

# *Harnessing cell phone traces to model the spread of Zika in Colombia*

Daniela Perrotta<sup>1</sup>, Enrique Frias-Martinez<sup>2</sup>, Miguel Luengo-Oroz<sup>3</sup>, Daniela Paolotti<sup>4</sup>, Michele Tizzoni<sup>4</sup>, Alessandro Vespignani<sup>4,5</sup>

(1) Max Planck Institute for Demographic Research (2) Telefonica Research (3) United Nations Global Pulse  
(4) ISI Foundation (5) Northeastern University

In today's highly-connected world, millions of people travel every day, creating routes and opportunities for the spread of infectious diseases worldwide. Population movements offer safe passage for disease vectors and animal pathogens, which can easily cross international borders and even reach new previously unaffected areas. Accurate data on human mobility is therefore of paramount importance. Epidemic models, if properly designed and informed with accurate data, allow to simulate the spatial propagation of infectious diseases and evaluate appropriate control measures and timely intervention strategies. Traditional data collected from census surveys is often inadequate due to lack of spatial and temporal resolution, or may be completely unavailable in developing countries. Novel sources of data and innovative techniques are used more and more to detect people's physical movements over time. In this context, mobile phone data, in the form of Call Detail Records (CDRs), implicitly brings a large ensemble of details on human activity, including human movements. Such digital traces left by mobile phone users' represent a relatively low-cost resource to draw a high-level picture of human mobility patterns at an unprecedented scale.

In this study, we focus on the human mobility patterns relevant to the epidemic spread of Zika during the outbreak that occurred in Colombia in 2015-2016. The aim of the study is to investigate the potential predictive power gained by integrating the human mobility derived from CDRs data into an epidemic modelling approach, as well as more traditional methods, such as census data and mobility models. Specifically, more than two billion encrypted and anonymized calls made by around seven million mobile phone users in Colombia have been used to reconstruct the aggregated mobility network describing the flows of people travelling daily among different locations. In addition to the mobile phone network, we apply the gravity model and the radiation model to synthetically infer the population movements in Colombia and generate the corresponding mobility networks.

From the networks' point of view and by considering the census data as a reference, our results show that the gravity model strongly underestimates the mobility flows, whereas the mobility determined by the radiation model and the CDRs data show comparable performances in terms of high correlations and good similarities metrics with census data, thus reproducing well the actual mobility between the 33 administrative units (i.e. departments) in Colombia.

To model the spread of Zika, we employ a metapopulation modelling approach to explicitly simulate the spatial spread of Zika as governed by the transmission dynamics of the virus through human-mosquito interactions and as promoted by population movements across the country. Notably, our approach integrates detailed data on the population, the spatial heterogeneity of the mosquito abundance and the consequent exposure of the population to the virus due to socio-economic factors in order to help provide a more realistic representation of the epidemic progression.

Given the same modelling settings (i.e. initial conditions and parameters), our approach allows to perform numerical simulations of the spatio-temporal progression of the disease by integrating one mobility network at a time to ultimately assess their predictive power by comparing the simulated epidemic profiles with the Zika cases officially reported by the National Institute of Health in Colombia.