Analyzing the Impact of Natural Disasters:
A Concept for Spatio-temporal Analyses of Social Media Data

Jens Kersten[1](jens.kersten@dlr.de), Anna Kruspe[1], Friederike Klan[1]
[1]German Aerospace Center (DLR), Jena, Germany

Introduction

Problem:
- Analysis of Twitter messages for crisis response activities
- Analyzing the dynamics of reactions caused by disasters taking into account space and time to reveal insights regardin glocal event impacts on population and environment

Contributions:
- Identification of suitable approaches for spatio-temporal analyses of Twitter data
- General workflow (Fig. 1) for spatio-temporal social media data
- Experiments with Twitter data recorded during hurricane Florence (September 2018)
- Formulation of future work directions

Proposed Workflow

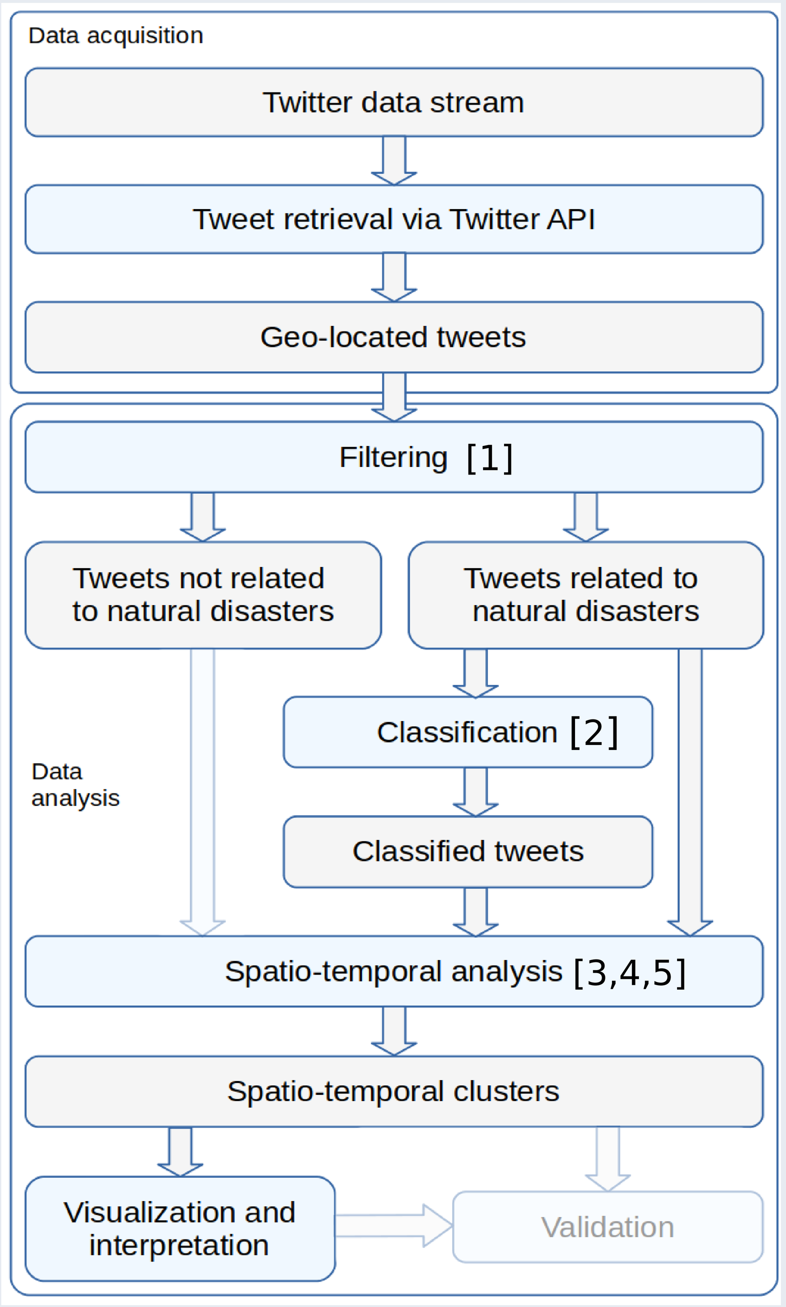

Fig. 1: Proposed workflow for the spatio-temporal analysis of social media data. Transparent elements are part of future work.

