8th International Conference on BIG DATA & Data Science for Official Statistics
BILBAO 2024
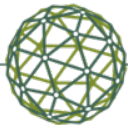
Informing Climate Change and Sustainable Development Policies with Integrated Data

BILBAO. SPAIN | 10-14 JUNE 2024 | #UNBigData2024

Correcting for biases in mobility indicators derived from Call Detail Records

Thomas Smallwood
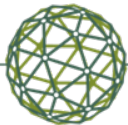Flowminder

The strengths and limitations of CDR data

Compared to survey and census data, MNO data have **several advantages**:

- **Automated** data collection by MNOs for billing purposes; no primary data collection required
- Once access to data is set up, **low maintenance costs**
- **Very high temporal granularity**
- **High spatial granularity**
- **Near-real time**, data become available within a few days

But also
limitations:

- Setting up partnerships can be a **lengthy process**
- **High set-up costs**
- **Lack of rigorous validation data**
- **How to quantify uncertainty?**
- **Selectivity and biases** of MNO data

The strengths and limitations of CDR data

Compared to survey and census data, MNO data have **several advantages**:

- Automated data collection by MNOs for billing purposes; no primary data collection required
- Once access to data is set up, low maintenance costs
- Very high temporal granularity
- High spatial granularity
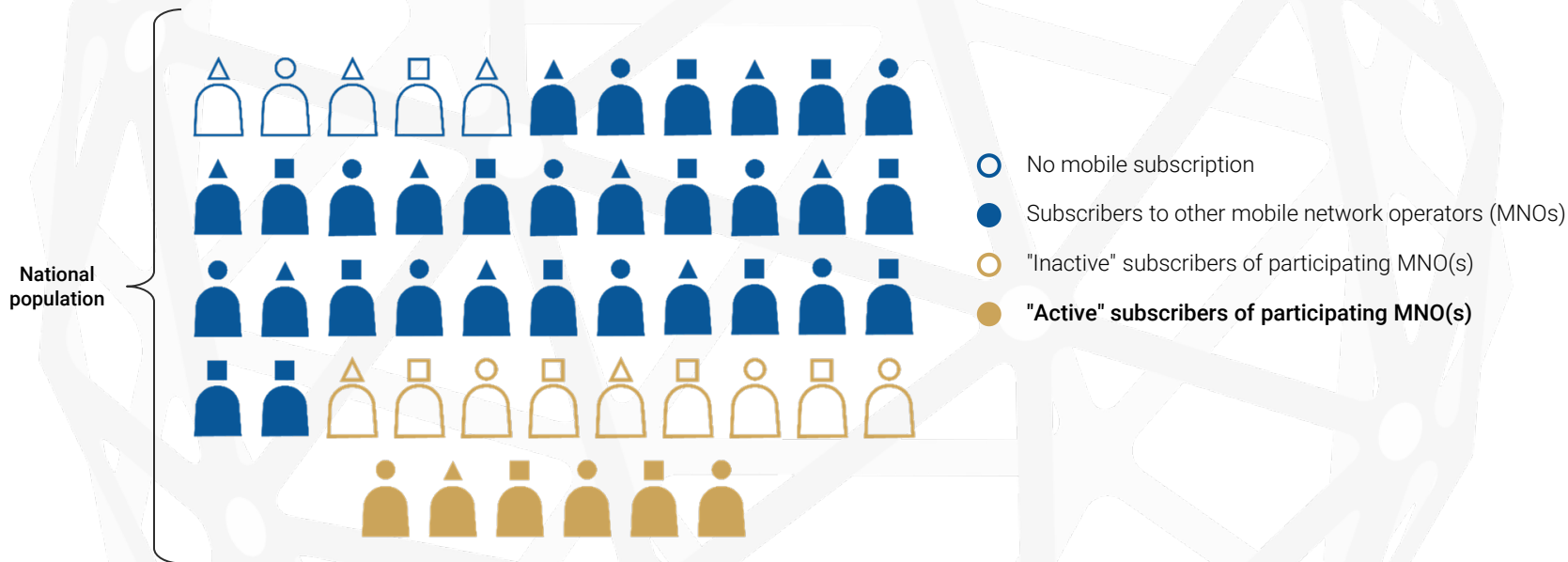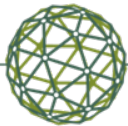- Near-real time, data become available within a few days

But also **limitations**:

