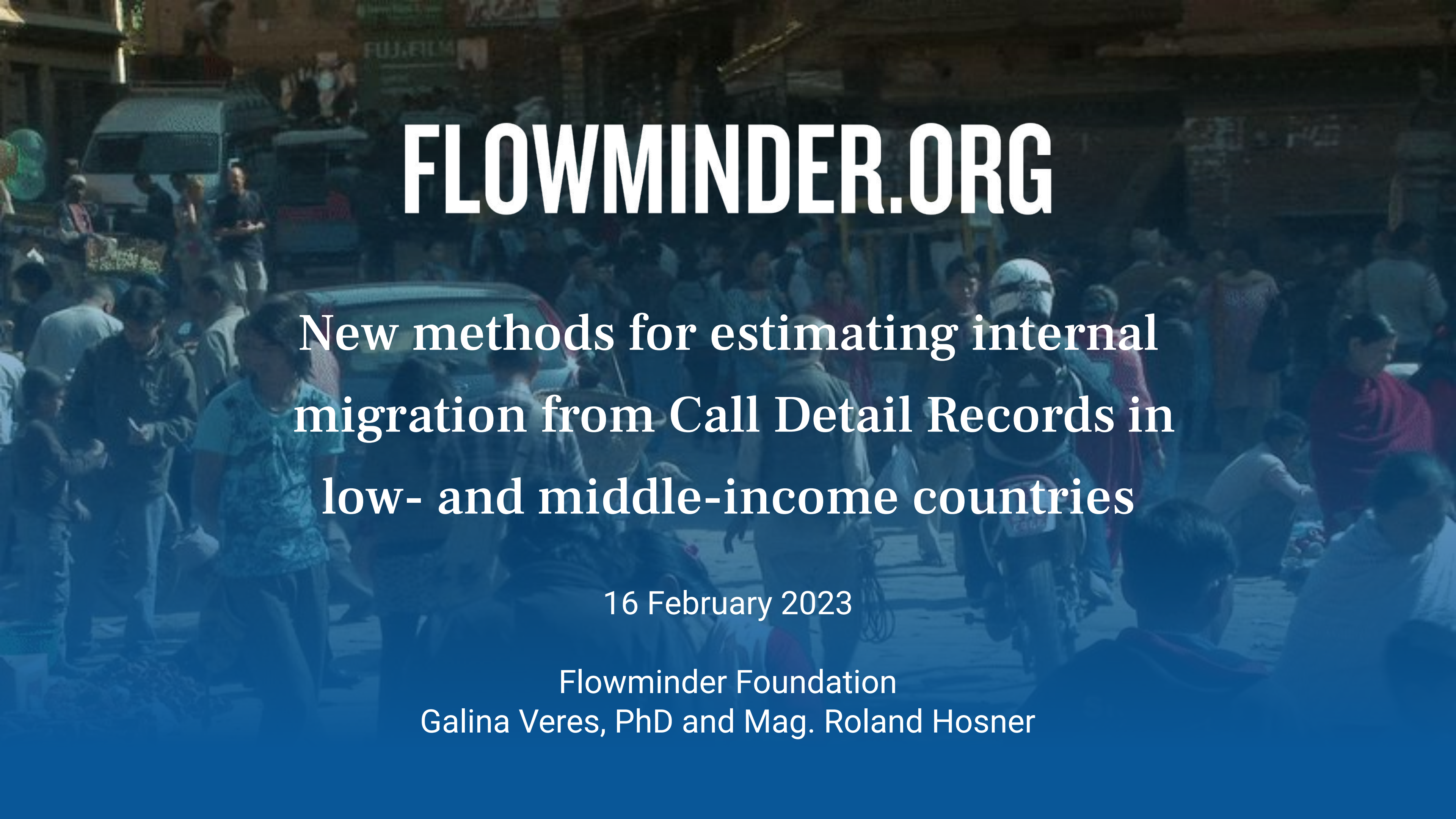


FLOWMINDER.ORG

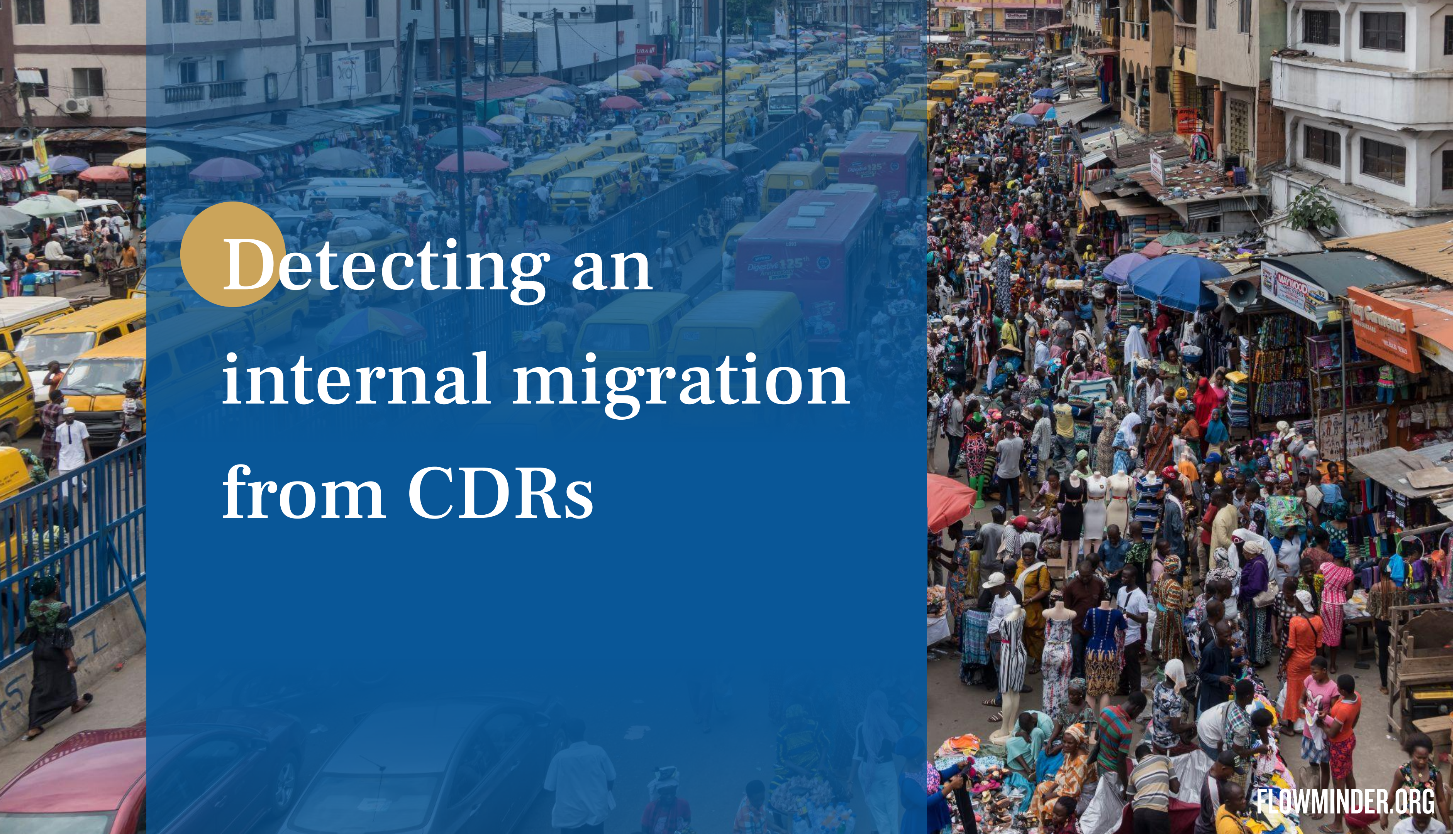
New methods for estimating internal migration from Call Detail Records in low- and middle-income countries

16 February 2023

Flowminder Foundation
Galina Veres, PhD and Mag. Roland Hosner

Outline

- Detecting internal migrations from CDRs
- Bias adjustment and scaling to the total population
- Producing monthly internal migration estimates for 3 countries



Detecting an internal migration from CDRs

