Tracking Hurricane Dorian in GDELT and Twitter

Innocensia Owuor[1], Hartwig H. Hochmair[1] and Sreten Cvetojevic[2]

[1] University of Florida
[2] George Mason University

Abstract. GDELT is a machine coded database of events that uses both foreign and domestic news feeds and contains over a quarter of a billion worldwide event records categorized into three hundred categories. This paper compares the spatial footprint of GDELT event mentions with those of event related geo-tagged tweets for Hurricane Dorian in the South-Eastern United States. Besides examining event related GDELT and Twitter data abundance, the study relates areas of elevated GDELT news and tweeting activities to the locations of the hurricane track over a six-day period, and statistically analyzes distances between daily GDELT event mentions and tweets, and the hurricane center on different days. It assesses the potential role of the geographic coverage of the cone in hurricane prediction maps on the level of event related news and tweeting activities. The study also discusses pros and cons of both data sources for event tracking with regards to data abundance, spatial and temporal resolution, and thematic accuracy.

Keywords: hurricane, Dorian, tweets, GDELT, events, spatial analysis.

1   Introduction

Social media platforms have revolutionized communication patterns between users and the propagation of news as these platforms facilitate the sharing of content and instant responses in quasi real time [1, 2, 3]. The widespread use of social media platforms, such as Twitter and Facebook, provides an additional source of news to its users, though social media still play a relatively limited role as a way of finding news compared to television, news media websites or print newspapers, even among 18 to 24 year olds [4]. GDELT (Global Database of Events, Language, and Tone) relies primarily on the second, more traditional type of information channels. It is a news repository based on monitoring news media in over 100 languages across the world that is updated every 15 minutes. It provides a rich archive of events since 1979 and includes scores about tones and emotions in the news besides location and time of the event. The GDELT project contains over 250 million geocoded events with an even higher number of event mentions and location references, along with a massive Global Knowledge Graph (GKG) that connects people, organizations, locations, themes and emotions underlying those events [5].

Social media platforms have been extensively used for the detection, monitoring and tracking of man-made or natural events (e.g., earthquakes, floods, or revolutions), and analyzing their impact on society and its responses [5, 6, 7, 8]. Several studies also addressed the relationship between geographic and social space when analyzing community interaction in social media platforms [10]. Other studies evaluated the spatial and temporal accuracy of event representation in GDELT [11]. However, potential differences in spatial and temporal coverage of an event between social media platforms (especially Twitter) and GDELT are largely underexplored. This makes it difficult to decide which platform is most suitable in order to obtain the best overview over a specific event (e.g. its spatial or temporal extent) and its consequences for the local population. To address this issue, this study uses Hurricane Dorian, a category 5 hurricane, which hit the Bahamas and part of the Southeastern U.S. in September 2019, as a showcase.

The study has the following objectives: 1) Compare the abundance and spatial resolution of data points (GDELT event mentions and tweets, respectively) related to this event for the Southeastern U.S. (Florida, Georgia, South and North Carolina); 2) examine the proximity of daily GDELT event points and tweets to the hurricane eye throughout a 6-day observation period; 3) explore the effect of the geographic coverage of the cone used in daily hurricane forecast maps on GDELT event mentions and tweeting activities. Twitter's relative data sparsity of geo-tagged tweets, with only 1-2% of tweets being annotated with geographical coordinates, have long been recognized as a challenge for geo-applications and analyses using geo-tagged tweets [12, 13]. In this context, this study will evaluate the potential of GDELT to represent a viable data alternative to tweets for the purpose of determining areas affected by Hurricane Dorian.

2 Literature Review

The enormous growth of data generated by social media results in an abundance of online information that facilitates event detection and related applications, such as emergency, traffic, and health and management, rescue operations [14], tracking of refugee movements [13], delineation of areas affected by local events, such as floods [15], or compliance assessment with governmental regulations before and during natural disasters [16]. A wide range of event detection and event extraction methods from social media and more generally, crowd-sourced geo-information data have been developed and classified over the years [14].

Whereas specifically tweets, and some other prominent crowd-sourced data sources, such as Flickr [17] or OpenStreetMap (OSM) [18] for event detection and management have been frequently discussed in the literature, the role of GDELT for mapping and tracking of events is less known in the geo-community. An earlier study addressing spatial aspects of the GDELT platform [19] analyzed 10,009 news media and 195,513 disasters appearing in the GKG between April 2013 and July 2014 in a hierarchical (mixed-effect) regression model, finding that disasters received more attention if they occurred in politically instable countries, if many people were

affected, if counts of kidnappings and killings were provided, and if disasters were aftershocks, radiation leaks, flooding, ice, or landslides. A validation of GDELT data for five event types occurring in Sudan or South Sudan between April 2013 and November 2013 found that, on average, 81.2 percent of the event codes in the database accurately reflect the nature of the articles, and that the geography of events in the database is correct 75.4 percent of the time [11]. A paper presenting a taxonomy of event extraction from Twitter [20] points out that the use of GDELT for the evaluation of event extraction from Twitter is limited by the coarse spatial level (city) and the lack of semantic elements in GDELT event messages.

Only a few studies analyze the spatial overlap between GDELT news coverage and that from crowd-sourced data. One study, for example, compared the spatial extent of the GDELT report density with the OSM edit density for the Moore Tornado in Oklahoma (November 2013) and Typhoon Haiyan (May 2013) [21]. Results revealed that the majority of OSM edits were within the location of where the natural disaster occurred while GDELT reports were predominantly located in the largest city near the disaster affected location.

Previous research showed that both social networks and news media rapidly increase their attention to natural disasters especially within the first 72 hours, as demonstrated for an earthquake in GDELT [7] or for a typhoon in Twitter [22], among other examples, such as hurricanes [23]. In the case of tweets, hashtags were found to be important semantic features since they help to identify the topic of a tweet and to estimate the topical cohesiveness of a set of tweets [24]. The frequency of disaster related tweets was found to be highest in the spatial proximity of a disaster [25].

