW

O. Danylo [1], I. Moorthy [1], T. Sturn [1], L. See [1, *], J.-C. Laso Bayas [1], D. Domian [1], D. Fraisl [1], C. Giovando [2], B. Girardot [2], R. Kapur [3], P.-P. Matthieu [4], S. Fritz [1]

[1] International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria - (danylo, moorthy, sturn, see, lasobaya, domian, fraisl, fritz)@iiasa.ac.at
[2] Humanitarian OpenStreetMap Team (HOT), Washington, D.C., USA – (cristiano.giovando, blake.girardot)@hotosm.org
[3] Imperative Space, UK - ravi.kapur@imperativespace.com
[4] European Space Agency, Frascati, Italy - pierre.philippe.mathieu@esa.int

Commission IV, WG IV/4

KEY WORDS: disaster assessment, image interpretation, rapid classification, building damage, crowdsourcing, mobile apps

ABSTRACT:

In 2016, Hurricane Matthew devastated many parts of the Caribbean, in particular the country of Haiti. More than 500 people died and the damage was estimated at 1.9 billion USD. At the time, the Humanitarian OpenStreetMap Team (HOT) activated their network of volunteers to create base maps of areas affected by the hurricane, in particular coastal communities in the path of the storm. To help improve HOT's information workflow for disaster response, one strand of the Crowd4Sat project, which was funded by the European Space Agency, focussed on examining where the Picture Pile Tool, an application for rapid image interpretation and classification, could potentially contribute. Satellite images obtained from the time that Hurricane Matthew occurred were used to simulate a situation post-event, where the aim was to demonstrate how Picture Pile could be used to create a map of building damage. The aim of this paper is to present the Picture Pile tool and show the results from this simulation, which produced a crowdsourced map of damaged buildings for a selected area of Haiti in 1 week (but with increased confidence in the results over a 3 week period). A quality assessment of the results showed that the volunteers agreed with experts and the majority of individual classifications around 92% of the time, indicating that the crowd performed well in this task. The next stage will involve optimizing the workflow for the use of Picture Pile in future natural disaster situations.

1. INTRODUCTION

Volunteered Geographic Information (VGI) is a rapidly growing field of research since it first appeared as a concept in the literature more than a decade ago (Goodchild 2007). One of the most successful examples of VGI is the OpenStreetMap (OSM) initiative, where large numbers of volunteers contribute to the development of an openly available map of the world (Jokar Arsanjani et al. 2015; Mooney and Minghini 2017). Although the original purpose of OSM was to provide freely available mapped features of the type commonly found in the databases of national mapping agencies (which generally charged money for the data), the value of OSM for disaster preparedness and response has now been clearly recognized (Soden and Palen 2014). After the Haiti earthquake in 2010, 600 OSM mappers built a base map of Haiti in just 3 weeks. After this event, the Humanitarian OpenStreetMap Team (HOT) was launched, turning what was an informal entity into what is now a successful non-profit organization. HOT undertook a series of missions to Haiti in the aftermath of the 2010 earthquake and become well established in the country after 1.5 years, disseminating the value of community-based mapping and community ownership of the data. They have also developed new software such as the HOT Tasking Manager to support mapping by multiple volunteers and the HOT Export Tool to facilitate extraction of portions of the OSM database.

At the same time, other innovative tools have emerged that are using new digital technologies in the context of disaster response, e.g. the use of social media (Vivacqua and Borges 2012; Bruns and Liang 2012), as well as mobile-based applications to help filter out and improve the information coming in during an event. For example, MicroMappers (https://micromappers.wordpress.com/) have developed a set of rapid tagging applications for classification of Twitter data (Text Clicker), to ensure that only the most relevant tweets reach first responders in an emergency, and for identification of damage from Unmanned Aerial Vehicles (UAVs) (Aerial Clicker) and geotagged photographs (Image Clicker). MapSwipe (https://mapswipe.org/) is another mobile application for rapid identification of features from very high resolution satellite imagery. Developed by Médecins Sans Frontières (MSF) and part of the Missing Maps initiative, volunteers look for evidence of where people are living so that during major disease outbreaks, MSF staff can be mobilized quickly to the areas where vaccines are needed. Picture Pile is another tool that has emerged for rapid image classification and assessment. It differs from the other applications in that it can provide pairs of images for users to examine. In the context of disaster response, it can be used to classify images before and after an event, facilitating the identification of damage.