FLOWMINDER.ORG

What is migration?

By migration, we understand a change of home location by a resident for at least one month

The spatial resolution of a home location is the sub-regional level, usually **administrative level 2 or 3** (depending on the country)

Challenges to detect internal migrations in low- and middle-income countries (LMICs)

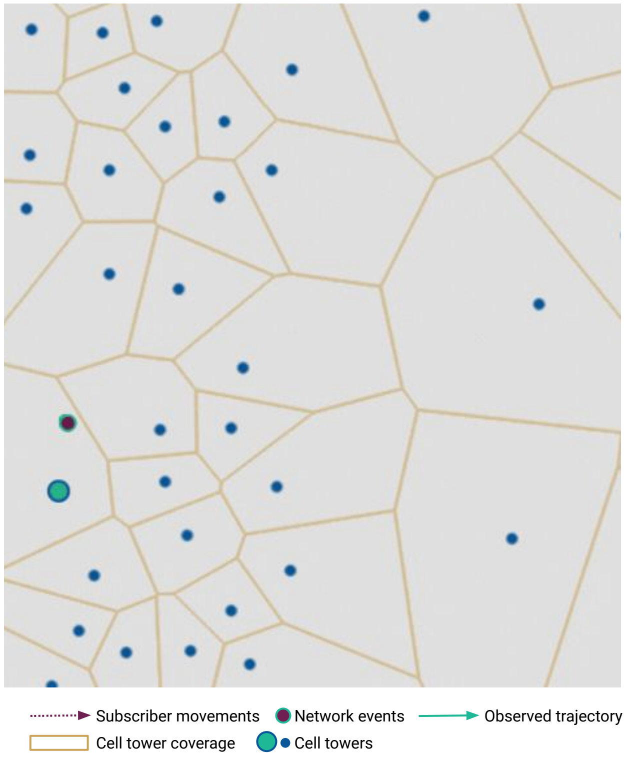
Traditionally, internal migration or residential mobility has been studied using **surveys** or **census**

