FLOWMINDER.ORG

UN Mobile Data Training Workshop: Overview of mobile phone data research

Tracey Li, PhD
tracey.li@flowminder.org

Research topics

Migration – migration statistics, seasonal migration, internal displacement

Human mobility – general mobility behaviour, mobility in post-disaster scenarios

Poverty mapping – spatial distribution of poverty in a country

Epidemiology – link between human mobility and disease prevalence

Urban analytics – transportation, commuting patterns, ‘hotspot’ analysis

Privacy – risk of de-identification, anonymisation techniques

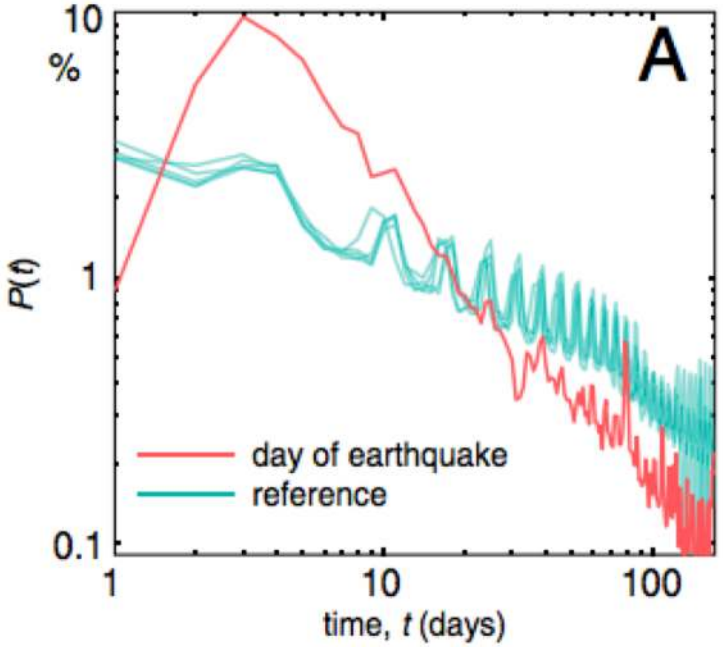
Understanding human mobility

Human movement during a disaster:

Differences between ‘normal’ movements and movement during a disaster are very easy to see using CDR data.

X. Lu, L. Bengtsson, P. Holme, PNAS, 2012

https://www.pnas.org/content/pnas/109/29/11576.full.pdf



Poverty mapping

Traditional methods to measure poverty rely on census data.

In many LMICs this data is unavailable or out-of-date.

Non-traditional datasets – mobile data and geospatial data– can be used instead.

Mobile data is most valuable in urban settings.

Mobile data features: basic phone usage, top-up patterns, social network metrics…

Poverty mapping

Poverty maps can be produced from a combination of CDRs and remote sensing data.

