Studying Human Behavior through the Lens of Mobile Phones during Floods

A. J. Morales[1], D. Pastor-Escuredo[1], Y. Torres[1], V. Frías-Martínez[2], E. Frías-Martínez[3], N. Oliver[3], A. Rutherford[4], T. Logar[4], R. Clausen-Nielsen[4], O. De Backer[4], M. A. Luengo-Oroz[4]
[1]Universidad Politécnica de Madrid, [2]University of Maryland
[3]Telefónica Research,[4]United Nations Global Pulse

Natural disasters affect hundreds of millions of people worldwide every year. Emergency response efforts depend upon the availability of timely information, such as information concerning the movements of affected populations. The analysis of Call Detail Records (CDR) captured from the mobile phone infrastructure provides new possibilities to characterize human behavior during critical events. In this study, we combine remotely sensed data and CDRs to understand how people communicated during severe floodings in the Mexican state of Tabasco in 2009. This research demonstrates that CDR data has the potential to provide useful information on human behavior for improved emergency management and humanitarian response. Our results could also serve as a potential proxy indicator for flood impact and risk awareness.

The lack of timely, accurate information about movements and communications of affected populations during natural disasters can limit the effectiveness of humanitarian response. However, the growing ubiquity of mobile phones has revealed new opportunities for accessing such information. Mobile phone data can provide valuable insights, in order to tackle issues related to economic and humanitarian development [2], such as understanding the behavior of affected populations during a natural disaster [1]. For example, recent research has demonstrated the potential of mobile phone data to help study population movements after an earthquake in Haiti [3] or to model malaria outbreaks in Kenya [4]. During these critical events, the patterns of collective human behavior are disrupted, as the population faces the ongoing emergency. Such effect is closely related to the emergence of large information cascades, since people tend to communicate with others, triggering chain reactions in the social network [5]. The geographical distribution of the activity can also be used to characterize the catastrophe. For instance, Twitter activity allows to locate an earthquake's epicenter with extraordinary accuracy by geographically measuring the volume of related tweets [6].

The goal of this study is to develop and apply methods to assess the suitability of CDR data to characterize the impact of floods on the population. Our vision is to build CDR-based decision-support tools to help the public sector better respond to floods and other natural disasters. We investigate the viability of using CDR data combined with other sources of information to characterize the floods that occurred in Tabasco, Mexico in 2009. In particular, we analyzed CDRs of the geographical area affected by the floods during a period of nine months (July 2009 to March 2010). The main technical contribution of this work is the development of a multimodal data integration framework that facilitates the combination of CDR data with data from other sources, in order to characterize changes in the communication patterns during the floods. We also contrast our results with external ground truth information. For a longer description of the research presented in this short paper please refer to [1].

The methodological framework proposed in this study is composed of the following steps: First, we evaluate the representativeness of the data by using the 2010 census [10] of Tabasco as the ground truth. For this purpose, we compare the census data with a data-driven social baseline that we built based on the location of the home antenna tower (HAT) for each phone, meaning the antenna tower most used at night during the baseline (BL) period [7]. Second, we integrate additional and diverse data sources to further understand the phenomena. We use remote sensing through medium resolution (15 to 60 meters) ETM+ Landsat7 [8] satellite images to detect and geographically confine the submerged land. Moreover we analyze the Tropical Rainfall Measuring Mission [9] data, in order to build a temporal series of precipitations and to understand the relationship between the natural phenomenon and mobile phone activity.

In order to detect abnormalities in the activity, we examine mobile phone activity data before, during and after the disaster. We propose the variation metric that relies on the comparison of the number of phones placing or receiving calls per antenna $x(t)$, against their characteristic variation obtained during the baseline period (BL). Mathematically, the variation metric is defined as the z-score from $x(t)$ –*i.e.* the normal distribution characterizing the baseline– and defined as $x_{norm} = (x(t) - \mu_{BL})/\sigma_{BL}$, where the pair (μ_{BL}, σ_{BL}) statistically characterizes the activity during the baseline period. By means of analyzing the normalized series, we establish a baseline understanding of emergency behavior which enables us to measure the rate of disaster recovery and to show how affected populations behave in response to flooding. The variations in the number of active phones connected to each cell tower reveal abnormal activity patterns in the most affected locations during and after the floods that could be used as signatures of the floods – both in terms of infrastructure impact assessment and population information awareness.