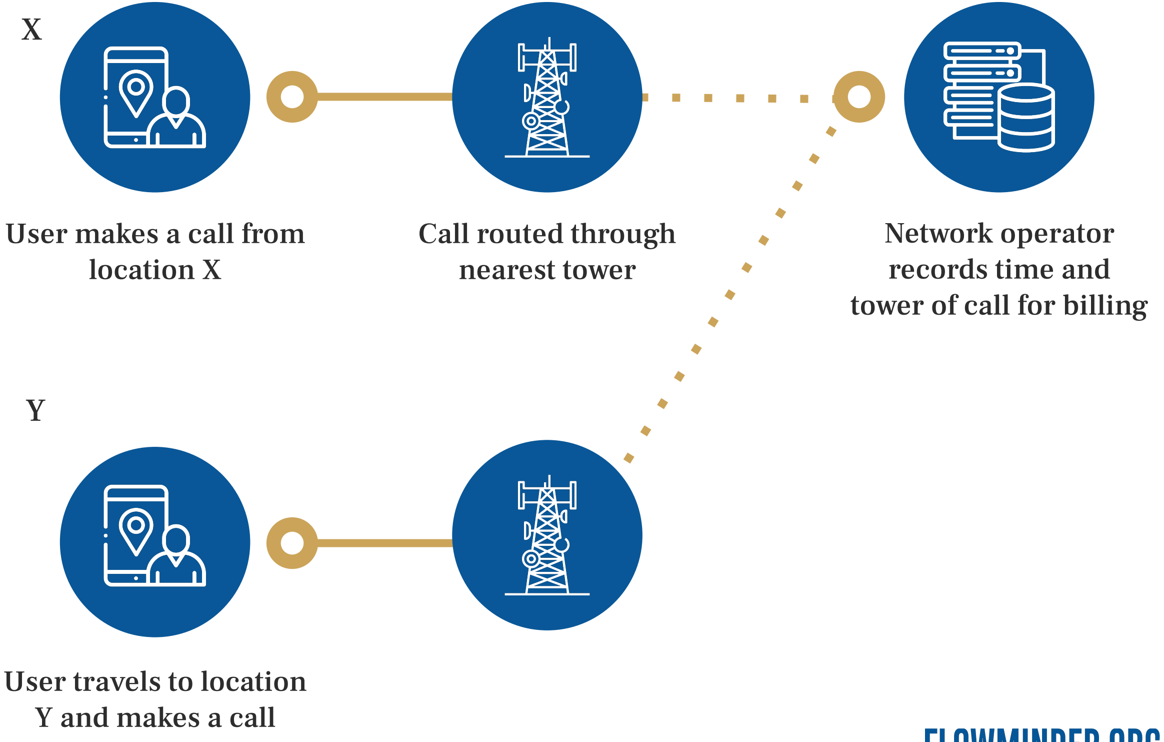
However, in LMICs, census or survey data are often **outdated** or **unavailable**

Conducting surveys in such countries can also be very challenging due to **inaccessibility** or **insecurity**

How can we estimate residents' mobility in such cases?

CDRs as alternative data source for near-real time estimates of population movements & changes in population density





User makes a call from location X

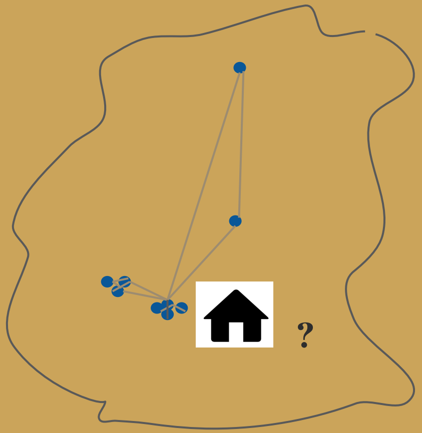
Call routed through nearest tower

Network operator records time and tower of call for billing

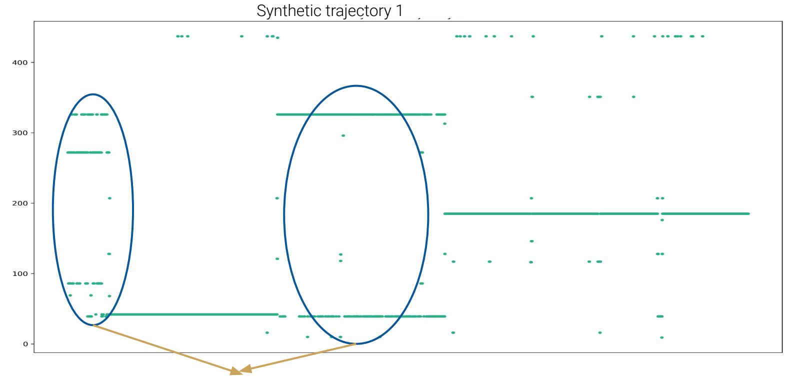
User travels to location Y and makes a call

