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

- Combining disparate data sources for improved poverty prediction and mapping.
Neeti Pokhriyal and Jacques, Proc. of National Academy of Sciences, 2017

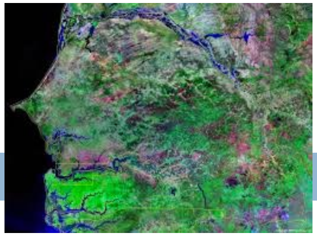
--Estimating and Forecasting Income Poverty and Inequality in Haiti Using Satellite Imagery and Mobile Phone Data, Neeti Pokhriyal, Omar Zambrano, Jennifer Linares, Hugo Hernandez, Working Paper, Inter-American Dev. Bank, 2020.

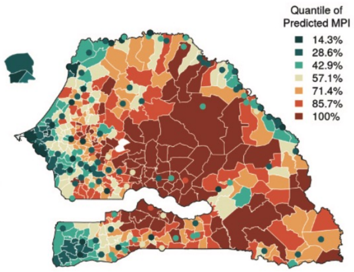
-Accurate Intercensal Estimates of Energy Access to Track Sustainable Development Goal 7, N. Pokhriyal, Emmanuel Letouź e, Soroush Vosoughi, EPJ Data Science, 2022.

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

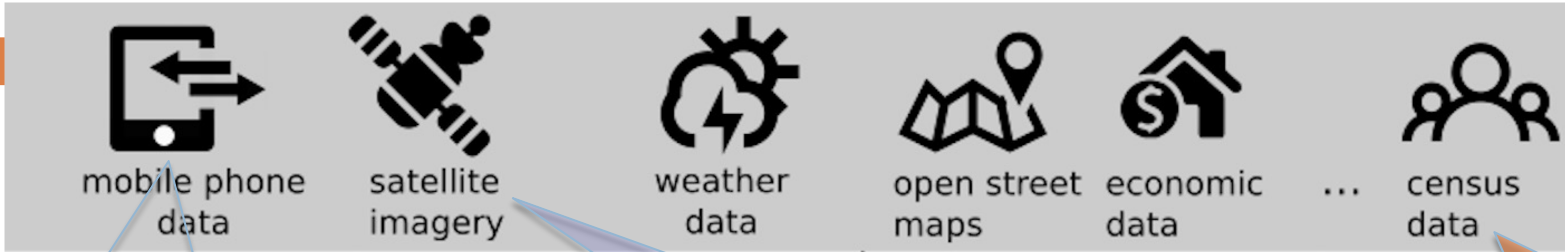
Context

- 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 Interevent 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

Blumenstock J, Cadamuro G, On R (2015) Predicting poverty and wealth from mobile phone metadata. *Science*
Eagle N, Macy M, Claxton R (2010) Network diversity and economic development. Science 328:1029–1031.

Disparate datasets used

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

Table 1. Summary statistics and characteristics of the data used - CDRs, environment, census, MPI poverty index.

Process overview

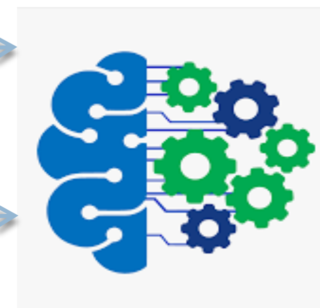
Feature
Extraction (converting raw
data into quantitative
features) and aggregation

