FAST INSIGHT ABOUT THE SEVERITY OF HURRICANE IMPACT WITH SPATIAL ANALYSIS OF TWITTER POSTS

K. Spasenovic[1*], D. Carrion[1], F. Migliaccio[1], B. Pernici[2]

[1]DICA, Geodesy and Geomatics, Politecnico diMilano, Milan, Italy (katarina.spasenovic, daniela.carrion, federica.migliaccio)@polimi.it
[2]DEIB, Politecnico di Milano, Milan, Italy, (barbara.pernici)@polimi.it

Commission IV, WG IV/4

KEY WORDS: Spatial analysis, social media, Twitter, crisis, disaster, emergency management, hurricane

ABSTRACT:

Social media could be very useful source of data for a people interested in disasters, since it can provide them with on-site information. Posted georeferenced messages and images can help to understand the situation of the area affected by the event. Considering this type of resource as a real-time crowdsource of crisis information, the spatial distribution of geolocated posts related to an event can represent an early indicator of the severity of impact. The aim of this paper is to explore the spatial distribution of Twitter posts related to hurricane Michael, occurred in the USA in 2018 and to analyse their potential in providing a fast insight about the event impact. Kernel density estimation has been applied to explore the spatial distribution of Twitter posts, after which Hot Spot analysis has been performed in order to analyse the spatiotemporal distribution of the data. Hot Spot analysis has shown to be the most comprehensive analysis, detecting the area of high impact. The Kernel density map has shown to be useful as well.

1. INTRODUCTION

Social media have shown a significant contribution in disaster relief. They could be very valuable sources of the on-site disaster-related information shared by the affected people. Since disasters are generally characterized by high need and low availability of information (Shklovski et al. 2010, Albuquerque et al. 2014), it seems obvious to consider social media as an additional data source for crisis management and rapid mapping (Spasenovic et al. 2019). The work presented by Spasenovic (2018) examined three social media platforms and showed that Twitter is currently one of the most popular social media used for the exchange of information connected to disasters. Focusing on the tweets that carry a geographic reference, a "picture" of what is happening in a specific place could be made mapping the posts over the area of interest. This data source can be considered as a real-time crowdsource of crisis information, the spatial distribution of geolocated tweets related to an event can represent an early indicator of the severity of impact. This raises a question: would it be possible to understand the affected zones and approximate level of impact by mapping and processing spatially the posts?

The use of social media is constantly growing. In the past few years, an increasing interest in analysing information produced by social media related to crises has been recorded in literature. Corbane et al. (2012) have explored the relationship between the spatial pattern of SMS messages and building damage in the Haiti disaster of 2010. Albuquerque et al. (2014) have explored the spatial distribution of social media with respect to authoritative data in order to identify useful information for disaster management. Sakaki et al. (2010) and Crooks et al. (2013) investigated the use of Twitter for detecting and estimating the trajectory of earthquakes in real time. Some of those studies are focused on spatial distribution of social media data in order to understand the event. Others analyse the spatial relationships between Volunteered Geographic Information (VGI) and other spatial information related to the event.

The aim of this paper is to explore the spatial distribution of Twitter posts related to a disaster and to analyse their potential in providing a fast insight regarding its impact. The focus of the analysis was on the tweets related to hurricane Michael that hit Florida (USA), in October 2018. Reference information related to the hurricane was obtained from National Oceanic and Atmospheric Administration (NOAA) report (Beven et al. 2019). The focus of this study is to explore the potential of Twitter crisis posts in showing affected areas. The available message data set consisted of 8169 tweets posted from 10th until 15th October. The tweets have been filtered based on their content and publication time. A set of spatial statistics have been computed with the open-source software QGIS, identifying where statistically significant spatiotemporal clusters were present. More precisely, the zones with significant concentration of relevant posts were corresponding to areas with high impact of the hurricane. Comparisons of the results with the reference data have been performed and discussed, showing the potential value of these data for crisis mapping.