FLOWMINDER.ORG

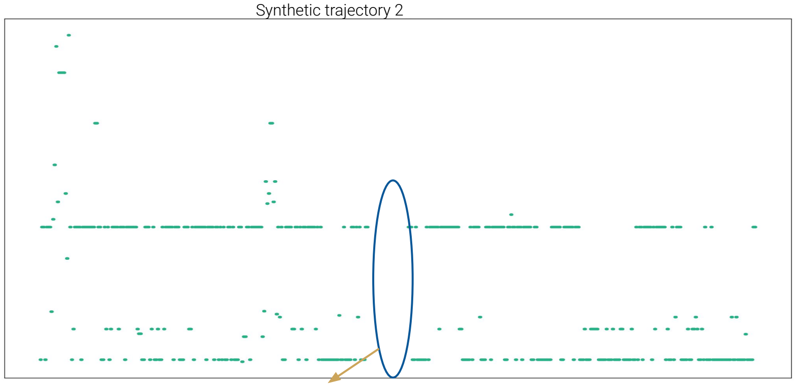
To detect migrations we need to detect home locations



Challenges in detecting home location using CDRs



Several potential home locations



Inactive period

Irregular calls

Infrequent calls

Short stays (1-3 days) overnight

Medium stays (4-7 days) overnight

Two or more similar frequency locations

Changing phone usage patterns

Ping-pong effect

Re-routing by a mobile operator

How do we know where people live? (1)

Monthly home location detection method

The **most frequent recent last-call-of-day location** in each 7 days window, moving by 1 day everyday

Daily

Assign **monthly home locations** for the current month

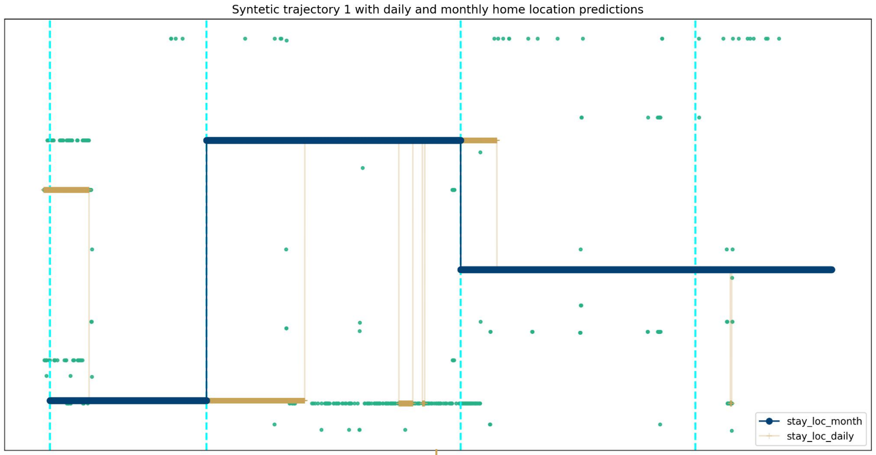
Monthly

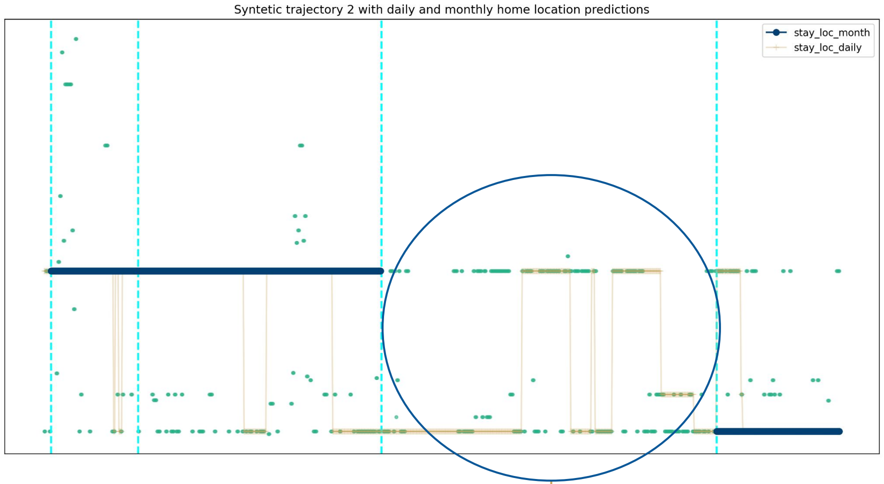
- If **the same location** is an **absolute majority of daily home locations** (more than half) in the current month → assign it as home location
- Else if **the same location** is **more than third of daily home locations** in the current month and **majority** in the previous month → assign it as home location
- Otherwise - assign ‘**unlocatable**’ for the current month

Home locations can be assigned to all subscribers or only to a set of active subscribers.

