Cell phone data and census microdata to model human movement and migration

A. Sorichetta et al.
Modeling Human Mobility in Space and Time

“Traditional”

- Census data
- Cross-border and traffic surveys
- Household travel history surveys

Tatem, 2014 (International Health)
Cellphone Call Data Records (CDRs)

User makes a call from location X

X

User travels to Y and makes a call

Y

Call routed through nearest tower

Network operator records time and tower of call for billing
Cellphone Call Data Records (CDRs)
More Accurate and Dynamic Assessments of Population Distributions

The combination of phone network data and satellite-based covariates produces accurate and dynamic population maps that are relatively insensitive to network usage biases.

Deville et al., 2014 (PNAS)
Cell Phone Ownership

Cell Phone Ownership Surges in Africa

Adults who own a cell phone

- U.S. 89
- S. Africa 89
- Ghana 83
- Kenya 82
- Tanzania 73
- Uganda 65


PEW RESEARCH CENTER

In 2011, there are 32 non-smart phones for every 1 smart phone

All phones in 2011:
500 000 000

Smart phones:
15 000 000

Non-smart phones:
485 000 000

In 2015, we expect Africa to have:

- 722 500 000 non-smart phones
- 127 500 000 smart phones

In 2015, there will be 5.6 non-smart phones for every 1 smart phone

2011
500 000 000

2015
850 000 000

3%
97%
15%
85%
Mobile Phone Data Access

PARTNERS

- Millicom
- Telenor Group
- Safaricom
- Airtel
- Ncell
- Digicel Haiti
- Ericsson
- Mic
- Orange
- Zantel
- Vodafone
Preserving Confidentiality!

Call Detail Records’ (CDRs) including low-resolution location data (nearest tower location) anonymized on separate server hosted by operator. Analyses are conducted under operator supervision always behind operator firewall.

Mobile operator firewall

Aggregated mobility estimates are exported and made open access for being potentially used with other mobility estimates and epidemiological data.

Compliance with GSMA data integrity guidelines: Data never leaves mobile operator’s system to avoid any privacy and/or commercial concerns.
Measuring migration
Seasonal Population Mapping

Erbach-Schoenberg et al., 2016 (Population Health Metrics)

Namibia Pop: 2.3 mill
MTC active subscriptions: 2.1 mill
Dynamic facility catchment populations

Erbach-Schoenberg et al., 2016 (Population Health Metrics)
Namibia closer to elimination than previously assumed?

% change in health district incidence through the year after accounting for dynamic catchment populations

Areas in red may have lower incidence than currently assumed using static catchment denominators

Erbach-Schoenberg et al., 2016 (Population Health Metrics)
Understanding short term mobility

Above normal inflow to each district
(negative numbers indicate less incoming people than normal)

Wilson et al., 2016 (PLoSCurr)
Very large flows from Kathmandu to other districts immediately after the earthquake...

Nepal Population Estimates as of 10th June 2015

2. Kathmandu Valley

Kathmandu Valley is here defined as the districts Kathmandu, Bhaktapur and Lalitpur. Kathmandu Valley is one of the most densely populated areas in Nepal and home to ca 2.8 m people [1].

Key findings:

- An estimated 390,000 people more than normal had left the Kathmandu valley—comparing May 1 with the day before the earthquake April 24 (ratio to the population: 14%).
- An estimated 247,000 persons less than normal had come into the area during the same period (ratio to the population: 8.8%).
- People leaving Kathmandu Valley went to a large number of areas, notably the populous areas in the south and the Central and West Development Regions.

Wilson et al., 2016 (PLoSCurr)
Population Displacements

Estimated population away from their home Section Communalement:

<table>
<thead>
<tr>
<th>HOME DEPARTMENT</th>
<th>GRANDE ANSE</th>
<th>SUD</th>
<th>NIPPESE</th>
</tr>
</thead>
<tbody>
<tr>
<td>POPULATION AWAY FROM HOME</td>
<td>77500</td>
<td>132000</td>
<td>51000</td>
</tr>
<tr>
<td>% AWAY FROM HOME</td>
<td>18%</td>
<td>17%</td>
<td>15%</td>
</tr>
</tbody>
</table>

24 October 2016, location of people away from their home Section Communalement (out of those living pre-hurricane in Grande Anse, Sud and Nippe only)[3]

[2] Of the people normally resident within the given Département, we estimate the total number away from their home Section Communalement on the given day.
[3] Section Communalement are left blank where insufficient data is available.
CDRs Pros and Cons

**Advantages**
- Massive sample size, impossible to achieve with travel history surveys
- National-scale data
- Long time series
- Relatively reliable source of origins/destinations and lengths of stay
- Provides information on social networks, wealth too

**Disadvantages**
- Bias in representation of national population movements
- Coverage gaps in the most rural areas
- No demographic information
- Cross-border measurements feasible but not easy
- Difficulties in sharing and accessing (mostly due to commercial and privacy concerns)

*Alternative datasets are required in order to quantify and map mobility across continental scales*
A comparison of the ranked estimates of movement (I)

Wesolowski et al., 2013 (PLoS ONE)
A comparison of the ranked estimates of movement (II)

Ruktanonchai et al., 2016
(Malaria Journal)
Modelling Internal Migration Using IPUMSI Census Microdata
Modelling Internal Migration in Africa

Henry et al., 2013 (Appl. Geogr.); Garcia et al., 2014 (Migration Studies)
Response Variable and Covariates

In order to consistently model internal migration across all countries only globally available datasets proving to be able to explain most of the variance in the gravity models of Garcia et al. were explored.

http://www.worldpop.org.uk/data/methods/
Modelling Framework

$MIG_{ij} = \frac{P_i^\alpha P_j^\beta}{d_{ij}^\gamma}$

With $\alpha$, $\beta$, and $\gamma$ being parameters, used to indicate the magnitude of the effect for each covariate, that are typically estimated in the statistical modelling framework.