Dataset

- 600,000 geo-located Tweets
- Hurricane Florence, Carolinas, USA (Fig. 2)
- Record period September 12-19, 2018
- 30,700 crisis-related tweets with geo-location after CNN-based filtering [1]

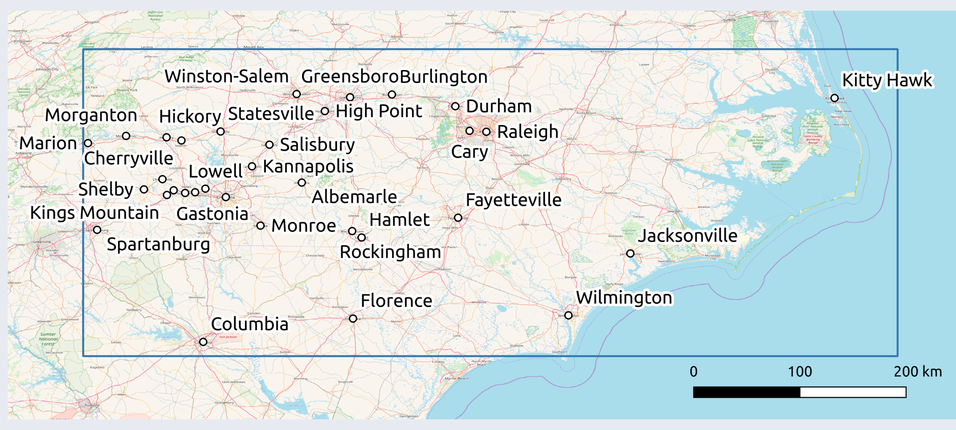

Fig. 2: Tweet acquisition area of interest (AOI).

Spatio-temporal Analyses

A. Local hot-spot detection (Getis-Ord G* [3])
- Hot- and cold-spot analysis (Fig. 3)
- Clustering of high and low tweet occurrences (crisis-related) on county level
- Occurrences normalized by population density [6]

B. Space-time kernel density estimation (ST-KDE [4])
- Non-parametric approach to compute a density function
- Describes the intensity of geographic events' distribution
- High activity areas (reflecting population densities): Charlotte, Raleigh and other cities (Fig. 4)
- Higher activities after landfall (Sep. 14)

C. STDBSCAN [5]
- Spatio-temporal extension of DBSCAN (Density-Based Spatial Clustering of Applications with Noise)
- Density associated to a point is obtained by counting the number of surrounding points
- Discovers clusters with arbitrary shapes
- Estimates the number of clusters