How do we know where people live? (2)

Monthly home location detection method: examples

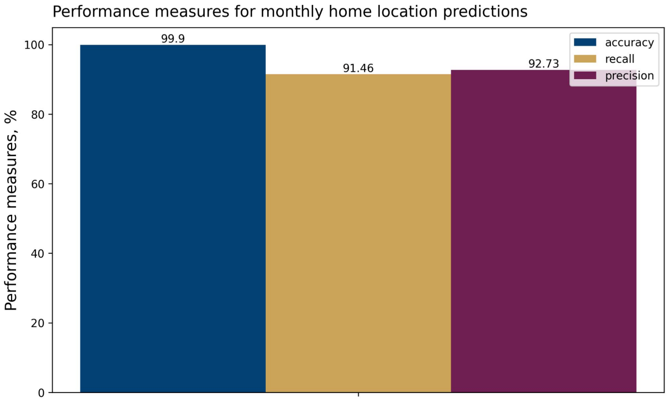


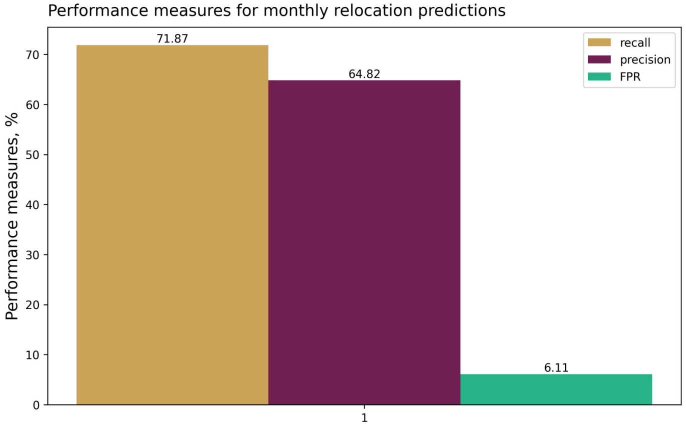


We can detect monthly home locations and months when relocations took place

Subscriber cannot be assigned monthly home location

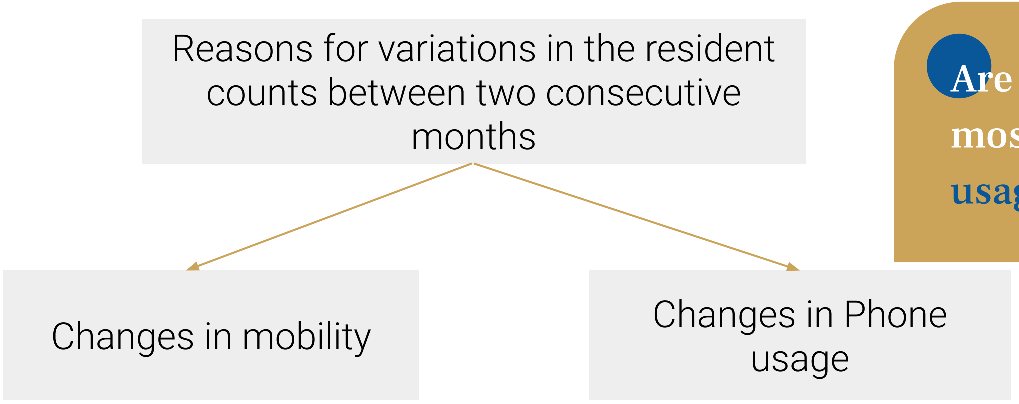
Validation of the monthly home location method





- **Validation** was done on **manually labelled 781 Digicel subscribers** in Haiti
- All algorithms were run on Haiti server for privacy protection.
- Performance measures are accuracy, recall, precision and false positive rate (FPR)
- **Monthly home location detection**
  - Accuracy, precision and recall are above 90%
  - FPR is 0.06%
- **Monthly home relocation detection**
  - Both the month of relocation and the location after relocation is taken into consideration.
  - Recall is ~70%, precision is ~65%, and FPR is ~6%