Satellites and Geographic Information Systems and Mobile phone data

census/surveys

Input

Target



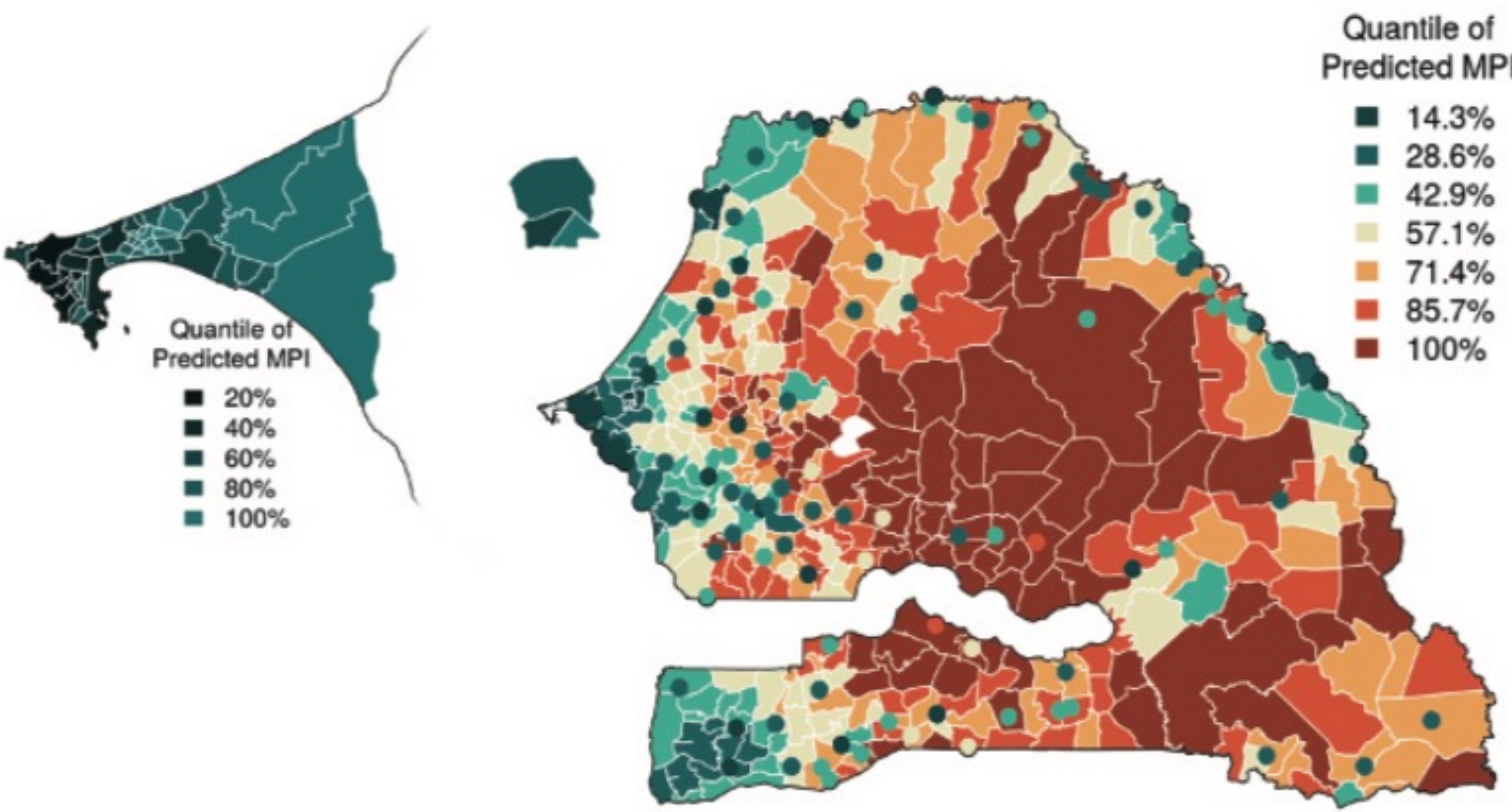
ML Model

<u>Objective:</u> 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.*

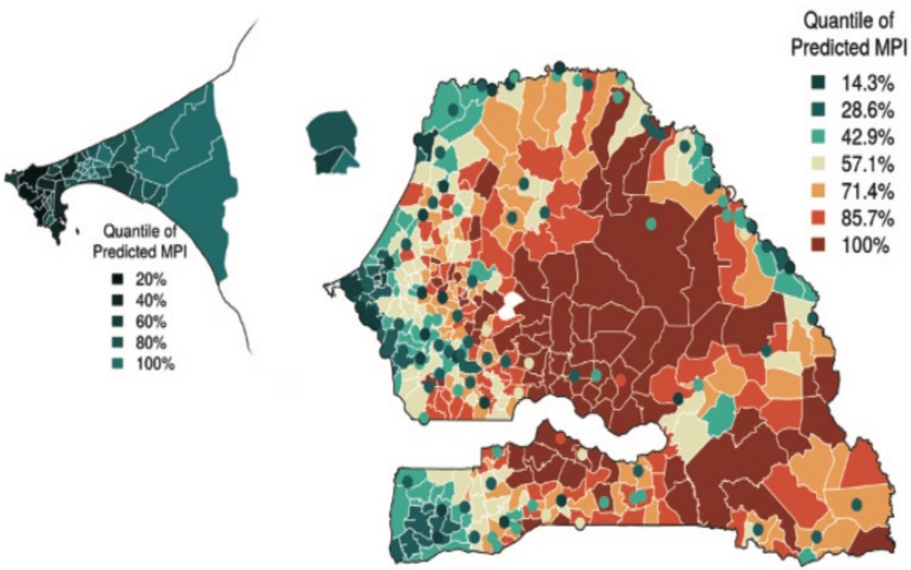
Rasmussen CE, Williams CKI (2006) Gaussian Processes for Machine Learning. (The MIT Press).