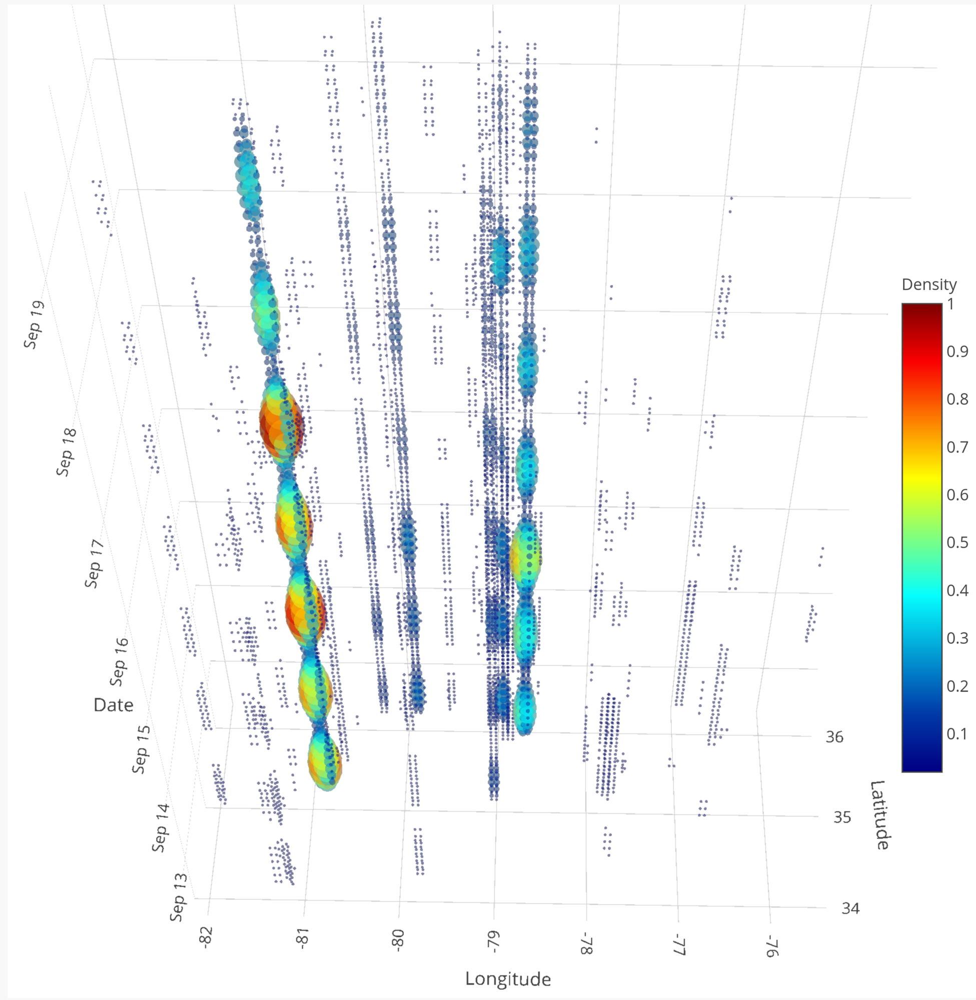

Fig. 4: ST-KDE result, spatial resolution = 5 km, temporal resolution = 0.1 days. Spatial and temporal bandwidths: 7.5 km, 0.3 days.

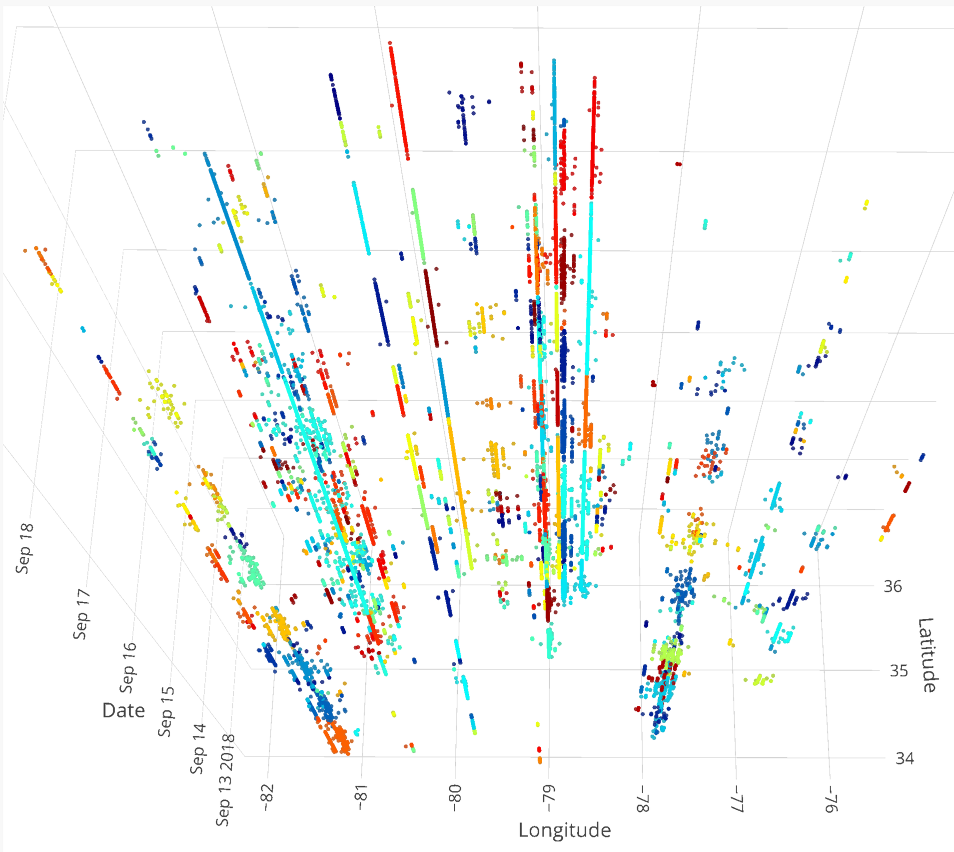

Fig. 5: STDBSCAN result, spatial threshold=10km, temporal threshold=1h, minimum number of points=5. The 430 clusters are randomly colored.