3   Data and Methods

This research covers four U.S. states, namely Florida, Georgia, North Carolina, and South Carolina, and analyzes hurricane related GDELT event mentions and tweets within that study area shared between September 1 and September 6, 2019. Dorian became a hurricane on August 28 north of the Greater Antilles, made landfall on the Bahamas on September 1 as a Category 5 Hurricane, and began moving northwestward on September 3, parallel to the coast of Florida, after weakening considerably. On September 6 it made landfall on Cape Hatteras in North Carolina at Category 1 intensity, after which it headed northeast as a tropical storm before dissipating near Greenland on September 10.

3.1   Tweets and GDELT event mentions

Geo-tagged tweets posted between September 1 and September 6, 2019 were downloaded for the study area through the Twitter Streaming API using the Tweepy python library and stored in a PostgreSQL database. Both tweets with precise coordinates as well as tweets geocoded with a place type city were downloaded. In June 2019, Twitter disabled precise location sharing from Twitter apps for mobile devices so that

precise location is available only through the camera by sharing a photo with the tweet[1]. Other tweets, when location sharing is on, share Twitter place (city level) position information by default, with an option of selecting a nearby point of interest (POI) on Twitter apps. For our research, most tweets used in the analysis came with Twitter place position information (i.e., city or POI). Although this information is spatially not as accurate as precise coordinate information, this was not viewed as a significant problem for the analysis given the geographic scope of the study which expands over several states. Tweets from automated Twitter accounts (bots) were subsequently removed using the Botometer web service [26], where Python was used to call the Botometer API for Twitter users in the analyzed dataset. The JSON object for each tweet contains among others a list of all hashtags that are used in a tweet. Hashtags provide a platform for the discussion of a specific topic and can be used to classify information [27]. Hashtag strings can be clicked to trigger a global search of tweets related to a topic of interest. To extract hurricane related tweets, tweets that contained the term "Dorian" or "Hurricane" in their hashtag or text were retrieved. This led to the retrieval of 8539 hurricane related tweets in the study area.

GDELT data can be mined in two ways. The first option is to run queries on Google Big Query for a fee, based on different available pricing schemes. A limited number of data can also be downloaded for free when a new account is opened. The second option is to download (for free) the three main tables of events, event mentions and the GKG that make up the GDELT database and which are updated every 15-minutes. The Events table contains event logs which are uniquely coded using a GLOBALEVENTID key, the Event Mentions table stores all the mentions that reference an event, and the GKG contains themes to which various events cited in the Event Mentions table belong to. These tables can then be integrated and analyzed in a database management system, such as PostgreSQL. For this study both methods were tested for the extraction of hurricane related GDELT event mentions, which returned the same results. However, about 200 GKG files could not be loaded into the PostgreSQL database due to data encoding errors. In Google Big Query, an SQL join between the three tables was run to retrieve all the event mentions about the hurricane. The SQL query is available for download on GitHub (see section 3.3). The extraction started by identifying all GKG records falling under the "HURRICANE" theme. Based on this, the Event mentions table was searched for HURRICANE related articles (mentions) during the study period, which finally led to searching the Events table for "HURRICANE" events. GDELT entries in the Events table were filtered to USCITY location type, which corresponds to the size of a US city or landmark. This is the finest spatial resolution for GDELT entries available, besides more coarse location types, such as COUNTRY, USSTATE, WORLDCITY, and WORLDSTATE. Each location in the GKG comes with a latitude/longitude pair which represents the centroid of the location.

The hurricane track position for every 12 hours between September 1 and September 6, 2019 was retrieved from the National Oceanic and Atmospheric

---

[1] https://twitter.com/TwitterSupport/status/1141039841993355264

Administration (NOAA) website[2]. Time stamps of tweets, GDELT event mentions and hurricane tracking points are all given in UTC and were therefore not converted to local time.

3.2 Analysis methods

Related to objective 1, descriptive statistics and maps will provide an overview of the abundance of daily hurricane related GDELT event mentions and tweets. Objective 2 determines how closely GDELT event mentions and tweets follow the hurricane path over the analyzed time period. Two general approaches were applied for this, as follows.

Since the hurricane eye was travelling mostly over the ocean, and GDELT news and tweets are primarily posted from land, we could not directly compare the locations between the news sources with the location of the hurricane eye. Instead, for the first approach, median distances were computed between GDELT/tweet locations and the location of the hurricane eye for all combinations of days between September 1 and 6. The hypothesis with this analysis is that the GDELT/tweet locations generally follow the hurricane eye closely on land, so that the distance between them and the hurricane eye is shortest when the positions of the same day are compared. For example, GDELT news/tweets from Sep 6 are expected to be closest to the September 6 hurricane tracking point. As opposed to this, the distance can be expected to be longer for other date combinations, such as when measuring distances from GDELT news/tweets shared on September 6 to the hurricane eye position on September 4 or 5. This hypothesis will be statistically tested through a series of Wilcoxon-signed ranked tests for GDELT and Twitter.

For the second approach, the number of GDELT event mentions and tweets per day are counted for all 413 counties in the four analyzed states. To reach a sufficient sample size only those counties with at least 30 tweets (n=48 counties) or at least 30 GDELT event mentions (n=23 counties), respectively, over the observed six-day period were retained for further analysis. Next, daily GDELT event mention and tweet numbers for these counties were corrected by a scaling factor to obtain the same total number of observations each day. Next, the mean number of GDELT event mentions and tweets was computed for each county across the six-day period, followed by the computation of the z-score for each day and county. The hypothesis related to this objective is that z-scores of GDELT news and tweet counts tend to be high for counties that are located near the hurricane eye on a given day, and that more distant counties have a lower z-score.