Variations in resident counts as a measure of mobility

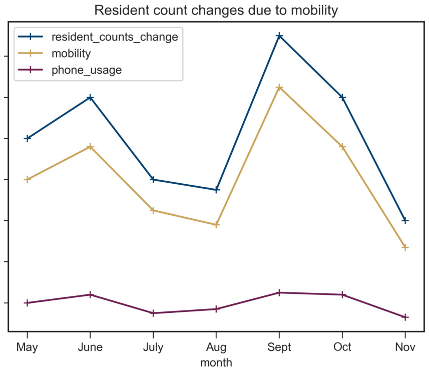


Reasons for variations in the resident counts between two consecutive months

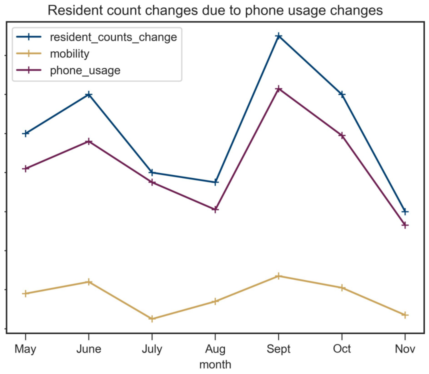
Changes in mobility

Changes in Phone usage

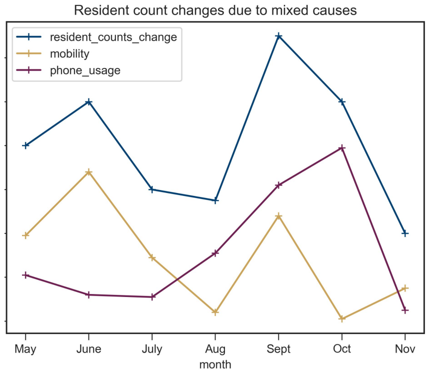
Are changes in resident counts influenced mostly by mobility or by changes in phone usage?



Resident count changes due to mobility



Resident count changes due to phone usage changes



Resident count changes due to mixed causes

Estimating residents from net flows, subscribers

$$\mathrm{est\_residents\_subscribers}_{an} = \mathrm{est\_residents\_subscribers}_{a(n-1)} + \mathrm{est\_netflow}_{an}$$

Where

- **$\mathrm{est\_residents\_subscribers}_{an}$** is the estimated number of resident subscribers in area a and month n
- **n=0** is a month corresponding to a baseline month (or a baseline period)
- **$\mathrm{est\_netflows}_{an}$** is the estimated netflow (difference between est_inflow and est_outflow) to area a in month n

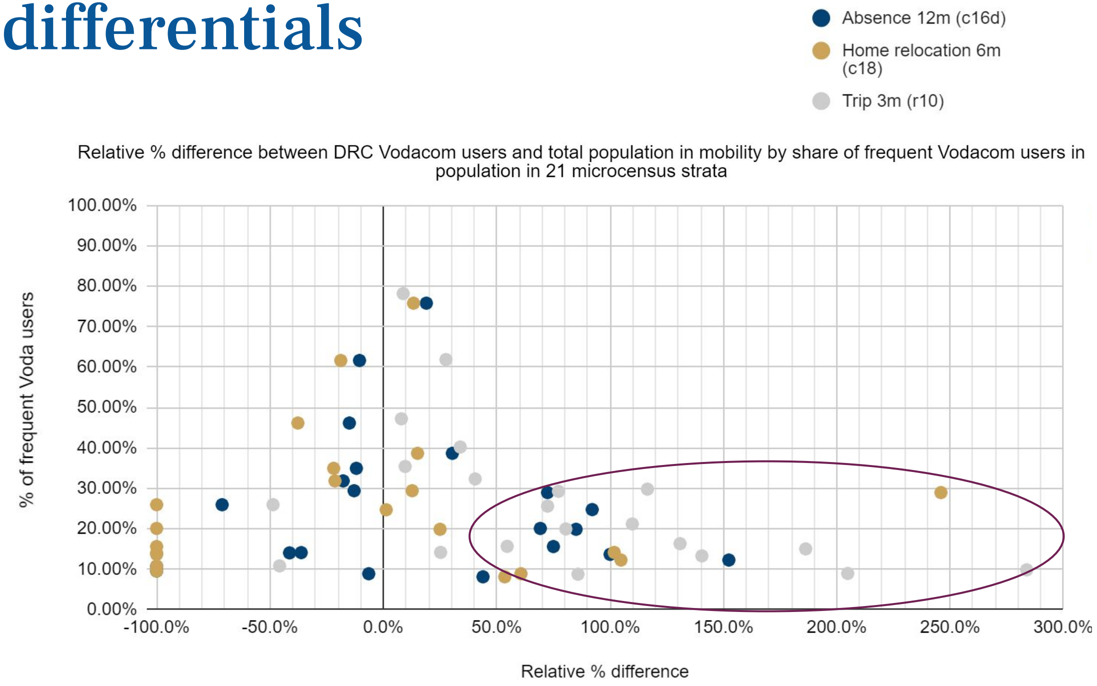
Bias-adjustment and scaling of mobility estimates from CDR aggregates

FLOWMINDER.ORG

MNO subscribers are
not a random sample of the
population, nor can be assumed to be.