Results and Cluster Analysis

G*-statistics
- High population density around Charlotte → low relative number of crisis-related tweets

ST-KDE
- Spatio-temporal density visualization
- Daily activity patterns and increased activity after the landfall

STDBSCAN
- Densely populated areas tend to produce clusters covering more than one day (Fig. 5)

Analysis of cluster contents: classification of tweets into 7 information classes [2] (Fig. 6)
- Large clusters
  → Similar distributions covering all classes
  → Similar keyword maps (Fig. 7)
- Peak of tweets related to affected individuals on Sept. 15
- More tweets related to donations and volunteering after the landfall

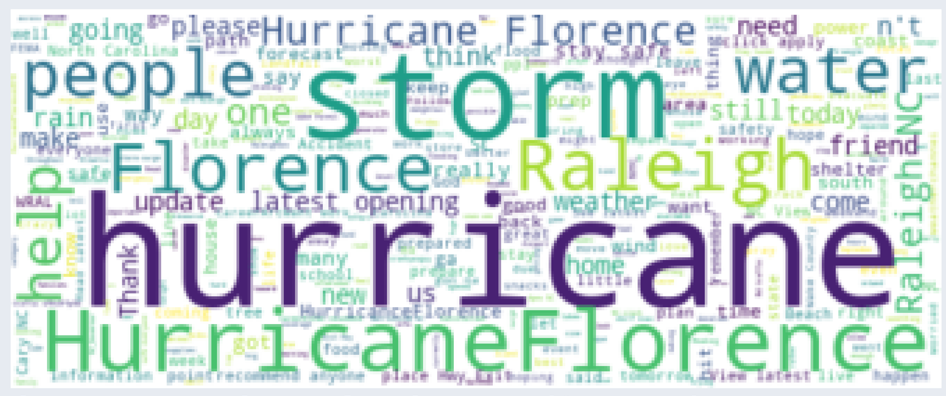

Fig. 7: Keyword map for cluster 4 (Sept. 15).

Conclusions and Outlook

- Workflow for spatio-temporal analysis and visualization of crisis-related tweets
- State-of-the-art CNNs used for filtering as well as tweet classification
- Case study: Hurricane Florence, Sep. 2018
- First qualitative results with G*-statistics, ST-KDE, and STDBSCAN
- Visualization of hot-spots and spatio-temporal patterns

Open problems and possible solutions
- Data sparsity using geo-located Tweets → Additional acquisition and use of tweets from keyword-based search
- Choice of method parameters → Systematic analyses required
- Large STDBSCAN clusters with similar contents → Incorporation of class labels for clustering + parameter tuning
- Validation of results → Benchmark data for testing + checking against (sub-) events reported in the news

References

[1] J. Kersten, et al., Robust filtering of crisis-related tweets, ISCRAM, Valencia, Spain, May 19-22, 2019, accepted, in press.

[2] A. Kruspe, et al., Classification of incident-related tweets: Tackling imbalanced training data using hybrid CNNs and translation-based data augmentation, TREC 2018, vol. 27.

[3] J. K. Ord and A. Getis, Local spatial autocorrelation statistics: Distributional issues and an application, Geographical Analysis, vol. 27, no. 4, pp. 286–306, 1995.

[4] J. Lee, et al., Exploring spatiotemporal clusters based on extended kernel estimation methods, International Journal of Geographical Information Science, vol. 31, no. 6, pp. 1154–1177, 2017.

[5] D. Birant and A. Kut, ST-DBSCAN: An algorithm for clustering spatial-temporal data, Data and Knowledge Engineering, vol. 60, no. 1, pp. 208–221, 2007.

[6] ESRI, https://www.arcgis.com/home/item.html?id=ff48bbae433442a38f6c635b8c7baf72, 2016 Population Density by Congressional District, accessed: March 3, 2019.