For objective 3, the potential effect of the spatial coverage of the hurricane forecast cone on activity levels of GDELT news sharing and tweeting is examined. The hypothesis is that areas under the cone, which have a higher chance to be hit by the hurricane in a matter of days according to forecasts, receive relatively more attention in GDELT and Twitter than areas outside the cone. To test the hypothesis, two consecutive dates are picked (September 2 and 3) which show little hurricane

---

[2] https://www.nhc.noaa.gov/archive/2019/DORIAN_graphics.php

movement in-between, but that show a significant change in the cone shape. In addition, a reference area close to the hurricane eye but outside the cone (on both dates) is chosen. The number of GDELT/tweet posts is counted for September 2 and 3 for the reference area and the cone area from September 2. If the hypothesis is true, one can expect a significant drop in GDELT/tweet counts in the September 2 cone area between September 2 and 3, relative to the drop in the reference area between both days.

3.3　Data and software availability

IDs of hurricane related tweets (in .txt file format) as well as records of GDELT news mentions (in .csv format) used in this study can be obtained from the data folder on https://github.com/InnocensiaO/Tracking-Hurricane-Dorian-in-GDELT-and-Twitter. Tweet IDs are shared as described in the Twitter API developer agreement and policy which gives permission for the distribution of this data for academic research purposes. The queries that were used to retrieve GDELT event mentions through Google Big Query can be found in the code folder. The hurricane cone graphics can be accessed at https://www.nhc.noaa.gov/archive/2019/DORIAN_graphics.php.

The computational environment used during this research included a Lenovo W530 laptop with a Windows 10 operating system, 24 Gigabyte RAM and a 2 TB hard disk. ArcGIS Pro 2.4.3 was employed to carry out spatial analysis and map creation. Statistical charts were created in R [28]. The R code can be found in the GitHub link above.

4　Results

4.1　Data abundance

Table 1 shows for the four analyzed U.S. states the daily counts of hurricane related GDELT event mentions and tweets between September 1 and September 6, 2019.

A total of 3340 event mentions were retrieved from the GDELT master database in comparison to a much higher number of 8539 tweets. Tweets are only more abundant, however, if tweets of place type city are included, since only 14.9% of geo-tagged tweets obtained for this study had exact coordinates. Among the 1274 tweets with exact latitude/longitude coordinates, all except for four came with images. Only three out of the 8539 city level tweets came from the Twitter Web Client source, all others came from Twitter apps for mobile devices. Both in GDELT and Twitter datasets daily record numbers are highest when the hurricane is close to the Miami Metropolitan Area (September 2 and 3), which can be expected due to the large population in that area. Numbers drop significantly once the hurricane reaches the northern fringes of the study area in North Carolina on Sep 6.

Table 1. Daily counts of hurricane related GDELT event mentions and tweets

Date	GDELT event mentions	Tweet counts	
	U.S. city level	Place type city or exact coordinates	Exact coordinates
9/1/2019	421	1673	229
9/2/2019	622	2276	360
9/3/2019	631	1604	249
9/4/2019	584	1177	189
9/5/2019	599	1183	162
9/6/2019	483	626	85
Total	**3340**	**8539**	**1274**

Fig 1 visualizes the count of hurricane related GDELT event mentions (a) and tweets (b) per city on September 3. The daily positions of the hurricane eye are plotted off the coast in the Atlantic Ocean. The map illustrates that on that day GDELT events and tweets are primarily found along the Florida East coast but only occasionally in the more northern states Georgia, South Carolina, and North Carolina. Furthermore, tweets, due to their relative abundance, appear in more cities than GDELT events, which is most noticeable along the Florida East coast.
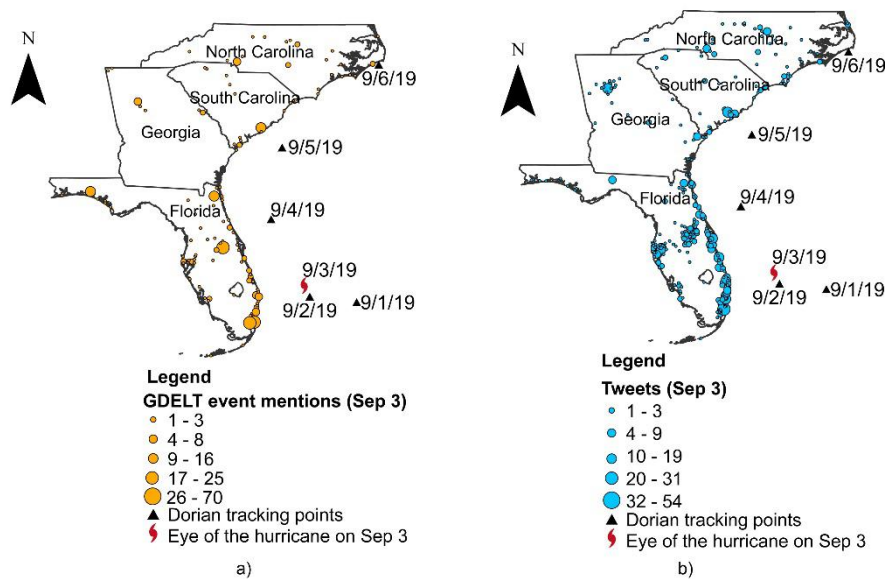
Fig. 1. GDELT event mentions (a) and tweets (b) per city on September 3.

4.2 Distance to hurricane tracking points

Fig 2 plots the median distances between GDELT event mentions (a) and tweets (b) for a given day (between September 1 and 6) and hurricane center points at a given day and time (provided in 12-hour increments). The latter is shown along the x-axis of each chart. The GDELT and Twitter sample size for each median computation can be obtained from Table 1. The line patterns reveal a drop in median distance during the first three days both in GDELT and Twitter, which can be expected as in these few days the hurricane eye moves towards the coast (compare Fig 1). Similarly, median distances increase towards the end of the analyzed time period where the hurricane eye shifts towards the fringe of the analyzed region. It is important to notice that the curves associated with different GDELT and tweet dates tend to have lowest points at different hurricane dates. This means that the average positions of GDELT or Twitter posts from such a day are also closest to the hurricane eye. For example, the GDELT September 5 curve has the lowest distance value among all six curves for the September 5 hurricane eye position. This pattern is more pronounced for Twitter than for GDELT which shows that Twitter more accurately follows the hurricane track. Exception to this are the distance curves for September 1 through 3 which nearly overlap in both figures. That is, the median distances for tweets between September 1 and 3 are similar since in the first 48 hours of the analysis the hurricane moved towards the Southeast Florida perpendicular to the coastline, and not in a lateral direction along the coast.