- Setting up partnerships can be a lengthy process
- High set-up costs
- Lack of rigorous validation data
- **How to quantify uncertainty?**
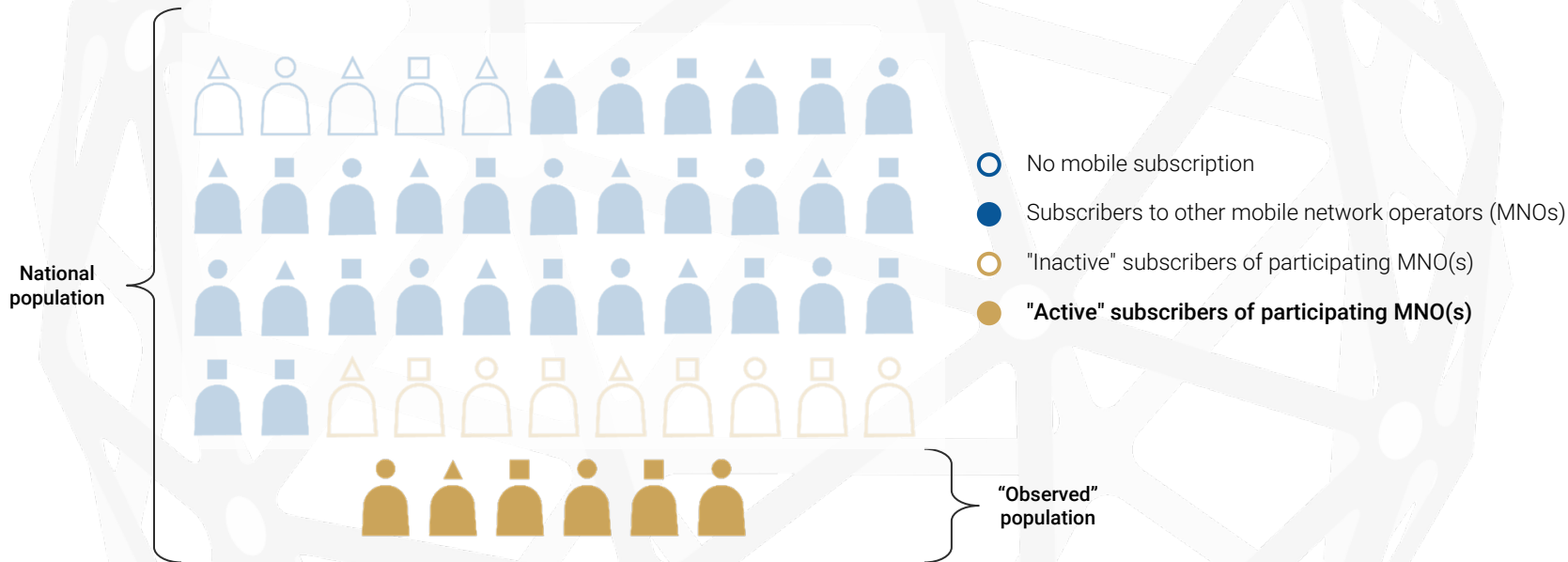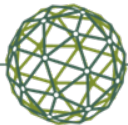- **Selectivity and biases** of MNO data

The subscribers of an MNO are not representative of the population as a whole, and cannot be assumed to be

National population

○ No mobile subscription

● Subscribers to other mobile network operators (MNOs)

○ "Inactive" subscribers of participating MNO(s)

● **"Active" subscribers of participating MNO(s)**

The subscribers of an MNO are not representative of the population as a whole, and cannot be assumed to be

National population

"Observed" population

○ No mobile subscription

● Subscribers to other mobile network operators (MNOs)

○ "Inactive" subscribers of participating MNO(s)