Results

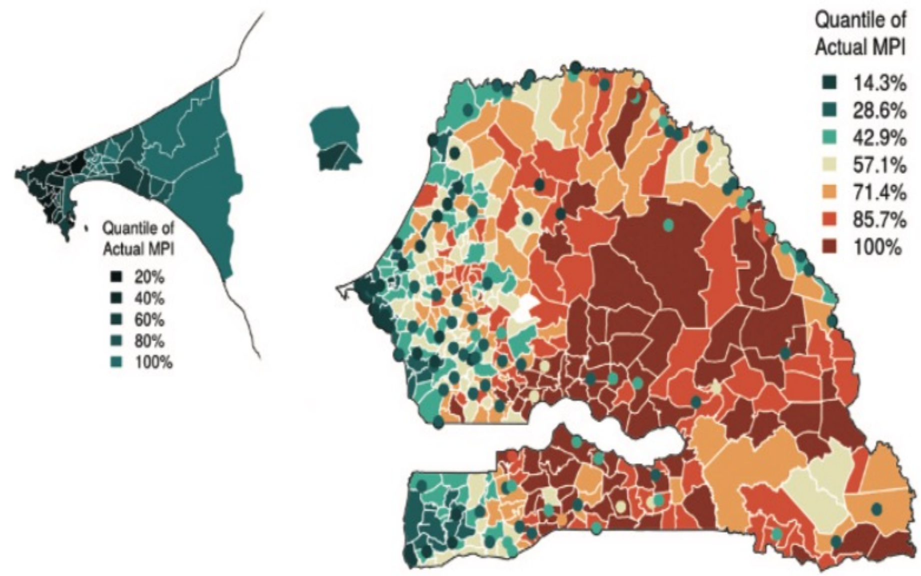


Dots on the map:121 urban centers. Rest are 431 rural communes

Estimated Poverty Map



Predicted using our model

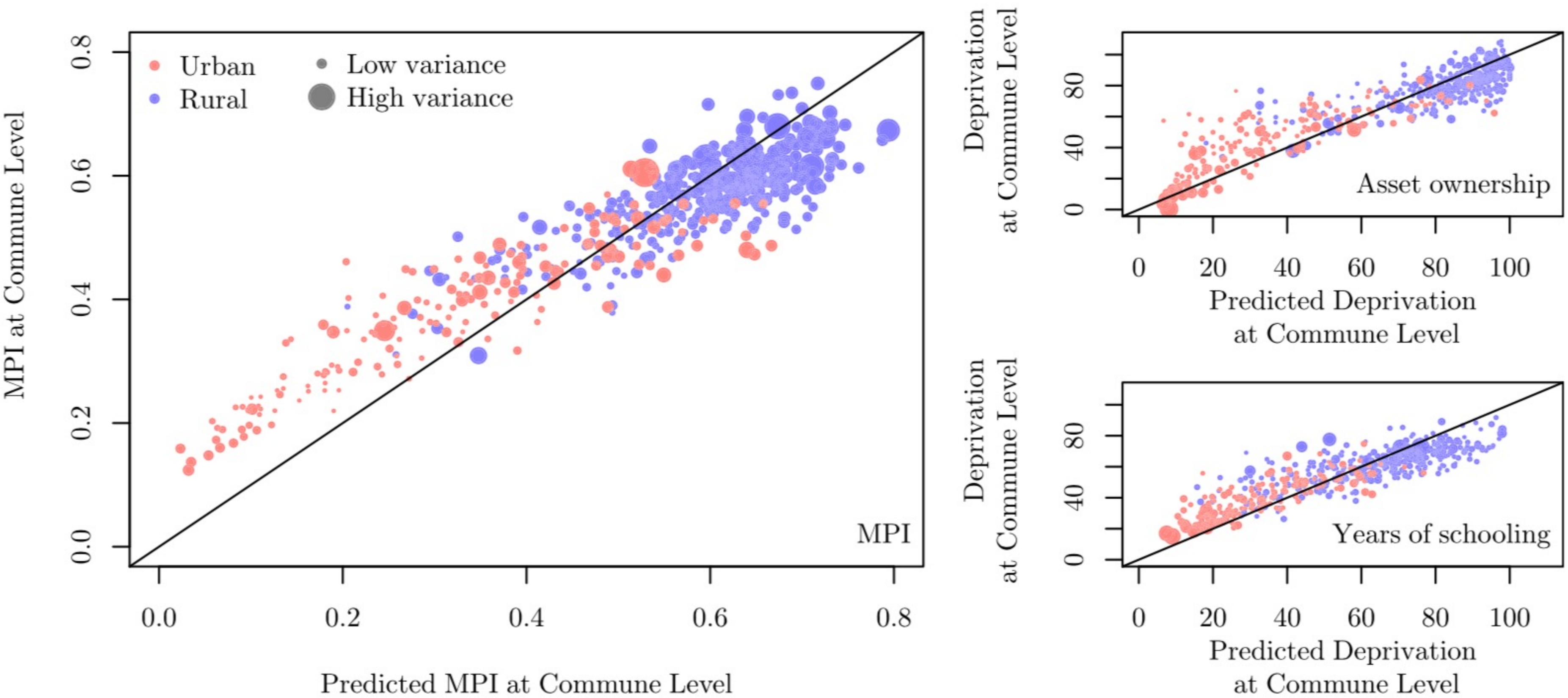


Estimated from the census

Quantitative Results

	Multi-source Data		CDR		Satellite	
Poverty Indicators & Dimensions	rank corr.	RMSE	rank corr.	RMSE	rank corr.	RMSE
MPI	0.88 (0.06)	0.08 (0.01)	0.86 (0.07)	0.08 (0.01)	0.80 (0.10)	0.10 (0.02)
H	0.85 (0.08)	10.79 (3.96)	0.84 (0.08)	10.76 (2.60)	0.75 (0.11)	13.65 (4.86)
A	0.85 (0.07)	4.71 0.96)	0.82 (0.08)	4.98 (1.14)	0.79 (0.08)	5.36 (0.75)
Education	0.84 (0.05)	11.84 (1.88)	0.81 (0.07)	13.08 (1.68)	0.74 (0.07)	14.98 (3.03)
Health	0.50 (0.16)	12.76 (2.12)	0.52 (0.12)	12.91 (1.92)	0.35 (0.23)	13.91 (2.32)
Standard of Living	0.75 (0.13)	14.82 (3.92)	0.74 (0.11)	15.24 (3.45)	0.64 (0.20)	17.88 (4.50)

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

Key policy recommendations - I

- 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