2. CASE STUDY AND DATA SET: HURRICANE MICHAEL TWITTER POSTS

2.1. The Hurricane Michael

In the period from 7th October till 11th October of 2018 the strong hurricane called Michael hit Florida Coast of the United States. According to the Saffir-Simpson hurricane wind scale (Saffir 1973, Simpson 1974), which is based on sustained (i.e. relatively long-lasting) winds, there are five possible categories (see Table 1).

---

* Corresponding author

Category	m/s	knots	mph	km/h
Five	≥70	≥137	≥157	≥252
Four	58 - 70	113 - 136	130 - 156	209 - 251
Three	50 - 58	96 - 112	111 - 129	178 - 208
Two	43 - 49	83 - 95	96 - 110	154 - 177
One	33 - 42	64 - 82	74 - 95	119 - 153

Table 1. Saffir-Simpson hurricane wind scale

As stated in the NOAA report done by Beven et al. (2019), hurricane Michael was a category 4 hurricane that made a catastrophic landfall near Mexico Beach and Panama City, Florida. It caused 16 deaths and about $25 billion damage in the United States (Beven et al., 2019). Before arriving to the United States, the cyclone hit the western part of Cuba as a category 2 hurricane, then it continued its path increasing its strength as shown in Figure 1. It started and ended as category zero, which is classified as tropical storm (and not hurricane), as also presented in Figure 1. The wind speed that characterizes a tropical storm is shown in Table 2, and it is below the limits of level 1 hurricane wind scale.
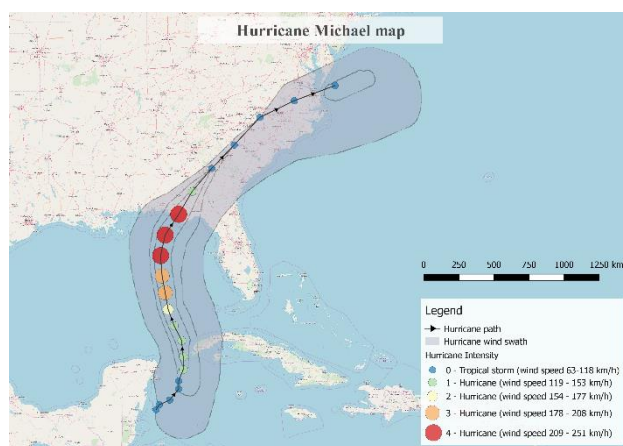

Figure 1. An overview map of the path of Hurricane Michael

Category	m/s	knots	mph	km/h
Tropical storm	18 - 32	34 - 63	39 - 73	63 - 118

Table 2. Tropical storm wind scale
(Saffir 1973, Simpson 1984)

2.2. The Twitter data set

The available Twitter data set contains 8 169 georeferenced short texts with image, tweets, posted within the territory of the United States. Beside the image and text message, every tweet contains several metadata fields, such as the time stamp (creation time: day, month, year and hour), hashtags (i.e. keywords preceded by #), URLs and integer unique ID. The time stamp of the data set is from 10th until 15th October 2018. The data set was provided by E2mC[1] . The extraction of the posts from Twitter was based on a set of hashtags relevant to this type of events. The location of tweets could be obtained explicitly or implicitly. Explicit method, known as geotagging, allows users to associate a location to the post, usually provided by the GPS of the device used for posting. If the location is not provided by the geotag it can be found implicitly by the CIME algorithm (Pernici et al.2018). The algorithm extracts from the post and its context

---

1 Data collected within the H2020 Evolution of Emergency Copernicus service (E2mC) project, www.e2mc-project.eu

(comments, shares, likes) the most accurate location that can be associated to it. The CIME algorithm was used to assign the location to 8 120 tweets, corresponding to 99.4% of the data set. Relevance filtering was performed on the data set and the output of the filtering was stored in the metadata as well. The Framework of automatic filtering applied on the data set has been presented in the paper of Barozzi et al. (2019), it includes:

- Redundant information reduction
  - Duplicate reduction
  - Retweets elimination
- Image filtering
  - Colour-based filtering
  - Elimination of pre-event images
  - Face detection
  - Text detection