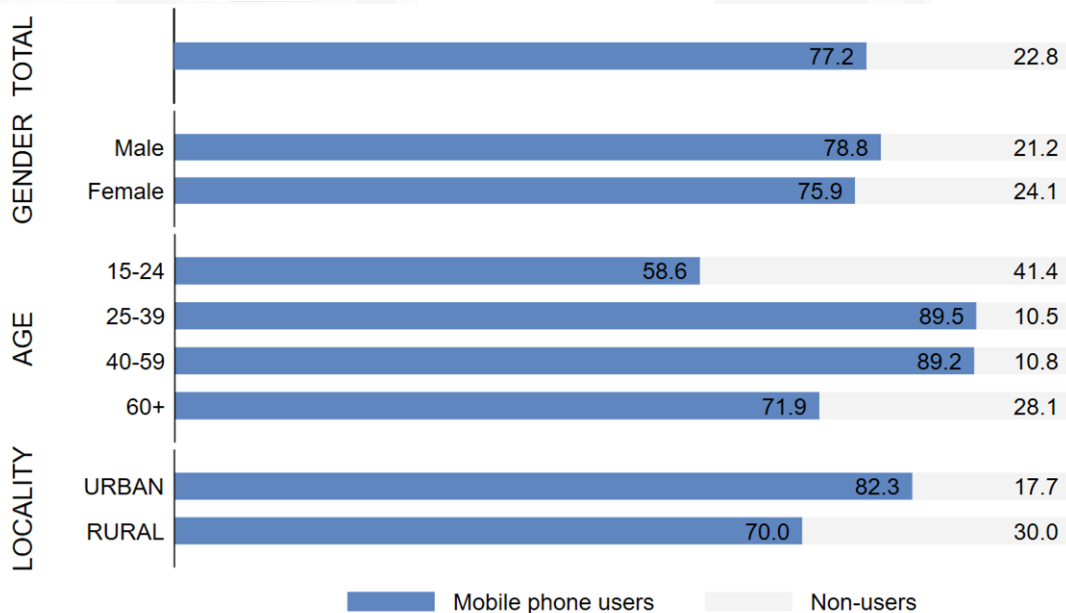● **"Active" subscribers of participating MNO(s)**
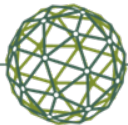
Mobile phone use by demographic | Ghana

Mobile phone use is:

- More common among **men** than among women

- More common among **persons of working-age** than among the elderly, and among children

- More widespread in **urban areas** than in rural areas

Percentage of mobile phone users (ages 15+) in Ghana
n = 31,284

		Mobile phone users	Non-users
GENDER	TOTAL	77.2	22.8
	Male	78.8	21.2
	Female	75.9	24.1
AGE	15-24	58.6	41.4
	25-39	89.5	10.5
	40-59	89.2	10.8
	60+	71.9	28.1
LOCALITY	URBAN	82.3	17.7
	RURAL	70.0	30.0

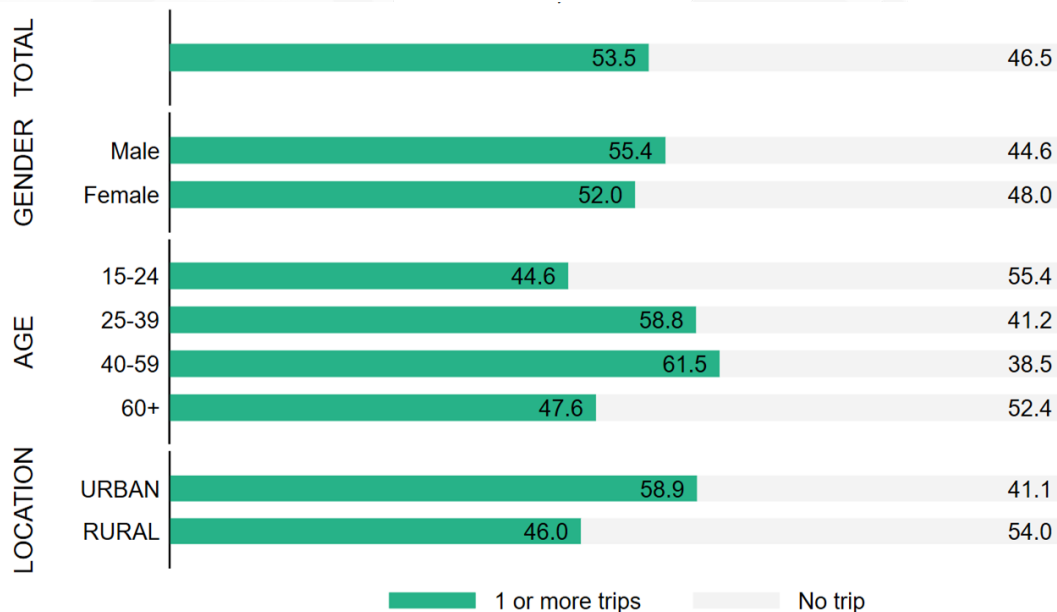Source: AHIES 2022, Q3, weighted(pop_weight)
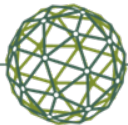
Mobility by demographic | Ghana

Mobility is:

- Higher among **men** than among women

- Higher among **persons of working-age** than among the elderly, and among children