This research demonstrates that mobile phone data has the potential to provide timely information about human

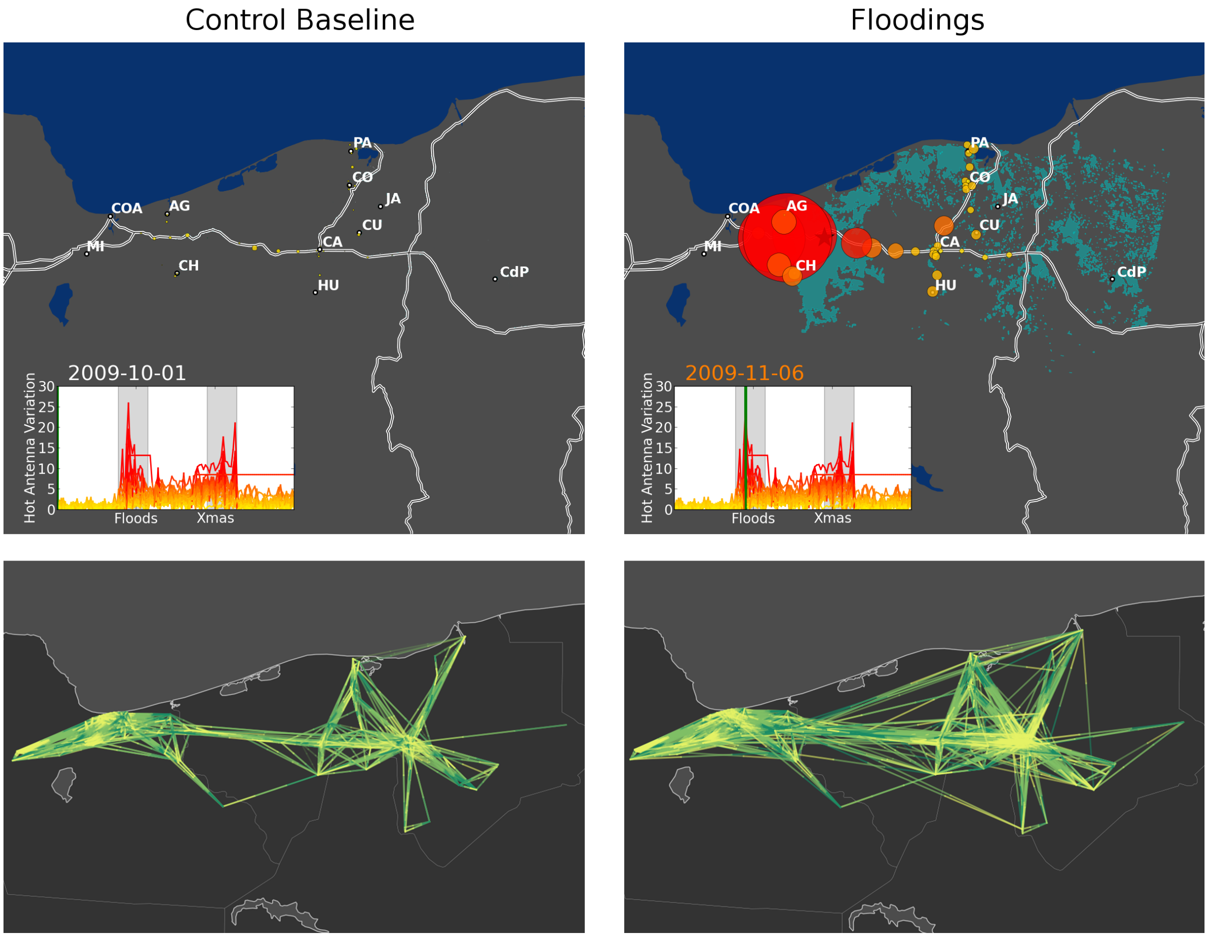

FIG. 1. Tabasco impact maps. The top panels show the absolute value of the antenna variation metric during an arbitrary day before floods (top left) and another one during floods (top right). In these panels, each antenna is represented by a circle with color and size proportional to the daily variation. The segmented flooded area has been colored in light blue. The insets display the temporal series of the antenna variation metric, and the green line indicates the day of observation. In the top right panel, antennas near the flooding area dramatically increased their variation during the floods. In the bottom panels, we show two visualizations of human displacement networks among cell towers. Towers are connected if a person makes two consecutive calls. We show an average week of the network before the floods (bottom left), and an aggregated network across the floods (bottom right). The edge color means the direction of the displacement, from green to yellow. It can be noticed that during floods (bottom right) the graph is denser, and more connections are established between the towers.

behavior for improved emergency management and humanitarian response. Insights gained from CDR analysis could also serve as a potential proxy indicator for flood impact and risk awareness. On one hand, mobile phone data can be highly representative of the population's behavior. A comparison between CDR data and census data yields a strong linear relation between official population statistics and population estimates computed from CDR data. Furthermore, civil protection warnings are not necessarily an effective way to raise awareness. In fact, a civil protection warning was issued on the day of highest rainfall in Tabasco in 2009. However, big spikes in phone activity were only observed in two cell phone towers along the most affected road when floods already showed initial impacts, meaning that the civil protection warning did not generate similar levels of awareness. This finding reveals important behavioral insights for emergency responders on how and when affected populations are made aware of a disaster. Finally, mobile activity can provide signals of flooding impact. When analyzed against the baseline activity, cell phone towers with the