The aim of this paper is present results from a first test that used Picture Pile for rapid damage assessment of buildings by

---

* Corresponding author

Hurricane Matthew. Hurricane Matthew was a category-5 hurricane that hit in early October 2016 and affected parts of Canada, the USA and several islands in the Caribbean. However, the most significant effects were felt in Haiti, where there were more than 500 fatalities and 1.9 billion USD of damage (Wikipedia 2018). Hence this event was chosen as a demonstration case because widespread damage would be visible from very high resolution satellite imagery.

After a brief description of the Picture Pile tool and some history of its application, the campaign for assessing building damage from Hurricane Matthew is outlined. This is followed by the results from the campaign, which includes the map of building damage produced by the volunteers. An initial quality assessment is then provided, including an analysis of the performance of the volunteers against experts and the results when considering the majority. Finally we consider the speed at which the task was completed, which could provide some indication of how quickly such a map could be created during a real event.

2. PICTURE PILE FOR RAPID IMAGE ASSESSMENT

2.1 The Picture Pile tool

The Picture Pile application is designed as a generic and flexible tool for ingesting imagery for rapid assessment and classification. The images can be very high resolution satellite images, orthophotos, images from UAVs or geotagged photographs. This tool represents a generalization of the Cropland Capture tool (Sturn et al. 2015) in which volunteers were shown boxes of 1 km^2 (at the equator) and were asked whether they could see evidence of cropland in the satellite image or geo-tagged photograph. A simple game mechanic was employed in which volunteers would swipe the image to the right if evidence of cropland was visible, to the left if no evidence was present and downwards to denote 'maybe', e.g. when images were cloud covered or features were difficult to distinguish.

Cropland Capture was used in a campaign that ran for 25 weeks and resulted in the collection of more than 5 million image classifications. A number of lessons were learned during this campaign including the need for control or expert data to assess quality rather than relying on only majority voting for determining accuracy (Salk et al. 2017). However, in general, the quality of the classifications by the crowd was high (Salk et al. 2015).

2.2 Campaign to map deforestation

The first use of Picture Pile in a campaign was one focussed on the identification of deforestation (see Figure 1). The campaign ran for several months during 2016/17, where the volunteers were asked to examine pairs of very high resolution satellite images from different years (purchased from DigitalGlobe) for evidence of forest loss (i.e. images from before and after any deforestation has occurred). The areas shown to the volunteers were pixels of 250 m^2 (at the equator) and volunteers were asked to determine if forest loss of more than 5% of the image was visible. Pairs of images were sampled from Tanzania and Indonesia using the forest loss gain product of Hansen et al. (2013) as a stratification layer.

In contrast to Cropland Capture, control data were collected by experts and used in the scoring of the volunteers since it was

shown that majority agreement is not always the most reliable indicator of quality (Salk et al. 2017). More than 5 million classifications were obtained from 1339 volunteers during the course of the campaign. The data are currently being analysed in various ways, including an assessment of the data quality and an evaluation of the forest loss gain product of Hansen et al. (2013) in Tanzania and Indonesia.

Figure 1. A screenshot from the Picture Pile tool used for identification of forest loss

3. METHODLOGY

For this study, Picture Pile was applied to mapping building damage from Hurricane Matthew. The methodology, from ingestion of initial images to running the campaign, is shown in Figure 2. Similar to the deforestation campaign, pairs of images were provided to the user, i.e. before the disaster and post-disaster. For the purpose of this test, the input data (or images) were obtained from previously released, easy to access, open data from Digital Globe[1] as well as Microsoft Bing. The second stage in the methodology was the filtering of images since the number to be processed was large. Images were first filtered to remove water, i.e. near coastal areas. The Global Urban Footprint product[2] and a road network from OSM were then used to select areas with a higher probability of having buildings. In total there were 37458 images used in the campaign for rapid classification by the volunteers, including the control or expert data set as described below. This was followed by a pre-processing stage in which images were converted from the Tile Map Service to a format needed for ingestion into Picture Pile.