- Higher in **urban areas** than in rural areas

Percentage of population (ages 15+) of Ghana w. Trips 5+km
n = 31,284

		1 or more trips	No trip
GENDER	TOTAL	53.5	46.5
	Male	55.4	44.6
	Female	52.0	48.0
AGE	15-24	44.6	55.4
	25-39	58.8	41.2
	40-59	61.5	38.5
	60+	47.6	52.4
LOCATION	URBAN	58.9	41.1
	RURAL	46.0	54.0

■ 1 or more trips   No trip

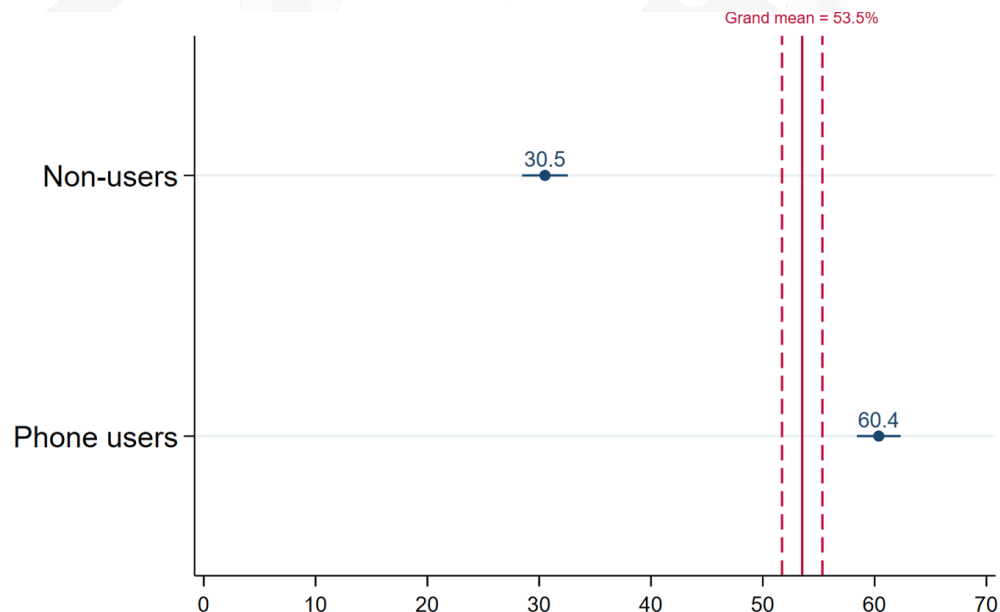Source: AHIES 2022, Q3, weighted(pop_weight)

Differences between mobile phone user and non-users create biases

As a consequence, mobility is often **different between mobile phone users and non-users**.

- On average, **mobility is higher among phone users** than among non-users
- Mobility estimates based on phone user data alone may **overestimate mobility**
- Not all mobility indicators show such large differences and some may show underestimations

Percentage of population (ages 15+) of Ghana w. trips 5+km
n = 30,928

Grand mean = 53.5%

Non-users — 30.5

Phone users — 60.4

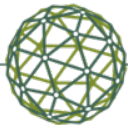Source: AHIES 2022, Q3, weighted (pop_weight)

The need for data triangulation (data fusion)

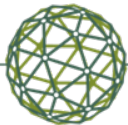CDR aggregates **alone** are therefore not sufficient for

- **Estimates on population counts or population density**

- **Estimates on population change** (births/deaths, immigration/emigration) or **population density change**

- **Extrapolation outside coverage areas** of CDR data

> **Survey (and census) data can and should be used!**

Addressing biases in CDR data

- MPD data is often **scaled from subscriber-level to population-level** using the **market share** represented in the data

- However, MPD data is **not a random sample of the population** and behaviours, including mobility, may **vary between groups**

- Flowminder has developed methods to **adjust for representation biases** and provide **population- scaled estimates** using:
  - **Demographic data** (e.g. census data)
  - **Phone usage** and **mobility** data from surveys

Calculating monthly mobility and population estimates

**Relocations and residents indicators**

Method for relocations estimates

- Relocations from area a to area b between month m and month n can be estimated from CDR aggregates of relocations (cdr_flow_{abmn}) between those areas and months, and from a **flow adjustment factor** and a **flow scaling factor**

- Flows are adjusted for the **number of SIMs per user** ($sims_{ab}$). The flow scaling factor is the inverse of the **share of MNO users** (mno_share_{ab}) in the flows:
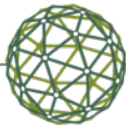
$$est\_flow_{abmn} = cdr\_flow_{abmn} * (1/sims_{ab}) * (1/mno\_share_{ab})$$