For the purpose of this research, only posts classified as relevant were selected from the provided data set. Relevant posts that were providing information about the hurricane impact were 254, which is 3% of the initial data set. An overview of the full dataset (relevant tweets – coloured dots and irrelevant tweets – black dots) is presented in Figure 2.
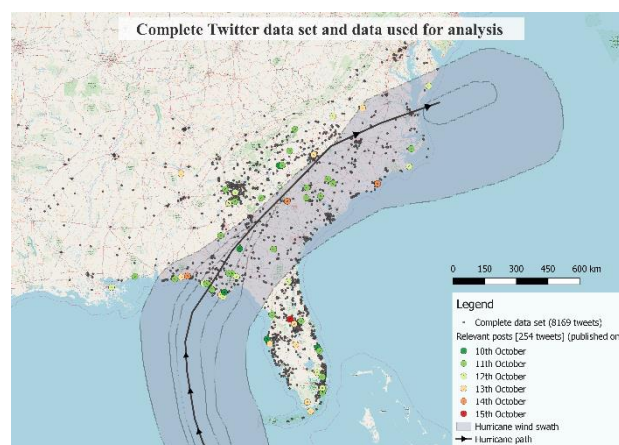

Figure 2. An overview of the complete tweet data set, overlaid to the hurricane path

3. METHODOLOGY AND RESULTS

3.1. Spatiotemporal analysis of tweets

This article explores possibilities for improving situation awareness during the emergency, analysing the spatial pattern of georeferenced social media data. According to Albuquerque et al. (2014), spatiotemporal characteristics of the event affect the spatiotemporal characteristics of social media messages. Studies that explored communication during the crisis have shown that social media messages coming from people close to the events may contain useful information (Acar et al. 2011, Bruns et al. 2012). The reasoning behind this result could be that people in affected areas feel the need to share information about the event situation that they are facing with. Based on this, the hypothesis is that high concentration of tweets located near the event could indicate the affected areas. Our approach thus explores spatiotemporal patterns, taking into consideration day and time of posts publication. Several spatial and spatiotemporal analyses have been done. Kernel density estimation has been applied to explore the spatial distribution of Twitter posts, and after that Hot

Spot analysis has been performed in order to analyse spatiotemporal distribution. The goal is to test distribution and pattern of useful tweets as a significant resource for the identification of the severity of the event.

3.2. Kernel Density computation

The density of Twitter posts has been analysed applying the Heatmap plugin in QGIS using Kernel Density estimation. This method calculates a magnitude-per-unit area starting from a point feature. The output heatmap visually highlights point pattern, in particular, social media posts distribution. Since the tweets used in the case study are spread over large area (see Figure 2), this tool could be very useful for finding and emphasizing the clusters. As an input parameter, the tool requires the size of the search radius to compute the cluster's density. For the purpose of this research, a bandwidth of 100 km has been defined, considering it as an approximation of the average town radius in the observed region (source: Florida League of Cities https://www.floridaleagueofcities.com/research/data-statistics/cmri-reports). We expect to find clusters in populated areas along the hurricane path (Figure 3), since Twitter posts are strictly dependent on population presence.
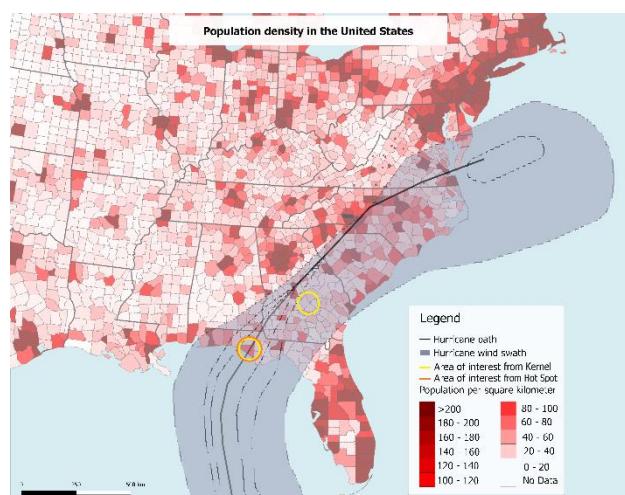

Figure 3. Population density in the United States in 2010 (yellow circles and the orange circle show areas with high impact detected respectively by Kernel Density and by Hot spot analysis)

The analysis has been performed on the 254 relevant tweets. Posts that are falling outside the wind swath have been considered as well, with the purpose of looking for a correlation between the data set spatial distribution and the hurricane path. The idea is to get the information about the event impact based on reactions on social media.