---

[1] https://www.digitalglobe.com/opendata/hurricane-matthew
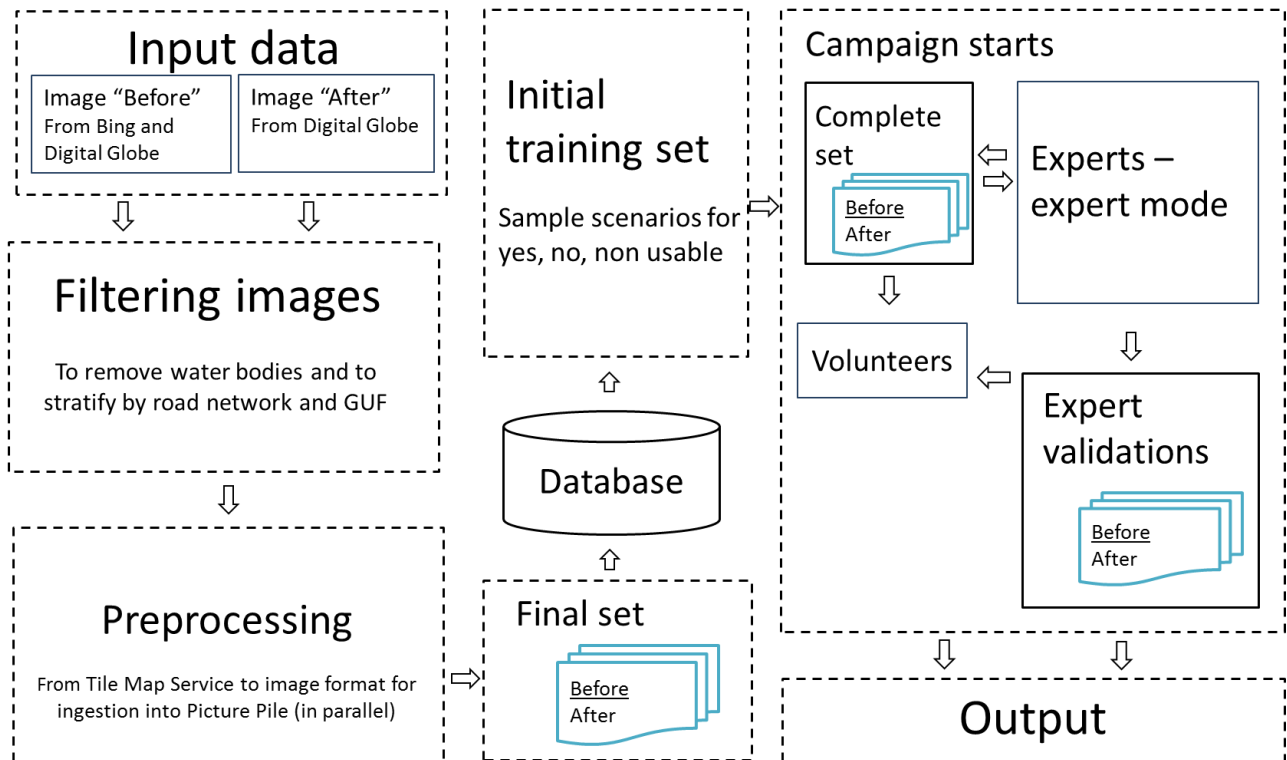[2] http://www.dlr.de/guf

Figure 2. An overview of the methodology used to assess building damage with Picture Pile

Prior to campaign launch, an initial training data set was created. Experts from HOT and IIASA classified 3000 images for volunteer training purposes, indicating examples with damaged and non-damaged buildings. From these classifications, the experts agreed completely on 1743 images as follows:

- 294 images had visible damage;
- 1143 images had no signs of damage;
- 306 images were deemed not usable, e.g. due to cloud cover or poor quality of the image.

A subset of these images was then annotated by the experts to create an initial training set, which was used to explain the task to the volunteers and train them in visual interpretation. For the purpose of this exercise, we focussed only on visible evidence of damaged buildings, not other indicators of damage such as damaged vegetation and flooded areas.

The images were reviewed and annotated to cover the following scenarios for training:
- Volunteers should answer "Yes" if:
  - Buildings are present showing visible damage;
  - Clouds cover part of the image but in the non-cloudy part, damaged buildings are visible.
- Volunteers should answer "No" if:
  - Buildings are present but no damage is apparent;
  - No buildings are visible anywhere in the image;
  - Damage to vegetation is visible, buildings are present but they do not appear to be damaged;
  - Debris is visible but it is not clear from the "Before" image whether or not a building was present.
- Volunteers should answer "Non usable" if:
  - Clouds cover the entire image;

  - Clouds cover part of the image and no buildings are visible in the non-cloudy areas;
  - Clouds cover part of the image and buildings with no damage are visible in the non-cloudy areas.

