MAPPING MULTI-DIMENSIONAL POVERTY BY COMBINING SATELLITE AND MOBILE PHONE DATA: CHALLENGES AND OPPORTUNITIES

EGM ON THE IMPLEMENTATION OF THE THIRD UN DECADE FOR THE ERADICATION OF POVERTY (2018-2027), UNECA, 10-12 MAY 2023

Neeti Pokhriyal, Ph.D.
American Association of the Advancement of Sciences (AAAS), Science and Technology Policy Fellow, National Science Foundation

Visiting Scholar, Department of Computer Science, Dartmouth College
Outline of the presentation

- Context
- Why combine satellite and mobile phone data?
- Case study on Senegal
  - Details of the data and methods
  - Results
- Lessons learnt and policy-significant challenges


-Virtual Networks and Poverty Analysis in Senegal, N. Pokhriyal et al. NetMob, MIT, 2015
Context

- Traditional ways to measure poverty
  - Costly; timely + high resolution updates – difficult
- *Rich census/surveys + alternate data*, like
  - mobile phone data
  - satellite and aerial imagery
  - weather stations
  - economic data
  - open street maps etc.
- **For**: Frequent, spatially finer and accurate prediction of poverty in data scarce situations
  - inter-censal times, conflicts/pandemics, policy evaluation etc.
**Data Ecosystem**

**Call Data Record** to capture how, when, where, and with whom the individual communicates.

**Basic phone usage** Call duration; Active days

**Regularity of calls** Inter-event time

**Diversity of contacts**

**Spatio-temporal variability** – number of antennas

---

**Food security** Temp.; Precipitation; Elevation, Soil

**Economic Activity** Nighttime lights; Land cover

**Accessibility to services** Proximity to urban centers, markets, main roads, schools/university, water tower, hospitals

---

**Training and Validating Machine Learning (ML) model**

Estimation of socio-economic deprivations

---


## Summary Statistics and Characteristics of the Data Used - CDRs, Environment, Census, MPI Poverty Index

<table>
<thead>
<tr>
<th>Summary Statistics</th>
<th>CDRs</th>
<th>Environment Data</th>
<th>Census</th>
<th>Poverty Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Timeline</td>
<td>Jan-Dec 2013</td>
<td>1960-2014</td>
<td>2013</td>
<td>2013</td>
</tr>
<tr>
<td>Total calls &amp; text</td>
<td>11 Billion</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Unique individuals</td>
<td>9.54 M</td>
<td>N/A</td>
<td>1.4 M</td>
<td>N/A</td>
</tr>
<tr>
<td>Spatial granularity</td>
<td>Antenna-level (1666)</td>
<td>vector data - 100 m -1 km</td>
<td>Household-level</td>
<td>Region-level (14)</td>
</tr>
<tr>
<td>Cost incurred in data collection &amp; preparation</td>
<td>Low/no cost (data exhaust)</td>
<td>Low/no cost (data exhaust)</td>
<td>USD 29 Million</td>
<td>Very high cost, and high human expertise</td>
</tr>
<tr>
<td>Frequency of update of data</td>
<td>Real-time</td>
<td>~1 year</td>
<td>10 years</td>
<td>3-5 years</td>
</tr>
</tbody>
</table>

**Table 1.** Summary statistics and characteristics of the data used - CDRs, environment, census, MPI poverty index.
**Objective:** Learn a relationship/mapping between inputs and outputs. Model helps to:
1) Predict output given an input.
2) Provides uncertainty with its poverty estimates – *measure of trust of our model.*

Results

Dots on the map: 121 urban centers. Rest are 431 rural communes.
Estimated Poverty Map

Predicted using our model

Estimated from the census
## Quantitative Results

<table>
<thead>
<tr>
<th>Poverty Indicators &amp; Dimensions</th>
<th>Multi-source Data</th>
<th>CDR</th>
<th>Satellite</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>rank corr.</td>
<td>RMSE</td>
<td>rank corr.</td>
</tr>
<tr>
<td>MPI</td>
<td>0.88 (0.06)</td>
<td>0.08 (0.01)</td>
<td>0.86 (0.07)</td>
</tr>
<tr>
<td>H</td>
<td>0.85 (0.08)</td>
<td>10.79 (3.96)</td>
<td>0.84 (0.08)</td>
</tr>
<tr>
<td>A</td>
<td>0.85 (0.07)</td>
<td>4.71 (0.96)</td>
<td>0.82 (0.08)</td>
</tr>
<tr>
<td>Education</td>
<td>0.84 (0.05)</td>
<td>11.84 (1.88)</td>
<td>0.81 (0.07)</td>
</tr>
<tr>
<td>Health</td>
<td>0.50 (0.16)</td>
<td>12.76 (2.12)</td>
<td>0.52 (0.12)</td>
</tr>
<tr>
<td>Standard of Living</td>
<td>0.75 (0.13)</td>
<td>14.82 (3.92)</td>
<td>0.74 (0.11)</td>
</tr>
</tbody>
</table>
Validation against ground truth (Census)
Discussion

- Need an aggregation mechanism to link the different datasets
  - Varying spatial granularity
  - Privacy concerns of mobile data
    - In this work, data remains private within its ecosystem

- Need to mitigate potential biases in data and outputs
  - Selection bias in mobile data; bias in satellite imagery
  - Better coverage of CDR data, data from more telecom providers, higher resolution satellite data might benefit the model

- Data Governance issues related to responsible data collection, data management, and data sharing need to be tackled and incentivized
Build public-private partnerships to discuss participatory mechanisms of responsible data sharing that can support providing accurate estimates of poverty while preventing the misuse of data or models.

To address inclusion, diverse digital datasets that are collected by the local governments in response to poverty eradication and intervention programs (via e-governance initiatives) should be explored in conjunction with satellite and mobile phone data.
Key policy recommendations - II

- Efforts that **scale** these methodologies to **other countries** and to **more intercensal** time periods
  - Build multilateral partnerships among the governments and statistical agencies of different countries
  - Sustained collaborations with researchers/academic institutions and participatory involvement with the local communities

- Need to develop **workforce and infrastructural, computing, and technical capacity** of the Statistical agencies for African countries
  - **Twin goals**: strengthen their capacities and enhance the traditional survey and census-based data collection to drive improved decision-making and facilitating recovery efforts from the crises.
Thank you

neeti.pokhriyal@gmail.com