Biases due to mobility differentials

- Based on survey data from the 2021 microcensus in the DRC, using three different mobility indicators, we identified some **large differences in mobility** between **Vodacom users** and the **rest of the population** (incl. non-phone-users)



DRC Microcensus 2021, final data (v2.2), weighted

- Across the three indicators and 21 microcensus strata, **15 parameters (out of 63) differed significantly** (i.e. more than the expected 5%)

FLOWMINDER.ORG

Flowminder has recently developed estimation methods to arrive at bias-adjusted & population-scaled estimates for

- Relocations from sub-region to sub-region, per month
- Residents per sub-region, per month

Bias-adjusted and population-scaled estimates

These estimates are based on

- CDR aggregates
- Primary & secondary survey data
- Existing population estimates
- Sub-region shapefiles

Method for monthly residents' estimates

- The estimate of residents in area a for month n (est_residents$_{an}$) is calculated as the sum of the baseline population for that area (est_base_pop$_{a}$) and by iteratively adding the cumulative sum of all net arrivals (est_netflow$_{amn}$) for all months between the baseline month and the current month, and by applying an area-specific rate of natural population growth (growthrate$_{a}$) to each monthly sum:

$$\text{est\_residents}_{a1} = \text{est\_base\_pop}_{a}$$ (Month 1 (baseline), m=0, n=1)

$$\text{est\_residents}_{a2} = (\text{est\_residents}_{a1} + \text{est\_netflow}_{a12}) * \text{growthrate}_{a}$$ (Month 2, m=1, n=2)

$$\text{est\_residents}_{a3} = (\text{est\_residents}_{a2} + \text{est\_netflow}_{a23}) * \text{growthrate}_{a}$$ (Month 3, m=2, n=3)

$$\ldots = \ldots$$

$$\text{est\_residents}_{an} = (\text{est\_residents}_{am} + \text{est\_netflow}_{amn}) * \text{growthrate}_{a}$$

- where the net arrivals estimate for area a between months m and n is the sum of all estimated inflows to that area minus all estimated outflows from that area:

$$\text{est\_netflow}_{amn} = \text{est\_inflow}_{amn} - \text{est\_outflow}_{amn}$$