From these expert classifications, a subset (or initial training 'pile') was created for use in the actual campaign where the interface is shown in Figure 3.
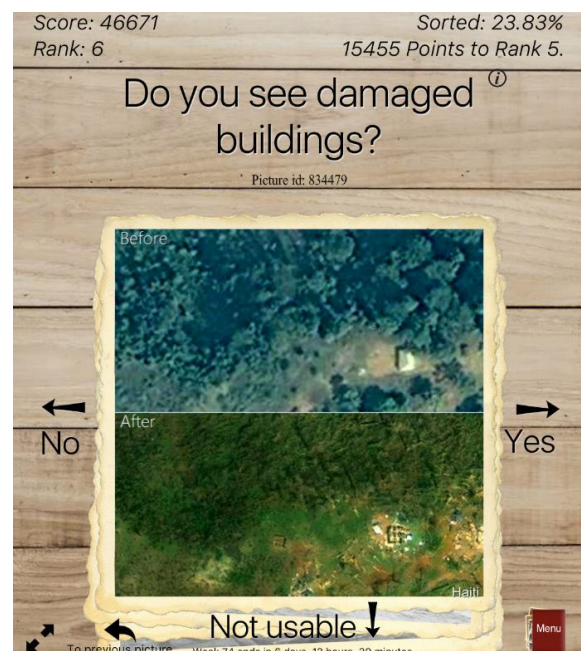

Figure 3. A screenshot of the Picture Pile tool used for assessment of building damage where a pre- and post-disaster image is presented to the volunteer

The training data set consisted of 263 expert images in which damage was visible, 258 images in which no damage was visible, and 41 unusable images for a total of 562 control images. Once this set was compiled, the campaign began (Figure 2), where the initial data were shown to the volunteers for training and quality assurance purposes. A simple scoring mechanism was also used. After a training period, volunteers would lose points if they classified a control point incorrectly. They would also receive feedback, i.e. an annotated image, so that training would continue throughout the campaign. Experts continued to provide more training data during actual campaign until all the images were classified. The outputs (Figure 2) are described in the next section.

4. RESULTS

4.1 Building damage and campaign statistics

The Hurricane Matthew Picture Pile campaign was publicly launched on April 28, 2017, at the European Geosciences Union (EGU) General Assembly in Vienna, Austria. Complementing this launch was a press release, additional promotion via social media channels including Twitter and Facebook and dissemination to other networks including a newsletter. The majority of volunteers joined the campaign on the launch day. Additional campaign promotion was carried out on May 4, 2017 by HOT on Twitter, which helped to recruit new volunteers to the campaign.

All of the image pairs were classified within the first week of the campaign. However, the volunteers were encouraged to continue the classification to help further improve the quality of the results, since the same image was given to more than one volunteer. Overall a total of 248,997 classifications were collected from 179 volunteers in less than three weeks, and half of the classifications were collected during the first five days of the campaign.
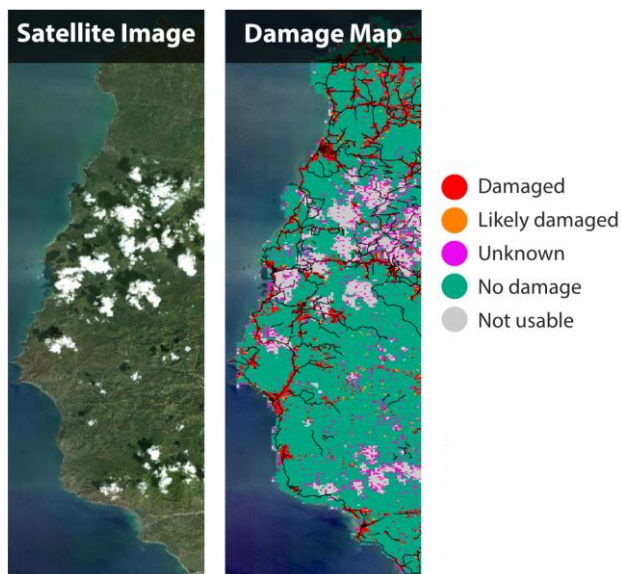

Figure 4. Results of the post-disaster building damage assessment. Left: Digital Globe (WorldView 3) satellite image of the area affected for 10 October 2016. Right: Result of post-disaster damage assessment using Picture Pile

Based on the crowdsourced classifications, an initial damage assessment map (Figure 4) was created based on the following criteria:

- **Damaged areas**: 4 or more volunteers agreed on visible damage to the buildings;
- **Likely damaged**: 3 volunteers agreed on visible damage to the buildings;
- **Unknown:** no majority agreement between volunteers on presence/absence of damage;
- **No damage**: 4 or more volunteers agreed on no damage to the buildings;
- **Not usable:** 4 or more volunteers agreed that the image was not usable due to cloud cover.