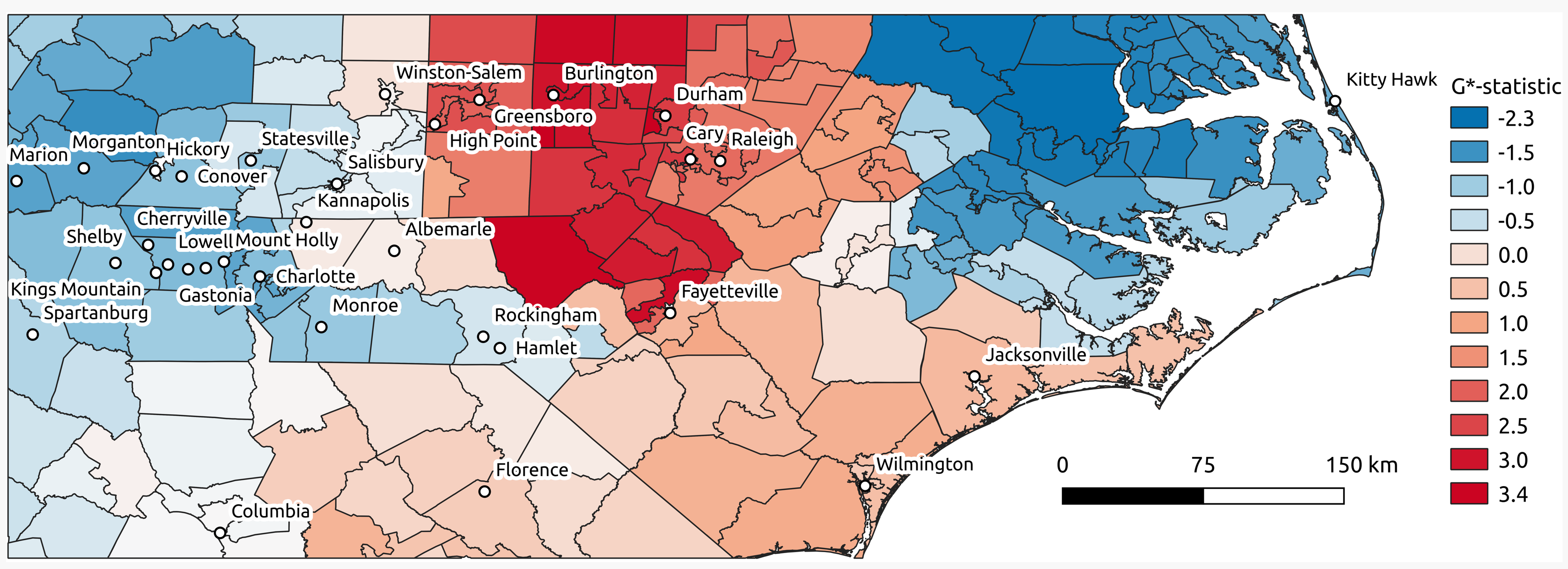

Fig. 3: Getis-Ord G* statistics result.

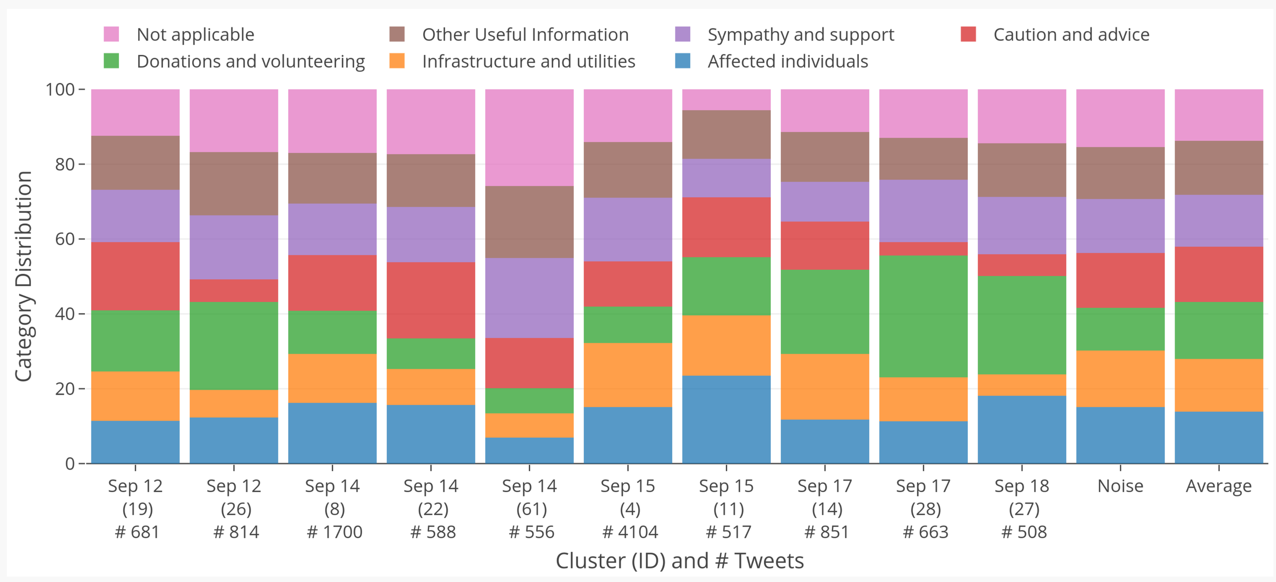

Fig. 6: Category distribution for the 10 largest STDBSCAN clusters compared to the average distribution of all clusters and of the ~4.800 noise points.

Deutsches Zentrum für Luft- und Raumfahrt
DLR