$p_{ij} = \frac{e^{\beta_0+\beta_1P_i+\beta_2P_j-\beta_3d_{ij}}}{1+e^{\beta_0+\beta_1P_i+\beta_2P_j-\beta_3d_{ij}}}$

where $p_{ij} = MIG_{ij}/TOT_j$; with $MIG_{ij}$ and $TOT_j$ representing the number of people residing in $j$ in the census year that were in $i$ 5 years prior to the census and the total population residing in $j$ in the census year, respectively.
Models Common Across all Countries

Multi-step approach to identify the model with the greatest predictive power

Best model was then selected using a leave-one-out cross-validation approach

R² values for all withheld countries were averaged and used to rank each models according to their predictive power averaged across all withheld countries.

Sorichetta et al., 2016 (Nature Scientific Data)
Internal Migration Flows in Africa

Sorichetta et al., 2016
(Nature Scientific Data)
Internal Migration Flows in Asia

Sorichetta et al., 2016 (Nature Scientific Data)
Internal Migration Flows in LAC

Sorichetta et al., 2016
(Nature Scientific Data)
### Validation/Uncertainty

<table>
<thead>
<tr>
<th>Continent</th>
<th>ISO code</th>
<th>$R^2$</th>
<th>Error p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFRICA</td>
<td>CMR</td>
<td>0.60</td>
<td>0.07</td>
</tr>
<tr>
<td>AFRICA</td>
<td>EGY</td>
<td>0.21</td>
<td>0.20</td>
</tr>
<tr>
<td>AFRICA</td>
<td>GHA</td>
<td>0.68</td>
<td>0.21</td>
</tr>
<tr>
<td>AFRICA</td>
<td>GIN</td>
<td>0.39</td>
<td>0.09</td>
</tr>
<tr>
<td>AFRICA</td>
<td>MAR</td>
<td>0.52</td>
<td>0.14</td>
</tr>
<tr>
<td>AFRICA</td>
<td>MLI</td>
<td>0.51</td>
<td>0.14</td>
</tr>
<tr>
<td>AFRICA</td>
<td>MWI</td>
<td>0.02</td>
<td>0.06</td>
</tr>
<tr>
<td>AFRICA</td>
<td>SEN</td>
<td>0.54</td>
<td>0.12</td>
</tr>
<tr>
<td>AFRICA</td>
<td>UGA</td>
<td>0.50</td>
<td>0.11</td>
</tr>
<tr>
<td>AFRICA</td>
<td>ZAF</td>
<td>0.49</td>
<td>0.23</td>
</tr>
<tr>
<td>AFRICA</td>
<td>ZMB</td>
<td>0.37</td>
<td>0.22</td>
</tr>
<tr>
<td>ASIA</td>
<td>ARM</td>
<td>0.11</td>
<td>0.16</td>
</tr>
<tr>
<td>ASIA</td>
<td>CHN</td>
<td>0.08</td>
<td>0.19</td>
</tr>
<tr>
<td>ASIA</td>
<td>FJI</td>
<td>0.16</td>
<td>0.28</td>
</tr>
<tr>
<td>ASIA</td>
<td>KGZ</td>
<td>0.23</td>
<td>0.08</td>
</tr>
<tr>
<td>ASIA</td>
<td>IND</td>
<td>0.11</td>
<td>0.15</td>
</tr>
</tbody>
</table>

Error p-value is here defined as the average probability that predicted migration values do not belong to the observed migration dataset.

Sorichetta et al., 2016 (Nature Scientific Data)
Limitations and Caveats (I)

• For consistency, internal migration flows were estimated using a fixed set of pull and push factors common to all countries;

• Use of census data from many years ago for some countries may have generated inaccurate estimates for the period considered in this study (i.e., 2005-2010).

• The model fit varied between countries and could be improved by considering additional locally-specific migration drivers;
Limitations and Caveats (II)

- Migration models were fitted using only a small sample (ranging between 0.07% and 10%) of the full census for each country;
- The spatial detail at which migration is captured and summarized varies by country;
- The role of some of the pull and push factors, may not have been captured at the spatial level at which they influence internal migration as recorded in the census;
Limitations and Caveats (III)

• Ancillary datasets used to represent pull and push factors are modelling outputs in themselves having a degree of uncertainty that will carry over into the migration estimates;

• Other types of migrations, such as seasonal movements and forced displacements, may be not captured by the model.
Next Step

• Modelling international migration among subnational administrative units in Africa, Asia, and LAC as a function of distance using an (Iterative Proportional Fitting) double-constrain multilevel spatial interaction modelling framework as described in Dennett & Wilson, 2013 (Environment and Planning A);

• Using IPUMSI-based estimates for internal migration;

• Using Abel & Sander, 2014 (Science) for international migration between countries.

Abel, Sorichetta et al., in preparation
Acknowledgements
For Further Information

www.worldpop.org
@WorldPopProject

FLOWMINDER.ORG
www.flowminder.org
@Flowminder