As can be seen, between the September 1 and 6 hurricane dates (i.e. the left-most and right-most position in the charts), the order of GDELT/Twitter distance lines flips. In other words, for the hurricane eye location on September 1, GDELT/Twitter posting locations from around that day (September 1-3) are closest to it, whereas for the September 6 hurricane eye position GDELT/Twitter positions from September 6 are closest.
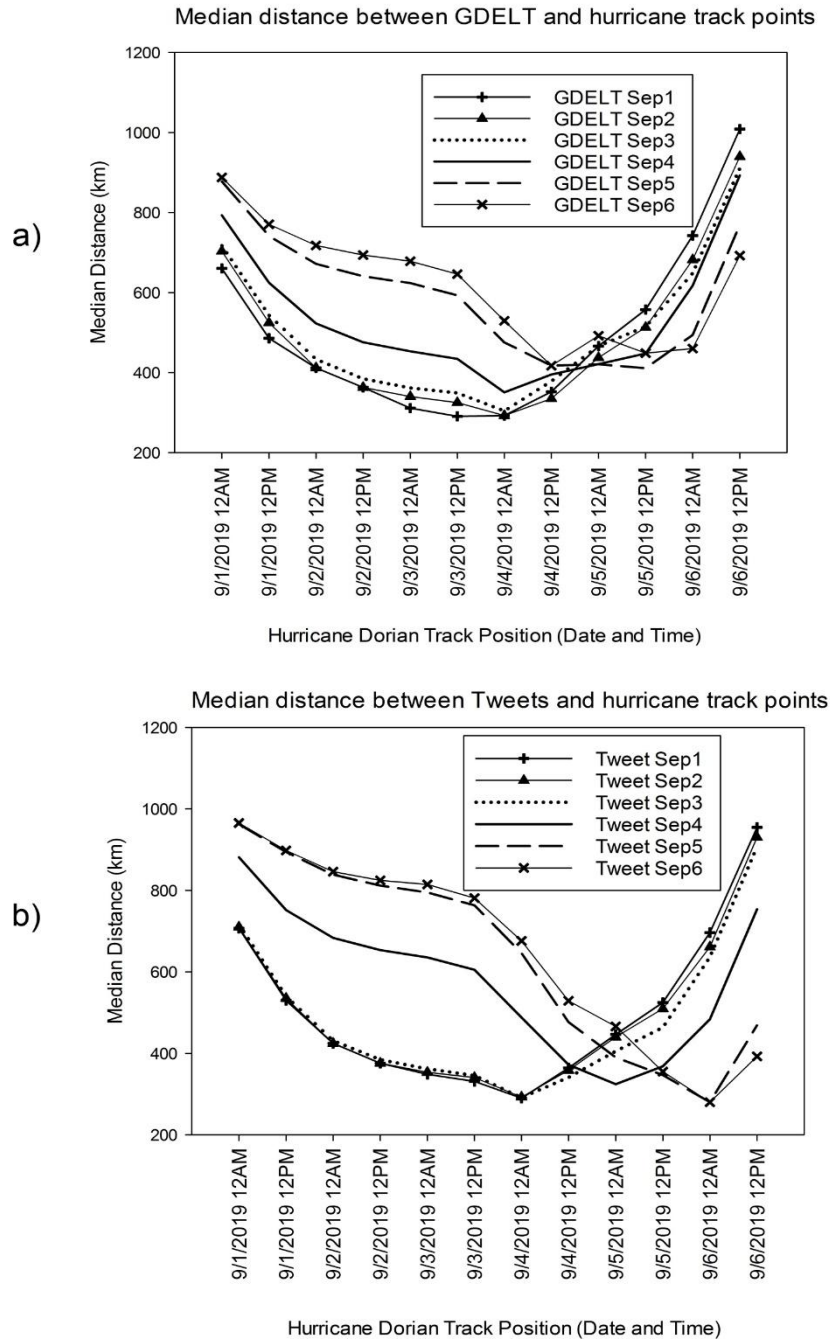
Median distance between GDELT and hurricane track points

a)

Median distance between Tweets and hurricane track points

b)

Fig. 2. Median distances from GDELT event mentions and tweets to hurricane Dorian eye locations on different dates

The boxplots in Fig. 3 provide a more detailed insight into the distribution of median distances between GDELT event mentions (a, b) and tweets (c, d) posted on September 1 through 6, and hurricane eye positions on September 1 and 6. The plots show that distances between GDELT positions and the September 1 hurricane location gradually increase from September 2 on (a), with the opposite trend for the September 6 hurricane location (b). For tweets, a similar pattern can be observed in Fig. 3c and d. This illustrates again the influence of the hurricane eye position on the locations from which GDELT event mentions and tweets are shared.
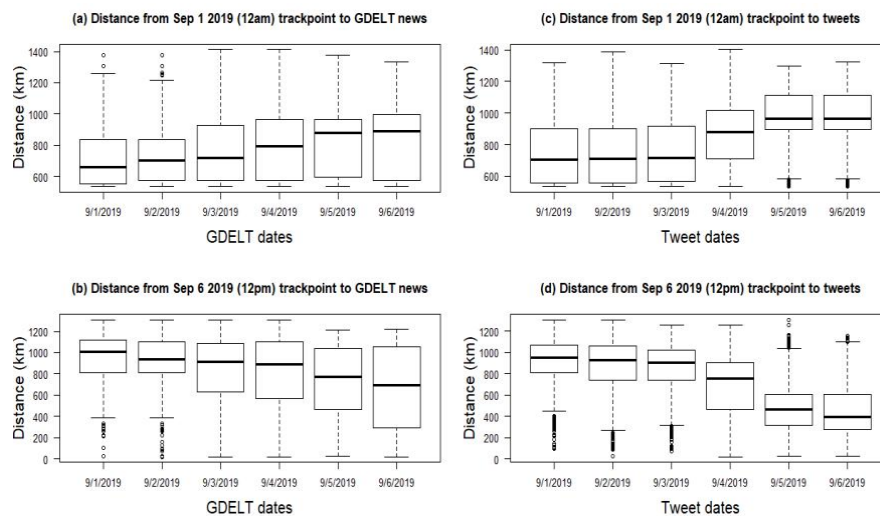