From Figure 4, one can see that most of the damaged areas detected by the volunteers correspond to the spatial distribution of the road networks (i.e. where settlements and buildings are located). The majority of the 'unknown' areas were due to volunteer uncertainty as a result of cloud cover. While in some cases the damage was visible on part of the image, volunteers sometimes reported such images as not usable, being conservative in their estimation.

4.2 Comparison of the crowd classifications with experts

As mentioned above, we used 562 locations as expert controls to ensure the quality of the crowdsourced data during the campaign (for training and as part of the scoring mechanic of Picture Pile) but also for post-campaign assessment. Table 1 provides a comparison of the expert control data with the individual classifications in the form of an error matrix. The total number of images classified in Table 1 is greater than 562 since the controls were seen by many individuals.

Individual	Experts			
	Not usable	Damage	No damage	%
Not usable	926	184	108	76.0
Damage	79	8602	280	96.0
No damage	163	466	6243	90.9
%	79.3	93.0	94.2	92.5

Table 1. Error matrix for agreement of individual volunteer ratings with the expert classification for images in the Hurricane Matthew Picture Pile campaign

The overall accuracy is 92.5%, indicating that the crowd performed well in relation to the experts. There was some minor confusion in which the crowd labelled unusable images as having no damage compared to the experts and vice versa but these types of errors would have little impact on the overall map of damage. However, 7% of the time, the crowd missed damaged buildings compared to the experts while 4% of the time, the crowd saw damage when none was found by the experts.

4.3 Comparison of the individual classifications with the majority rating

A second quality evaluation was undertaken, comparing the individual classifications with the majority rating (Table 2). The results were similar to the comparison with the experts, i.e. overall agreement was 92.2%. In contrast to the expert comparison, this time there was less confusion where an individual labelled unusable images as having no damage

compared to the majority and vice versa. However, there was higher confusion between images that contained damaged buildings when there were none (when taking the majority into account) and vice versa, of about 13%.

Individual	Majority			
	Not usable	Damage	No damage	%
Not usable	29836	529	6866	80.1
Damage	207	30694	4343	87.1
No damage	3425	3994	169103	95.8
%	89.2	87.2	93.8	92.2

Table 2. Error matrix for agreement of individual volunteer ratings with the majority classification of all volunteers for images in the Hurricane Matthew Picture Pile campaign

4.4 Time taken to complete the task

The user interface of Picture Pile has been designed and built for rapid classification of imagery. Figure 5 shows the distribution of time spent by the volunteers to complete the classification task. We excluded any classification that took more than 60 seconds as this may indicate that the volunteer took a break during the validation session; there were 657 such cases (or 0.26% of all classifications).
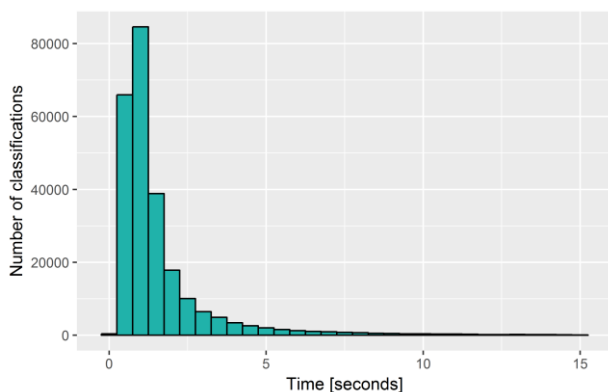

Figure 5. Frequency distribution of the number of classifications by the average time that volunteers spent on the classification task

On average each user spent 1.76 seconds per classification while 99.7% of the classifications were completed by 179 volunteers within 122 hours. Given that this was a test and that most volunteers were from the Geo-Wiki network of volunteers and not HOT's much, more extensive network, it would be possible to complete such a task with thousands of volunteers in the matter of 1 or 2 days. More precise calculations of the amount of human resources needed could be undertaken so as to provide HOT with a recruitment strategy for this type of task using Picture Pile during an actual event.

5. DISCUSSION

The results showed that volunteers were able to classify images rapidly with a high accuracy, which resulted in a usable map of building damage that could be used as one piece of information guiding emergency and disaster response. The workflow established during this project represents an implementable protocol for future post-disaster damage assessments with

Picture Pile but the challenge is to fully operationalize the approach. Much of the image preparation and expert data collection were laborious and will need to be automated in future applications.