J. Steele et al, Royal Society Interface, 2017

https://royalsocietypublishing.org/doi/pdf/10.1098/rsif.2016.0690

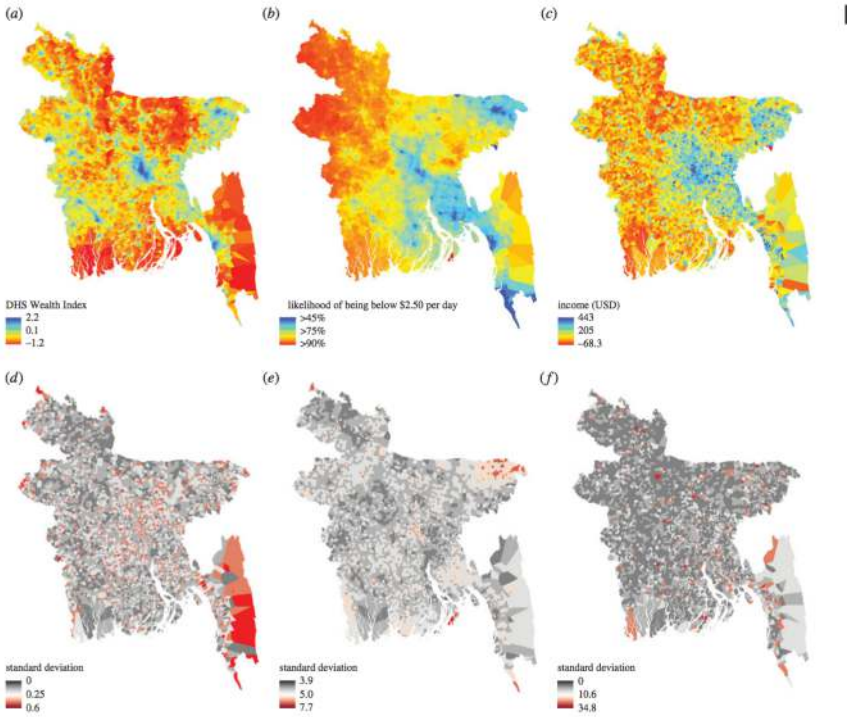


Figure 2. National level prediction maps for mean WI (*a*) with uncertainty (*d*); mean probability of households being below $2.50/day (*b*) with uncertainty (*e*); and mean USD income (*c*) with uncertainty (*f*). Maps were generated using call detail record features, remote sensing data and Bayesian geostatistical models. The maps show the posterior mean and standard deviation from CDR–RS models for the WI and income data (*a*,*c*), and the RS model for the PPI (*b*). Red indicates poorer areas in prediction maps, and higher error in uncertainty maps.

Epidemiology

Infectious disease models often use population models to predict the spread of disease e.g. cholera.

Models incorporating population movements derived from mobile phone data have been shown to produce better results than models using traditional population models.

Epidemiology

Movement maps produced from CDR can be used to predict the spread of diseases.

L. Bengtsson et al, Nature Scientific Reports, 2015,
https://www.nature.com/articles/srep08923

Figure 1: Mobile phone mobility network.

The average absolute number of mobile phones moving between the study areas (October 15 to December 19, 2010). Thicker, bluer lines indicate larger number of travelers. The original outbreak location (Mirebelais), the Artibonite River (dark blue) and Port-au-Prince (PAP) are depicted (visualisation using Gephi and ArcGIS).

Urban analytics

There are several use cases for studying behaviour in urban environments:

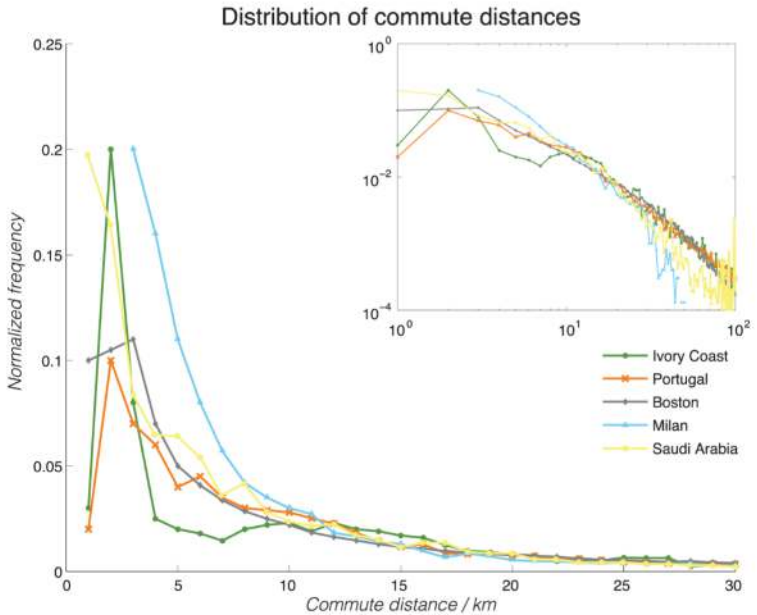
- Transportation usage
- Design and optimisation of transport networks
- Commuting patterns
- Urban planning
- …

Urban analytics

Mobile phone data can be used to compare cities in terms of mobility behaviours e.g. commuting.

K. Kung et al, PLOS ONE, 2014,

https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0096180



Privacy

How easy is it to re-identify an individual from ‘anonymous’ data?

Y. A. de Montjoye et al, Scientific Reports, 2013, https://www.nature.com/articles/srep01376?ial=1

”We study fifteen months of human mobility data for one and a half million individuals and find that human mobility traces are highly unique. In fact, **in a dataset where the location of an individual is specified hourly, and with a spatial resolution equal to that given by the carrier's antennas, four spatio-temporal points are enough to uniquely identify 95% of the individuals**. We coarsen the data spatially and temporally to find a formula for the uniqueness of human mobility traces given their resolution and the available outside information. This formula shows that the uniqueness of mobility traces decays approximately as the 1/10 power of their resolution. Hence, **even coarse datasets provide little anonymity**.”

Towards standardisation

Towards standardisation

Many people and organisations work with mobile phone data, for research and operational purposes.

Initiatives are beginning to work towards standardisation e.g. the UN Global Working Group Handbook on the use of Mobile Phone Data for Official Statistics.

What should be standardised?

Here are some things that could be standardised:

- Language and terminology
- Commonly-used analysis methods
- Methods for estimating uncertainty of commonly-used metrics
- Measures of data quality
- Quality assurance tests of the raw data
- Measures of effectiveness of anonymisation techniques
- Test data sets for simulating specific scenarios
- Training materials for mobile phone data analysis

What should be standardised?

Which items do you think it is most important to standardise?

Why?