Fig. 3. Boxplot of distances between GDELT event mentions (a, b) and tweets (c, d) posted on September 1 through 6 (horizontal axis) and the eye of the hurricane on Sep 1 and 6

A series of Wilcoxon signed rank tests was carried out to determine if differences between median distances on consecutive days in Fig. 3 were statistically significant or not. Each row in Table 2 corresponds to the comparison of two adjacent box plots in the different charts shown in Fig. 3. With the exception of some date pairs at the beginning and end of the study period, differences in median distances were significant, which means that the center of GDELT and tweet activities, respectively, shifted between days relative to the hurricane eye position on a given day (September 1 or 6). Tweets had more differences that were statistically significant, indicating that the tweets tracked the hurricane path better than GDELT did.

Table 2. Wilcoxon signed rank test results (p<0.05 shown in boldface).

		GDELT event mentions		Tweets	
		Date/time of hurricane position			
		Sep1 12AM p-value	Sep6 12PM p-value	Sep1 12AM p-value	Sep6 12PM p-value
GDELT/	Sep1-Sep2	0.2640	0.2020	0.5210	**0.0021**
tweet	Sep2-Sep3	**0.0006**	**0.0036**	**0.0006**	**0.0000**
dates	Sep3-Sep4	**0.0042**	**0.0267**	**0.0000**	**0.0000**
	Sep4-Sep5	0.0525	**0.0000**	**0.0000**	**0.0000**
	Sep5-Sep6	0.8543	**0.0338**	0.8507	**0.0065**

4.3 z-score maps

Fig 4 and 5 show daily z-score maps for GDELT event mentions and tweets from September 1 through September 6. The position of the hurricane eye on the corresponding day is visualized as a red triangle. Resulting z-scores for each map were classified into five categories using Jenks natural breaks classification to facilitate an easier visual comparison between the different days. Class 1 covers the lowest range of z-scores and class 5 the highest. The maps show that z-scores for GDELT event mentions and tweets on a selected day tend to be highest in counties that are close to the eye of the hurricane. This pattern is more discernible for tweets (Fig 5) than for GDELT events mentions (Fig 4), suggesting that tweets are more suitable for event tracking. A possible explanation for this is the larger numbers of hurricane related tweets than GDELT events mentions (compare Table 1).
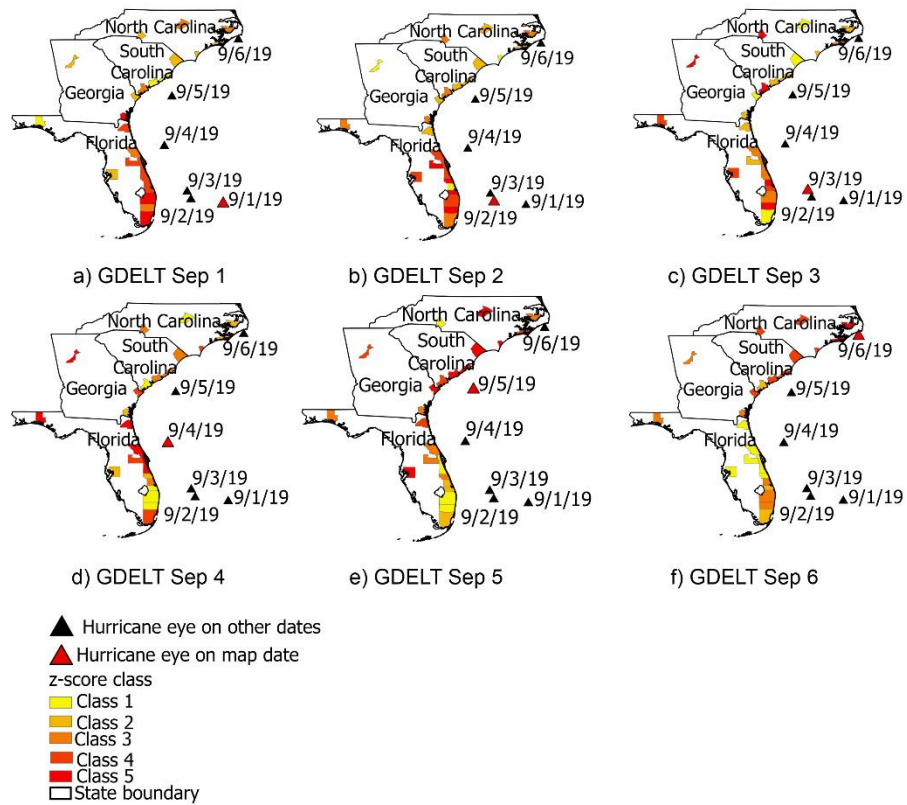
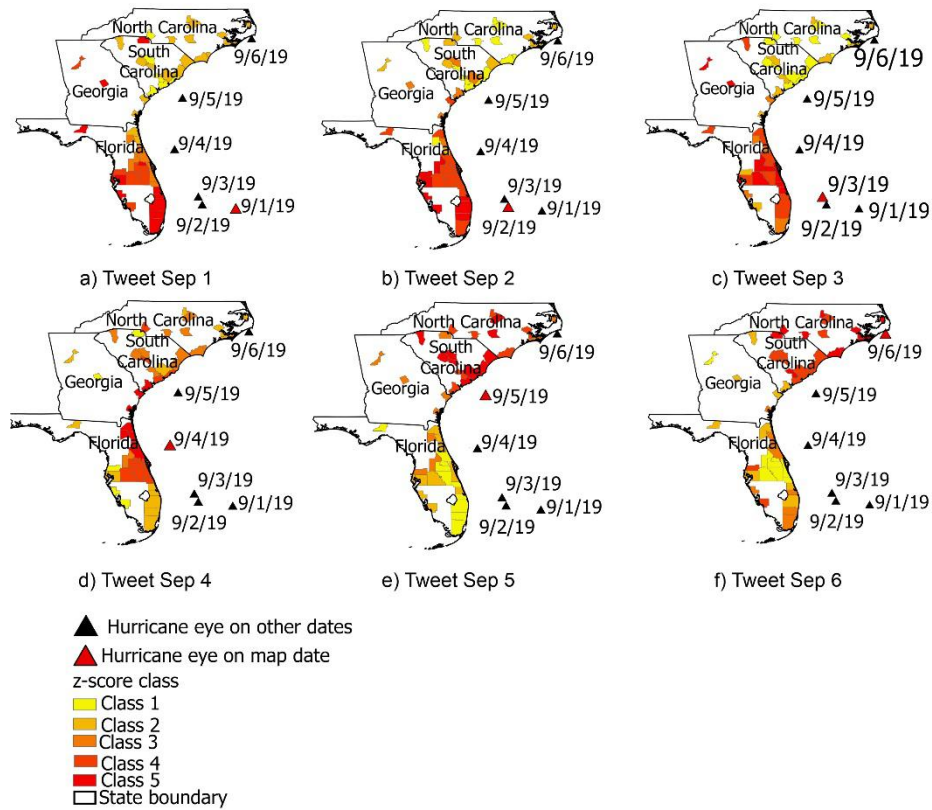**Fig. 4.** z-score maps of GDELT event mentions.