Looking at Figure 4, it is possible to see that tweets are spread over a large area of the East Coast of the United States. Kernel analysis, based on spatial distribution of Twitter posts, has highlighted two areas with high concentration of messages. The first area with high density is in Panama City and the second area is close to Douglas. Looking at the posts distribution in the small scale map (Figure 4) it is very difficult to understand where the high concentration of tweets is: even though the posts do not share the same location, at a small scale it seems that they overlap. Taking a closer look to the distribution of posts in the higher density cluster detected by the Kernel analysis, corresponding to Panama City (Figure 4) it is possible to see that there is a significant number of tweets well spread in the city.
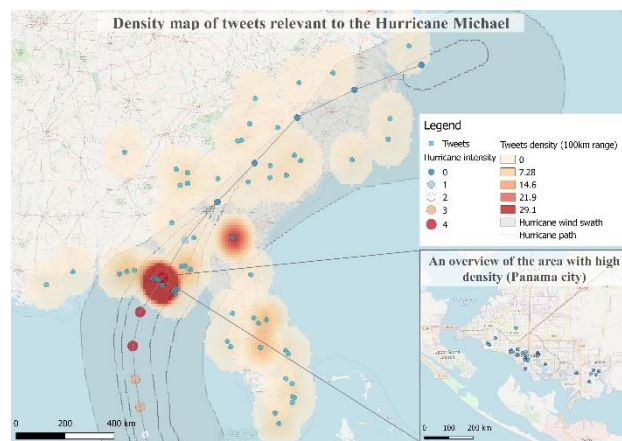

Figure 4. Kernel density map

The second area that presented a significant concentration of posts is positioned outside of the populated areas, as shown in Figure 3 with yellow circle. Analysing this area on a larger scale map (Figure 5), it has been found out that post locations are in the agricultural areas and they are distanced from populated areas. Checking the content of those tweets it has been discovered that they have been posted for the most part (21 out of 29) by telecommunication companies, so, it should be further investigated why the CIME algorithm assigned that location.
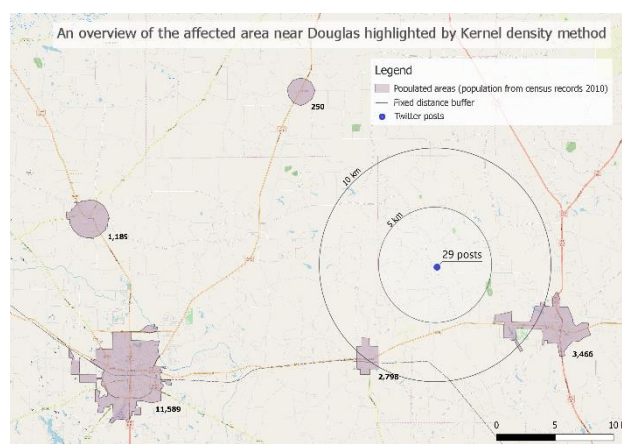

Figure 5. An overview of the populated area near Douglas and Twitter posts highlighted by the Kernel density method

Comparing the obtained results with the hurricane information (Figure 4) it is possible to see that the area with the highest density is close to the central path of hurricane Michael, when it corresponded to category 4 (wind of 209 km/h). Panama City was the first land area that was hit by the hurricane and according to the NOAA records in that area the hurricane intensity was the highest. On the other hand, the second area with a significant concentration is where the hurricane intensity was decreasing, and it was becoming a tropical storm.