- Note: Parameters for the subset of mobile households/individuals only available at admin1 by admin1 level

Method for monthly residents estimates
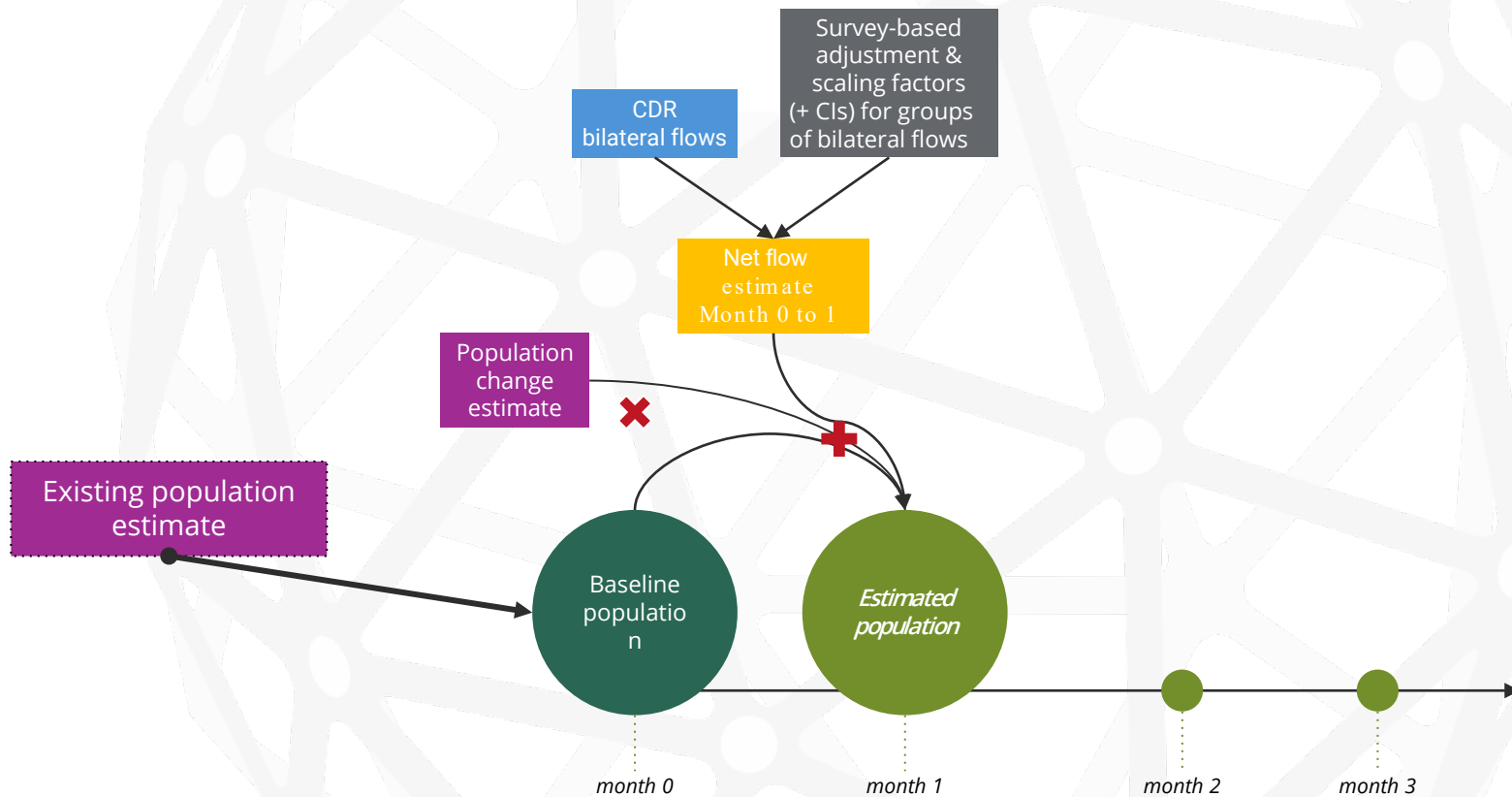
Data sources by country

Democratic Republic of the Congo

WorldPop 2020: gridded population estimates

Micro-census 2021: covering 7 provinces, phone users and non-users (led by FM)

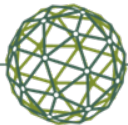**Phone survey 2021**: targeting phone users across the country from all MNOs (commissioned by FM)