Fig. 5. z-score maps of tweets.

The Pearson correlation between GDELT event mentions and tweets in 224 counties with at least one recorded point in either data source results in r = 0.78 (p < 0.001). This shows that both news sources share similar spatial patterns, but that they may still be able to complement each other since there is no perfect match. Fig 6 shows a corresponding scatter plot of the count of GDELT event mentions versus tweets in these counties. One outlier in Miami-Dade County revealed an unexpectedly high number of GDELT contributions, possibly due to a high number of news stations in that area. Removing this outlier leads to an increased r = 0.83 (p < 0.001).
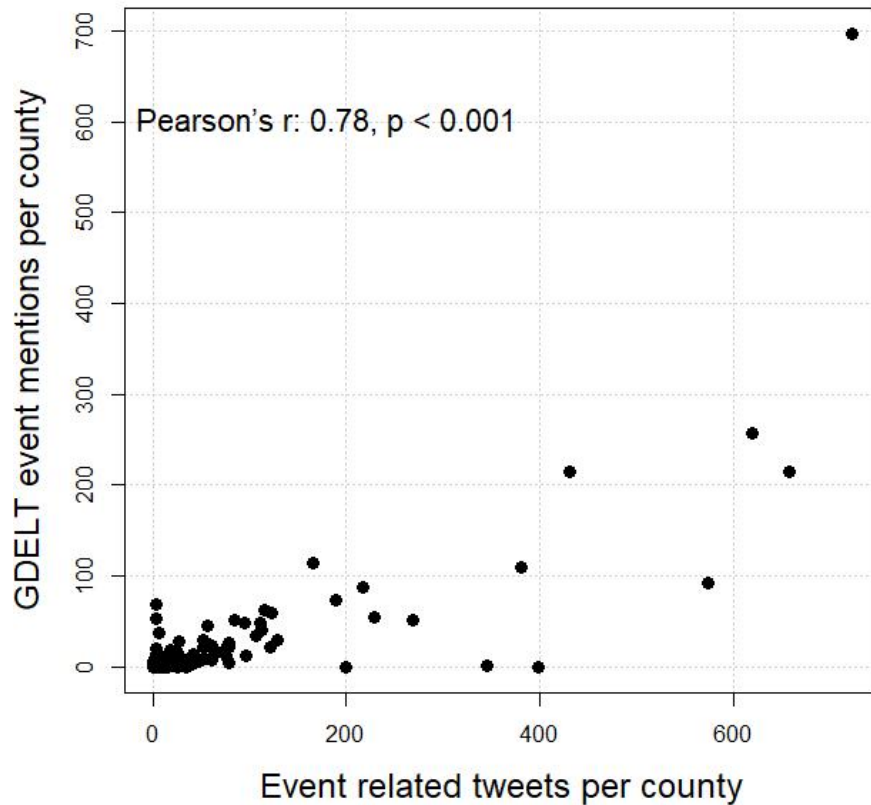
Fig. 6. Number of GDELT event mentions and tweets per county among counties with at least one contribution between September 1 and 6

4.4 Hurricane cone map and event data

Fig 7a and c show the locations of GDELT event mentions and tweets on September 2 within the hurricane prediction cone from September 2 (green dots) and within the reference area (red dots). The upper row in Table 3 reports corresponding counts of GDELT event mentions and tweets in these two areas, respectively, as well as the ratio of counts between the two areas. Fig. 7b and d map the same type of information for September 3, where, however, now the green dots show the locations of GDELT event mentions and tweets within the September 3 hurricane cone. What can be observed is that, although on September 3 the cone turned significantly away from the coast towards the ocean, the pattern of GDELT event mentions and tweets along the coast barely changes. This implies that the coverage of the cone does not noticeably affect news and tweeting activities. Instead, it appears that the distance to the hurricane eye is the driving factor of news and social media activities. If the hypothesis related to objective 3 was true, fewer GDELT mentions and tweets should