3.3. Hot Spot analysis

According to Cliff & Ord (1973), analysis of spatially distributed variables can lead to unreliable statistical interpretation when proper corrections for spatial effects are not considered in the model. The Kernel density method presented above was focused on the spatial distribution only, indicating areas with high density of posts. Having in mind that the reaction of affected people (in this case tweets) is correlated to the time when an event occurs,

the time of posting should be considered in the model. Applying QGIS plugin for Hot Spot analysis developed by Oxoli et al. (2016), the temporal information of posting could be taken into account as well, inserting an additional element to interpret the output of the Kernel density method. The computations implemented into Hot Spot plugin are the Getis-Ord Gi*statistic (Getis and Ord 1992), the Local Moran's I statistic (Anselin 1995) and the Bivariate Local Moran statistic (Wartenberg 1985). The analysis was performed considering as correlation attribute the day of posting. The time of the day did not show to be a good reference with respect to this analysis, having an erratic behaviour. The Hot Spot tool requires distance interval (search radius) and distance step (increments in search radius) as input parameters: they have been defined based on the average size of towns. The resulting cold spots present high concentration of the tweets corresponding to the beginning of the event, while hot spots are tweets clusters corresponding to the end of the event. For this specific case, we have that cold spot in Figure 6 shows the presence of posts that have been published during first few days of the event, when the hurricane had its highest intensity and when it started hitting the US coast.
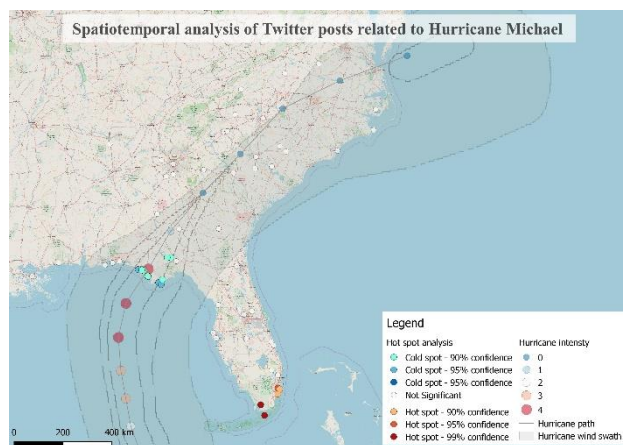

Figure 6. Hot spot analysis of Twitter posts (cold spots present concentration of posts published at the beginning of event; hot spot present concentration of posts published on the last days of the event)

The location of cold spot confirms the output of the Kernel analysis, which singled out Panama City as major place of impact. The combination of the two analyses shows that tweets close to each other temporally and spatially could be used as an indicator for the impact. The second area that had been spotted by the Kernel density method has not been detected by the Hot Spot analysis. This allowed to filter posts which were located to have close positions by CIME algorithm, but which did not correspond to a significant cluster of posts related to the reaction of people to the intensity of the hurricane.

4. DISCUSSION AND CONCLUSION

The work presented in this paper aims to explore the spatial distribution of Twitter posts related to a hurricane and to see if spatial and temporal analyses could be helpful for indicating the affected areas. The analyses have been performed on the 254 georeferenced tweets, out of 8 169 crawled tweets that were considered as relevant for the hurricane Michael event of 2018. For the purpose of these analyses, information of the geolocation and day of posting were used. Taking into account only the location of tweets, Kernel density analysis has been performed. As a result, two areas have been identified: the first area was in

Panama City and the second was near Douglas. This method has shown to be useful but not relevant with respect to the purpose of this work, because checking the results with referenced data it has been found out that the area near Douglas had not been really hit with high impact. Indeed, according to the meteorological records that area has been hit when the hurricane lowered its intensity to the level of a tropical storm. Adding temporal parameters and applying Hot Spot analysis has shown better selection of the clusters with respect to the aim of this work, i.e. selecting the most affected areas. This spatiotemporal analysis has highlighted only Panama City as the area of high impact. Having a cold spot in this area pointed out the location of the most significant impact was located but showed that in the first days of the event people activity on Twitter related to the event was high.

These first spatiotemporal analyses of social media data showed that they could provide fast information about the severity of the impact of a disaster, as other examples in literature showed (Corbane et al. 2012). The results will be tested on other test cases.

