Identifying Disaster-Impacted Areas with Machine Learning and Geospatial Analysis Using Social Media Data

Clemens Havas and Bernd Resch
Department of Geoinformatics–Z_GIS, University of Salzburg, Austria

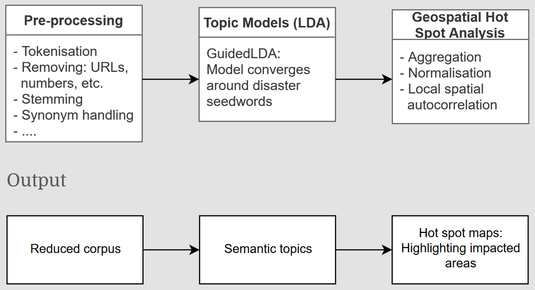
AAG Annual Meeting, April 3-7, 2019, Washington, DC

Introduction

Emergency management relies on up-to-date information about the impact of a disaster in an area. However, current emergency systems rely on data sources that do not provide a complete geospatial and temporal view of the disaster or have a temporal lag due to activation or orbital constraints. To fill these important information gaps, georeferenced social media posts are analyzed in near-real time. The Evolution of Copernicus Services (E2mC) project aims at demonstrating the technical and operational feasibility of the integration of social media analysis and crowdsourcing by developing a prototype of the innovative Copernicus Witness, a new Copernicus Emergency Management Service component. One module of this component focuses on analyzing the social media data stream and providing a "big picture" of the area of interest.

Extraction of Relevant Information

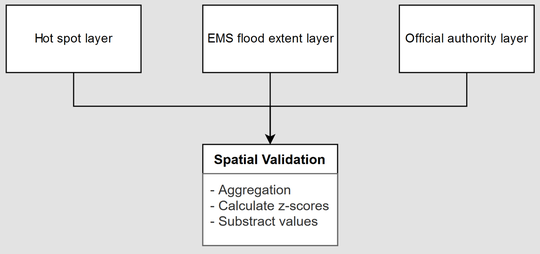
Z_GIS' Geocrawler: Software to crawl georeferenced social media posts of multiple social media networks e.g. Twitter, Flickr, YouTube, Foursquare.



Output

Result Validation

- Spatial validation of hot spot and EMS flood extent layer
- Adding official authority dataset for "ground truthing"



Use Cases

	Start	End	G-Tweets	Area[km²]
Hurricane Harvey	2017-08-25	2017-09-07	100,187	263,282
Hurricane Florence	2018-09-12	2018-09-29	1,063,166	2,931,804

Validation Harvey

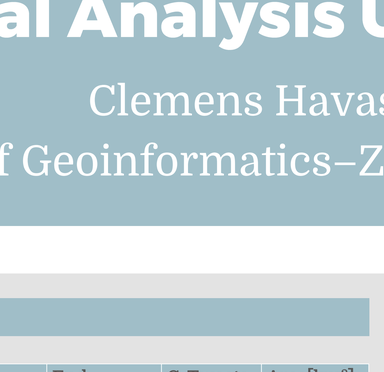
- U.S. Geological Survey 2,123 geo-high water marks
- Emergency Management Service flood outlines

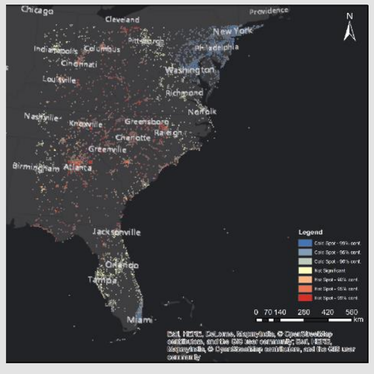
Hurricane Florence 2018 (EMSR311)

- Sept 12th : Crawler started
- Sept 14th : Hurricane made landfall
- Sept 15th : First hot spot maps

Areas impacted by hurricane Florence (coastal regions): Hot spots
→ display trajectory