be observed in the September 2 cone on September 3 than on September 2 relative to the reference area. However, that ratio increased from 1.19 to 1.31 for GDELT and from 1.23 to 1.89 for tweets, which means that people were still actively tweeting from the same areas on September 3 although the cone had already shifted away towards east. Therefore, the hypothesis needs to be rejected since the cone area has no apparent effect on GDELT news mentions and tweeting activities.
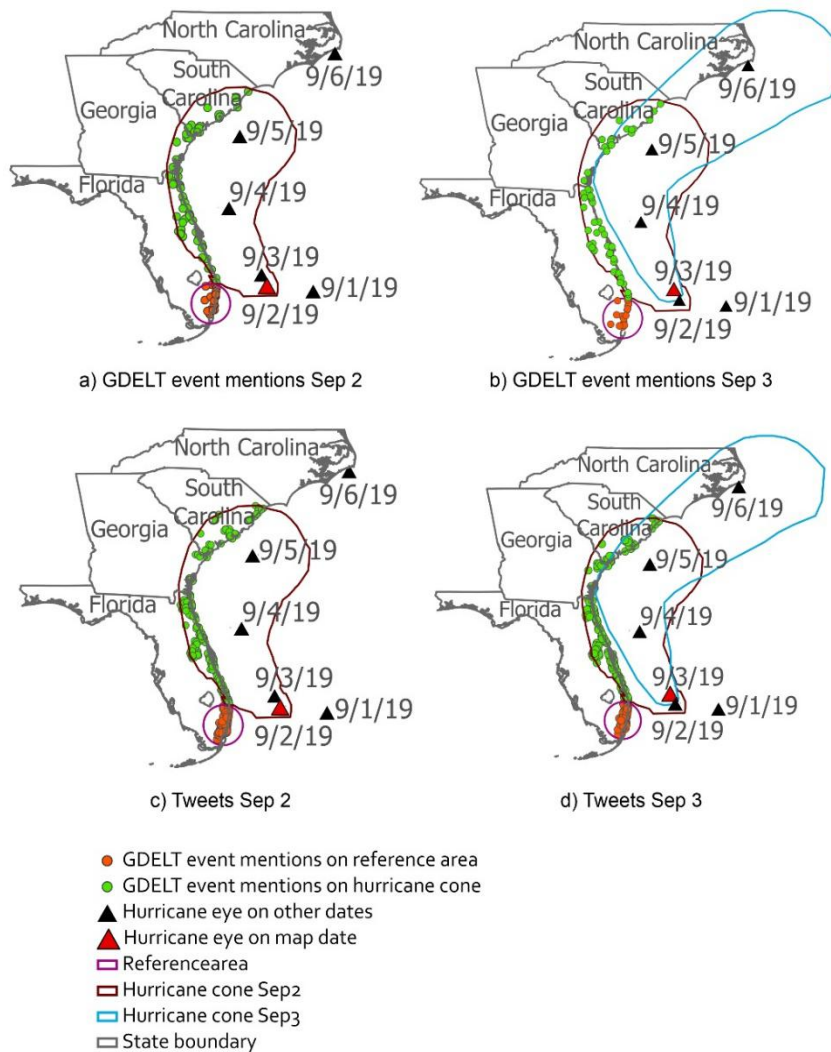

Fig. 7. GDELT event mentions and tweets inside the September 2 (brown) and September 3 (blue) hurricane cones and the reference area (magenta) on September 2 and 3, respectively.

Table 3. Counts of GDELT event mentions and tweets in September 2 hurricane cone and reference area

	GDELT event mention			**Tweets**		
Date	**Sep 2 cone**	**Reference area**	**Ratio**	**Sep 2 cone**	**Reference area**	**Ratio**
9/2/2019	230	193	1.19	758	618	1.23
9/3/2019	201	154	1.31	613	324	1.89

5   Discussion and Conclusions

GDELT event mentions and tweets are two examples of big data platforms with free data access that provide significant potential for tracking of natural disasters, such as hurricanes. The goal of this study was to compare the suitability of these sources for the tracking of Hurricane Dorian over a six-day period. The Twitter data count was about by a factor 2.6 higher compared to the GDELT event mentions when using city level data, hence Twitter is more suitable than GDELT in this regard. Twitter also provides the option to use tweets with exact coordinates, which facilitates more accurate event mapping than GDELT does, however, at the cost of a lower sample size. Distance analysis between GDELT mentions/tweets and the hurricane center across the six days suggests that the average position of data points both platforms generally follows the hurricane path on land, and that Twitter is more accurate in following the hurricane path than GDELT. A combination of both GDELT and Twitter will lead to an increased sample size and could therefore also improve spatial and temporal accuracy for event detection and monitoring, which will be explored as part of future work.

The Twitter streaming API provides tweets in quasi-real time, which makes it possible (though technically challenging through presence of bots, etc.) to monitor and map events in real-time. For GDELT the update rate is every 15 minutes. Depending on the type of analysis this may or may not be an issue. Extraction of tweets linked to a particular event poses some challenges, as this is typically done using hashtag filters or text mining, which may miss tweets that are not tagged accordingly [29]. This issue aside, tweets can, however, be used to study the spatio-temporal characteristics of any topic of interest (e.g. food, health, tourism, politics, travel). A challenge with extracting topical data from GDELT is that direct searching of keywords on GDELT event mentions is not possible through Google BigQuery. Instead, one needs to first pick a theme from a long list of pre-defined themes in the GKG table that the event of interest may fall under. One obvious advantage of GDELT over tweets is its historic dimension since its events date back to 1979, whereas access to older tweets is technically more challenging if not impossible.

Previous research has shown that during an event, areas closest to the event receive the most attention. This trend is generally confirmed through z-score maps in this study, more so for tweets than for GDELT event mentions. This suggests that tweets are a more reliable data source for the spatial analysis, e.g. tracking or regional impact assessment, of natural disasters. An expected association between the spatial coverage of a hurricane prediction cone and GDELT news/tweeting activities in these areas could not be confirmed. Future work will expand this analysis to other event types, such as protests, elections, sports (e.g. Olympic games) and other natural disasters, such as flooding.