highest variation metric in the number of calls made during the floods were located in the most affected locations (see top panels in Fig. 1). Note that mobility patterns also changed significantly near the main cities and the capital as the ground transport system and the physical size of the cities constrain how the people could move during the floods (see bottom panels in Fig. 1).

In summary, aggregated and anonymized mobile phone data can be used to assess risk awareness, understand the effect of public communications such as disaster alerts and measure the direct impact of floods on the population. The research findings described in this paper show that CDR data could be a beneficial source of information for both emergency management and resilience assessment. Analyzing mobile activity during floods could be used to potentially locate damaged areas, efficiently assess needs and allocate resources (for example, sending supplies to affected areas). Identifying cell phone towers in the most affected areas of flooding might also serve to improve and target public communications and safety alerts, as well as help measure the effectiveness of such early warning announcements.

[1] Pastor-Escuredo, D., Morales, A. J. et al., *Flooding through the Lens of Mobile Phone Activity.* IEEE Global Humanitarian Technology Conference, GHTC 2014.

[2] Decuyper, A. et al., *Estimating Food Consumption and Poverty Indices with Mobile Phone Data*, arXiv:1412.2595, (2014)

[3] Bengtsson, L. et al., PLoS Med **8** (2011), no. 8, e1001083. *Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: A post-earthquake geospatial study in haiti*,

[4] Wesolowski, A. et al., Noor, Robert W. Snow, and Caroline O. Buckee, *Quantifying the Impact of Human Mobility on Malaria*, Science **338** (2012), no. 6104, 267–270.

[5] Bagrow, J. P. et al., *Collective Response of Human Populations to Large-Scale Emergencies*, PLOS ONE **6** (2011), no. 3, e17680.

[6] Sakaki, T. et al. *Earthquake shakes twitter users: real-time event detection by social sensors*, Proceedings of the 19th international conference on World wide web (New York, NY, USA), WWW '10, ACM, 2010, pp. 851–860.

[7] Becker, R. et al., *Human mobility characterization from cellular network data*, Commun. ACM **56** (2013), no. 1, 74–82.

[8] http://earthexplorer.usgs.gov/

[9] http:// http://trmm.gsfc.nasa.gov/

[10] http://www.censo2010.org.mx/