The method also relies on experts to provide training data and for quality assurance. A group of experts would need to be mobilized quickly during real events to create this data set. However, it might be possible to reuse training data once collected over several campaigns. Similarly, outputs would need to be produced quickly in an automated manner.

Despite the need to improve certain aspects of the workflow, this method has a number of advantages over other methods. For example, there is no method that currently utilizes pairs of images to aid in building damage recognition. The interface is also easy to use and promotes rapid assessment through the mechanics of the interface. The scoring mechanism and initial training data are also designed to improve the overall quality. Similarly, gathering data from multiple volunteers at the same place provides information about the uncertainty of building damage at different locations. Finally, the scoring element provides some gamification to the application, which is not normally seen in humanitarian applications but nevertheless provides additional incentives to the volunteers.

In this example, Picture Pile was used for a hurricane event but it would also be possible to use this same approach for damage assessment from other events such as flooding or landslides. This would simply involve modifying the methodology to filter the imagery to focus on flood prone areas or those identified as having higher landslide risk. In fact, we have had previous discussions regarding how HOT could use Landsat imagery to identify potential landslides and Picture Pile could then be used to confirm the locations using VHR imagery as part of a larger HOT workflow on identifying areas that may need assistance.

6. CONCLUSIONS

This paper demonstrates how satellite imagery linked with a rapid image classification application such as Picture Pile could potentially support humanitarian efforts. Through a simple classification mechanic involving a yes/no/maybe question, volunteers helped to detect damaged buildings over a large region affected by Hurricane Matthew in a short period of time. Based on the results of this crowdsourcing campaign, overall agreement between both the volunteers and the experts and the volunteers and the majority was high, supporting the validity of such a crowdsourced approach for rapid post-disaster damage assessments. Work is ongoing to determine how Picture Pile can be used more operationally in future events.

ACKNOWLEDGEMENTS

REFERENCES

Bruns, A., and Liang, Y.E., 2012. Tools and methods for capturing Twitter data during natural disasters. *First Monday*, 17(2), April 2012.

Goodchild, M.F., 2007. Citizens as sensors: The world of volunteered geography. *GeoJournal*, 69, pp.211–221.

Hansen, M.C., Potapov, P.V., Moore, R., et al., 2013. High-resolution global maps of 21st-century forest cover change. *Science*, 342, pp.850–853, doi: 10.1126/science.1244693.

Jokar Arsanjani, J., Zipf, A., Mooney, P., and Helbich, M. (eds) 2015. *OpenStreetMap in GIScience*. Springer International Publishing, Cham, Switzerland.

Mooney, P., and Minghini, M., 2017. A review of OpenStreetMap data. In: Foody, G.M., See, L., Fritz, S., et al. (eds) *Mapping and the Citizen Sensor*. Ubiquity Press, London, UK, pp. 37–59.

Salk, C.F., Sturn, T., See, L., et al., 2015. Assessing quality of volunteer crowdsourcing contributions: lessons from the Cropland Capture game. *International Journal of Digital Earth*, pp.1–17, doi: 10.1080/17538947.2015.1039609.

Salk, C.F., Sturn, T., See, L., and Fritz, S., 2017. Limitations of majority agreement in crowdsourced image interpretation. *Transactions in GIS*, 21, pp.207–223, doi: 10.1111/tgis.12194.

Soden, R., and Palen, L., 2014. From crowdsourced mapping to community mapping: The post-earthquake work of OpenStreetMap Haiti. In: Rossitto, C., Ciolfi, L., Martin, D., and Conein, B. (eds) *COOP 2014 - Proceedings of the 11th International Conference on the Design of Cooperative Systems*, 27-30 May 2014, Nice, France. Springer International Publishing, Cham, Switzerland, pp. 311–326.

Sturn, T., Wimmer, M., Salk, C., et al., 2015. Cropland Capture – A game for improving global cropland maps. Foundation of Digital Games Conference (FDG 2015), 22-25 Jun 2015. Pacific Grove, CA.

Vivacqua, A.S., and Borges, M.R.S., 2012. Taking advantage of collective knowledge in emergency response systems. *Journal of Network and Computer Applications*, 35, pp.189–198, doi: 10.1016/j.jnca.2011.03.002.

Wikipedia, 2018. Hurricane Matthew. https://en.wikipedia.org/wiki/Hurricane_Matthew (15 May 2018).