Ghana

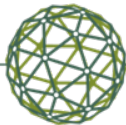**Census 2022**: phone users and non-users, population estimates (GSS)

Annual Household Income and Expenditures Survey (AHIES) 2022: phone users and non-users (GSS)

Phone survey 2022: targeting phone users across the country from all MNOs (commissioned by FM, conducted by GSS)
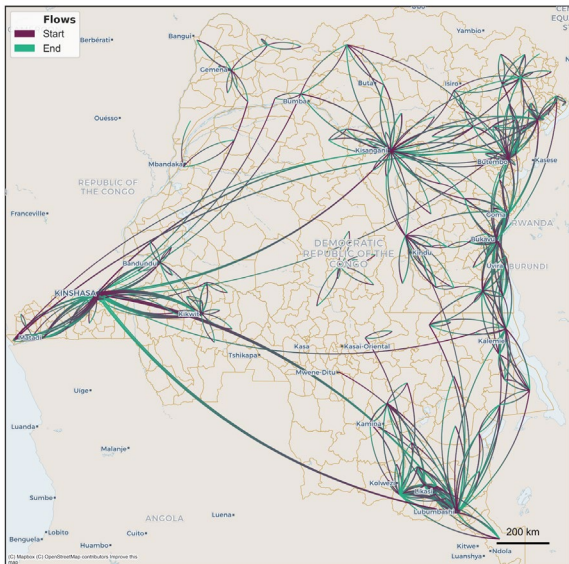
Calculating monthly mobility and population estimates
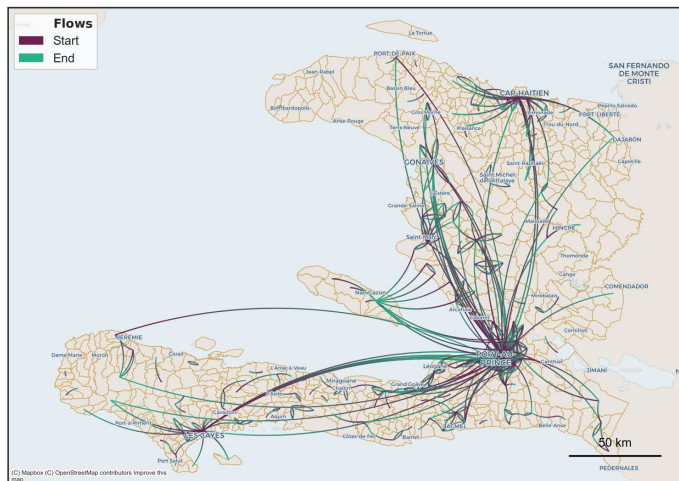
Examples of implementation

Monthly relocations between sub-regions

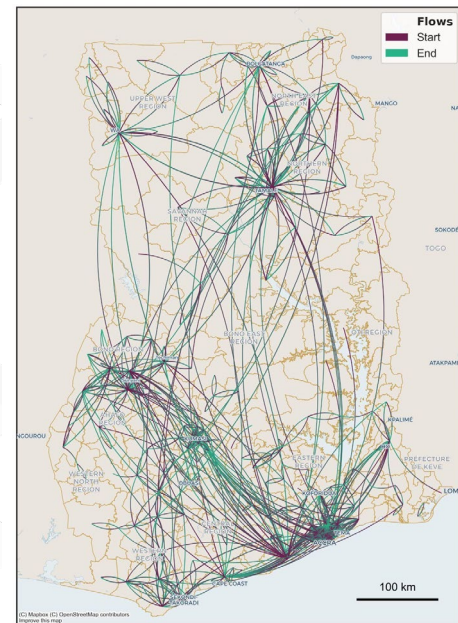## Democratic Republic of the Congo

Note: estimated top 1,000 flows between health zones, median, Nov 2021 - Dec 2022
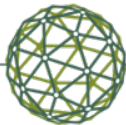
Haiti

Note: top 500 flows between communal sections, median, Feb 2020 - Feb 2022

Ghana

Note: unscaled top 1,000 flows between districts, median, Jan - July 2021

BILBAO 2024   #UNBigData2024

Next steps

Ensure regular updates to parameters (i.e. pop. change %, mobility, phone use) and **onboarding additional MNOs**

Refinement of adjustment and scaling factors

Testing estimations with **validation** data

Paper forthcoming!

Thank you very much!

For any questions, please contact:

thomas.smallwood[at]flowminder.org