Low activity in Atlanta and Washington: Cold spots





Overview of disaster-impacted areas on the east coast

From South Carolina to Cleveland affected by the hurricane: Hot spots

New York and Florida: Cold spots

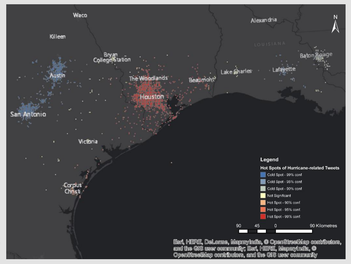
Hurricane Harvey 2017 (EMSR229)

6,706 hurricane-related tweets

Houston: and Coastal regions: Hot spots

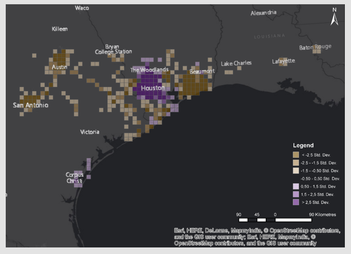
San Antonio (SATX) and Austin: Cold spots

Large number of weather station's bots → partly creating hot spots



Validation

Purple: Hot spot map; Brown: EMS flood extent; Green: High water marks



The major part of Houston was **not** identified by EMS

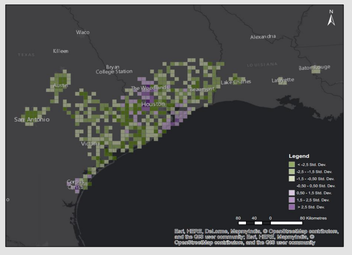
- Difficulty to provide results with SAR data

Brown areas in San Antonio and Austin: Witness identified non-affected areas (EMS did not deliver unambiguous results)

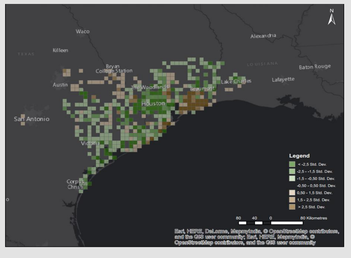
Both layers: Flooded areas around Houston

Hot spots: Stronger focus on Houston

High water marks: Stronger focus on the area around Houston



SATX, Austin, Lafayette and BRLA in green because cold spots in hot spot layer



EMS did not analyze the full area around Houston especially along the coast

EMS highlights the area around Beaumont

"Checkerboard" pattern

Conclusion

- Extraction of latent relevant semantic topics and classification of social media posts
  - ➢ Identification of dynamic keywords e.g. hashtags

word	probability
hurricaneharvey	0.2520
humid	0.0941
weather	0.0746
pressur	0.0667
feel	0.0612
temperatur	0.0276
cloud	0.0163
current	0.0154
houston	0.0135
baromet	0.0110

- Hardly any prior information needed to identify disaster-impacted areas and distinctly showing them with hot and cold spots
- Multimodality: Application of semantic and geospatial methods on a temporal and geographical filtered dataset to create relevant information
- Spatial validation by comparing hot spot layer, EMS flood extent layer and an official authority layer.
  → Showing the information value of hot spot maps.
- Fill the information gap: Create hot spot maps in the first hours of a disaster and continuously update them

Outlook & Further Reading

- Automatic determination of the „best" hyperparameter set
  - Evaluation of geospatial metrics
  - Use of geospatial metrics to assess the quality of topic modelling
- Focus on different kinds of natural disasters
- Comparison of analysis results over time

C. Havas et al., "E2mC: Improving Emergency Management Service Practice through Social Media and Crowdsourcing Analysis in Near Real Time," Sensors, vol. 17, no. 12, 2017.

B. Resch, F. Usländer, and C. Havas, "Combining machine-learning topic models and spatiotemporal analysis of social media data for disaster footprint and damage assessment," Cartogr. Geogr. Inf. Sci., pp. 1–15, 2017.

Acknowledgements

We would like to thank:
- the European Commission's Horizon 2020 program for supporting the project "E2mC - Evolution of Emergency Copernicus services", reference number 730082.
- the Doctoral College GIScience [DK W1237-N23] at the Department of Geoinformatics - Z_GIS, University of Salzburg, Austria, funded by the Austrian Science Fund (FWF) for their support.
- Dr. Wendy Guan from Harvard University's Center for Geographic Analysis for her support through providing us with Twitter data for our study.