References

1. Sakaki, T., Okazaki, M., and Matsuo, Y. Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors. In Proceedings of the 19th international conference on World Wide Web (pp. 851-860). New York, NY: ACM (2010).
2. Middleton, S. E., Middleton, L., and Modafferi, S. Real-Time Crisis Mapping of Natural Disasters Using Social Media. IEEE Intelligent Systems, 29 (2), 9-17 (2014).
3. Cvetojevic, S. and Hochmair, H. H. Analyzing the spread of tweets in response to Paris attacks. Computers, Environment and Urban Systems, 71 (September 2018), 14-26 (2018).
4. Nielsen, R. K. and Schrøder, K. C. The Relative Importance of Social Media for Accessing, Finding, and Engaging with News. Digital Journalism 2(4), 472-489 (2014).
5. GDELT The GDELT Project. https://www.gdeltproject.org/about.html, last accessed 12/28/2019
6. Crooks, A., Croitoru, A., Stefanidis, A., and Radzikowski, J. #Earthquake: Twitter as a Distributed Sensor System. Transactions in GIS, 17 (1), 124–147 (2013).
7. Su, Y., Lan, Z., Lin, Y., Comfort, L.K. and Joshi, J. Tracking Disaster Response and Relief Following the 2015 Nepal Earthquake, 2016 IEEE 2nd International Conference on Collaboration and Internet Computing (CIC), Pittsburgh, PA, 2016, pp. 495-499, doi: 10.1109/CIC.2016.075 (2016).
8. Kharroub, T. and Bas, O. Social media and protests: An examination of Twitter images of the 2011 Egyptian revolution. New Media & Society, 18 (2), 1973–1992 (2016).
9. Ahmouda, A. and Hochmair, H. H. Using Volunteered Geographic Information to Measure Name Changes of Artificial Geographical Features as a Result of Political Changes: A Libya Case Study. GeoJournal, 83 (2), 237–255 (2018).
10. Takhteyev, Y., Gruzd, A., and Wellman, B. Geography of Twitter networks. Social Networks, 34 (1), 73–81 (2012).
11. Racette, M. P., Smith, C. T., Cunningham, M. P., Heekin, T. A., Lemley, J. P., and Mathieu, R. S. Improving situational awareness for humanitarian logistics through predictive modeling. 2014 Systems and Information Engineering Design Symposium (SIEDS), Charlottesville, VA. IEEE (2014).
12. Kotzias, D., Lappas, T., and Gunopulos, D. Addressing the Sparsity of Location Information on Twitter. Workshop of the EDBT/ICDT 2014 Joint Conference, Athens, Greece (2014).
13. Hübl, F., Cvetojevic, S., Hochmair, H. H., and Paulus, G. Analyzing Refugee Migration Patterns using Geo-tagged Tweets. ISPRS International Journal of Geo- Information, 6 (10), 302 (2017).
14. Steiger, E., de Albuquerque, J. P., and Zipf, A. An Advanced Systematic Literature Review on Spatiotemporal Analyses of Twitter Data. Transactions in GIS, 19 (6), 809–834 (2015).
15. Shelton, T., Poorthuis, A., Graham, M., and Zook, M. Mapping the data shadows of Hurricane Sandy: Uncovering the sociospatial dimensions of 'big data'. Geoforum, 52, 167-179 (2014).
16. Martín, Y., Li, Z., and Cutter, S. L. Leveraging Twitter to gauge evacuation compliance: Spatiotemporal analysis of Hurricane Matthew. PLoS ONE, 12 (7), e0181701 (2017).

17. Chen, L. and Roy, A. Event Detection from Flickr Data through Wavelet-based Spatial Analysis. 18th ACM conference on Information and Knowledge Management, Hong Kong, China. ACM (2009).

18. Dittus, M., Quattrone, G., and Capra, L. Mass Participation During Emergency Response: Event-centric Crowdsourcing in Humanitarian Mapping. In 20th ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW 2017). New York, NY: ACM (2017).

19. Kwak, H. and An, J. A First Look at Global News Coverage of Disasters by Using the GDELT Dataset. In L. M. Aiello & D. McFarland (Eds.), Social Informatics (LNCS 8851, pp. 300-308). Berlin: Springer (2014).

20. Deng, J., Qiao, F., Li, H., Zhang, X., and Wang, H. An Overview of Event Extraction from Twitter. 2015 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery, Xi'an, China IEEE (2015).

21. Parker, S. M. Patterns of Geographical Content Contributions Made by the General Public in Response to Natural Disaster Events (Diploma Thesis, George Mason University, Fairfax, VA) (2015).

22. David, C. C., Ong, J. C., and Legara, E. F. T. Tweeting Supertyphoon Haiyan: Evolving Functions of Twitter during and after a Disaster Event. PLoS ONE, 11 (3), e0150190 (2016).

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

24. Atefeh, F. and Khreich, W. A Survey of Techniques for Event Detection in Twitter Computational Intelligence, 31 (1), 132-164 (2015).

25. Herfort, B., Albuquerque, J. P. d., Schelhorn, S.-J., and Zipf, A. Exploring the Geographical Relations Between Social Media and Flood Phenomena to Improve Situational Awareness. In J. Huerta, S. Schade & C. Granell (Eds.), Connecting a Digital Europe Through Location and Place (Lecture Notes in Geoinformation and Cartography, pp. 55-71). Berlin: Springer (2014).

26. Davis, C. A., Varol, O., Ferrara, E., Flammini, A., and Menczer, F. BotOrNot: A System to Evaluate Social Bots. WWW '16 Companion: Proceedings of the 25th International Conference Companion on World Wide Web, Montréal, Québec, Canada. ACM (2016).

27. Chong, M. Sentiment Analysis and Topic Extraction of the Twitter Network of #Prayforparis. Proceedings of the Association for Information Science and Technology, 53 (1), 1-4 (2016).

28. R Core Team. R: A language and environment for statistical computing. https://www.r-project.org/ (2019).

29. Cheng, T. and Wicks, T. Event Detection using Twitter: A Spatio-Temporal Approach. PLoS ONE, 9 (6), e97807 (2014).