Social media platforms are very powerful source of information, and lately they are growing very fast. Few decades ago, it was barely impossible to imagine social media such as Twitter to be exploited for emergency management. Nowadays they are part of everybody's life and by selecting the right content, valuable information can be obtained. Future work should thus monitor the activity of users that provide information related to the crisis events, exploit and compare the presented methods on events that have a higher number of responses through social media and analyse the correlation between damage and posts.

REFERENCES

Acar, A., Muraki, Y., 2011. *Twitter for crisis communication: lessons learned from Japan's tsunami disaster.* International Journal of Web Based Communities, 7 (3), 392–402. doi.org/10.1504/IJWBC.2011.041206

Anselin, L., 1995. Local Indicators of Spatial Association – LISA. *Geographical Analysis*, 27 (2), 93-115. doi.org/10.1111/j.1538-4632.1995.tb00338.x

Albuquerque, P., Herfort, B., Brenning, A., Zipf, A., 2015. A geographic approach for combining social media and authoritative data towards identifying useful information for disaster management. *International Journal of Geographical Information Science*, 29 (4), 667-689, doi.org/10.1080/13658816.2014.996567

Barozzi, S., Fernandez Marquez, J.L., Shankar A.R., Pernici, B., 2019. Filtering images extracted from social media in the response phase of emergency events. *Social Media in Crises and Conflicts, 16th ISCRAM Conference*, Valelncia, Spain.

Beven, L. J. II, Berg, R., Hagen,, A., 2019. Tropical cyclone report hurricane Michael (AL142018), National Hurricane Center, United States.

Bruns A., Burgess, J., Crawford, K., Shaw, F., 2012. # qldfloods and@ QPSMedia: Crisis communication on Twitter in the 2011 south east Queensland floods. Brisbane: ARC Centre of Excellence for Creative Industries and Innovation.

Cliff, A., Ord, J. K. 1973. Spatial autocorrelation. Pion, London.

Crooks, A., Croitoru, A., Stefanidis, A., Radzikowski, J., 2012. *#Earthquake: Twitter as a distributed sensor system*. Transactions in GIS, 17 (1), 124–147. doi.org/10.1111/j.1467-9671.2012.013 59.x

Getis, A., Ord, K., 1992. The Analysis of Spatial Association by Use of Distance Statistic*s. Geographical Analysis* 24, 189-206.

Oxoli, D., Prestifilippo, G., Bertocchi, D., & Zurbarán, M., 2017. Enabling spatial autocorrelation mapping in QGIS: The hotspot analysis plugin. *Geoingegneria Ambientale E Mineraria*, *151*(2), 45–50.

Pernici, B., Francalanci,, C., Scalia G., Corsi, M., Grandoni, D., Biscardi, M.A., *Geolocating social media posts for emergency mapping*, Social Web in Emergency and Disaster Management, Los Angeles, CA, USA, 2018.

Sakaki, T., Okazaki, M., and Matsuo, Y., 2010. *Earthquake shakes Twitter users: real-time event detection by social sensors*. Proceedings of the 19th international conference on World wide web, New York: ACM Press, 851–860.

Saffir, H. S., 1973: Hurricane wind and storm surge. *Military Engineering*, 423, 4–5.

Shklovski, I., Burke, M., Kiesler, S., Kraut, R., 2010. Technology adoption and use in the aftermath of hurricane Katrina in New Orleans. *American Behavioral Scientist* 53 (8), 1228–1246. doi.org/10.1177/0002764209356252

Simpson, R. H., 1974: *The hurricane disaster potential scale. Weatherwise*, 27, 169, 186.

Spasenovic, K., Carrion, D., 2018. Crisis maps quality check and a proposal of integration with data from social media sources, Politecnico di Milano, Milan, Italy. hdl.handle.net/10589/142367

Spasenovic, K., Carrion, D., 2019. Quality check of crisis maps produced over five years by Copernicus EMS, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W13, 1263-1268, doi.org/10.5194/isprs-archives-XLII-2-W13-1263-2019.

Wartenberg, D., 1985. Multivariate spatial correlation: a method for exploratory geographical analysis. *Geographical Analysis*. 17(4), 263-283. doi.org/10.1111/j.1538-4632.1985.tb00849.x