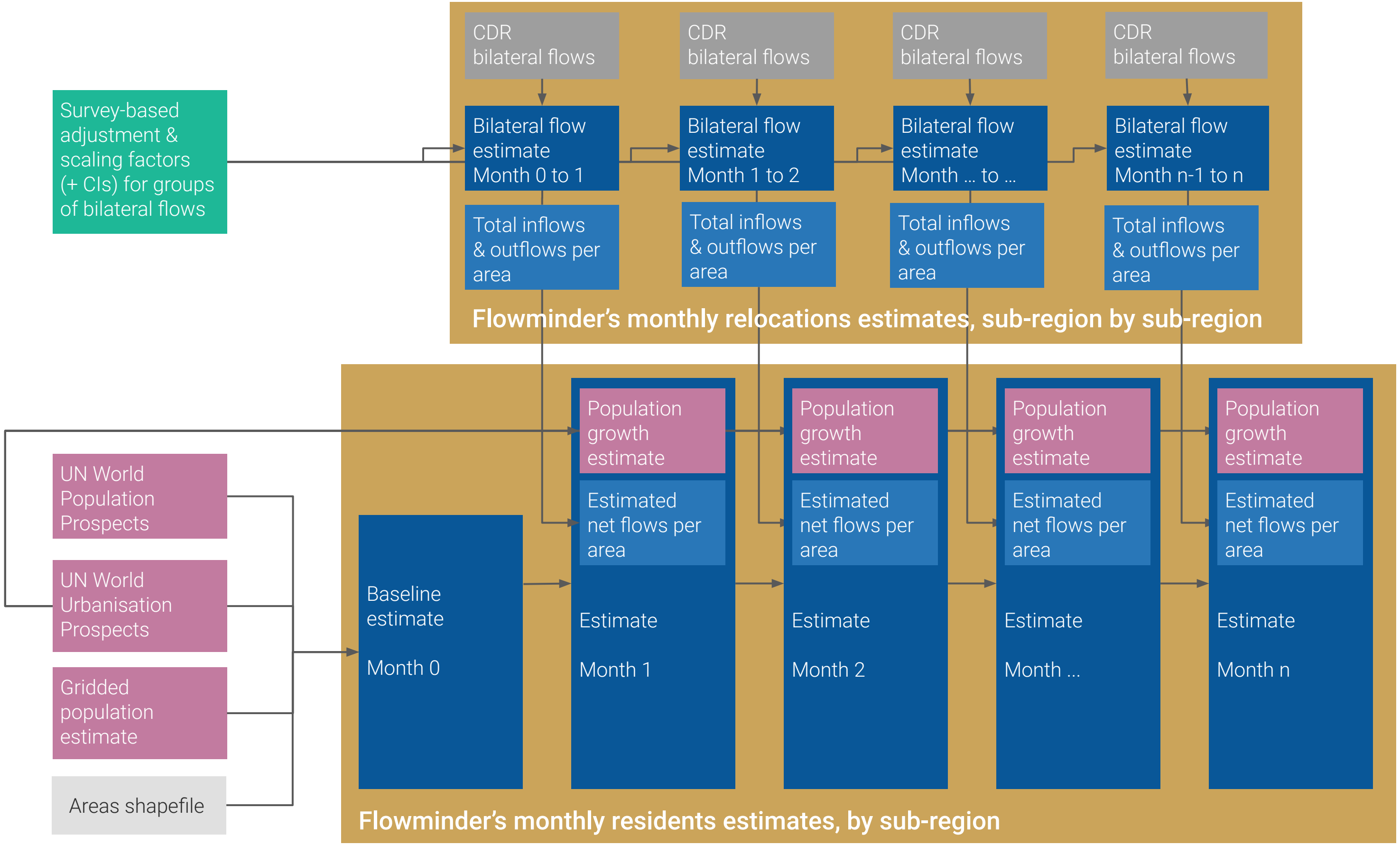
Method for relocations' estimates

- 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 users per SIM ($users_{ab}$) and 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:

$$est\_flow_{abmn} = cdr\_flow_{abmn} * (users_{ab}/sims_{ab}) * (1/mno\_share_{ab})$$

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

Method for monthly residents' estimates



Survey-based adjustment & scaling factors (+ CIs) for groups of bilateral flows

CDR bilateral flows

Bilateral flow estimate Month 0 to 1

Total inflows & outflows per area

CDR bilateral flows

Bilateral flow estimate Month 1 to 2

Total inflows & outflows per area

CDR bilateral flows

Bilateral flow estimate Month … to …

Total inflows & outflows per area

CDR bilateral flows

Bilateral flow estimate Month n-1 to n

Total inflows & outflows per area

Flowminder's monthly relocations estimates, sub-region by sub-region

UN World Population Prospects

UN World Urbanisation Prospects

Gridded population estimate

Areas shapefile

Baseline estimate

Month 0

Population growth estimate

Estimated net flows per area

Estimate

Month 1

Population growth estimate

Estimated net flows per area

Estimate

Month 2

Population growth estimate

Estimated net flows per area

Estimate

Month …

Population growth estimate

Estimated net flows per area

Estimate

Month n

Flowminder's monthly residents estimates, by sub-region

General caveats

Change estimates strongly depend on **baseline population number** - differs greatly between data sources

Limits to **granularity of survey estimates -** admin3 by admin3 would require very large survey sample sizes
or even census data

Currently only **cross-sectional survey data** used
(longitudinal data needed)

Lack of **validation data**: mobility and population estimates at admin3 are rare for LMICs

We continue to seek more data sources & **develop new methodologies** and methods.

Next steps

- Method refinement of **home relocation detection**
  - detection of relocations on daily/weekly basis
  - detection of short and medium stays
- Use **census data** (where available) for estimation models, or for **validation.**
- Test further **estimation models** (e.g. Machine Learning, Small Area Estimation, extrapolation)

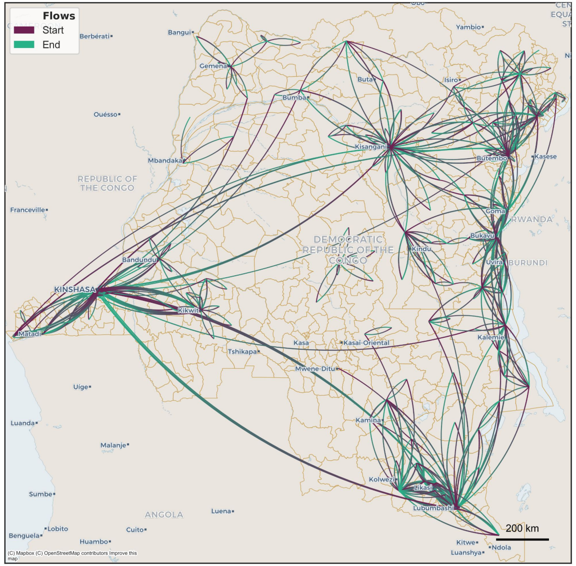
Adjustment and scaling of CDR time series aggregates are not trivial, but ultimately require highly complex estimation models based on multiple longitudinal & cross-sectional data sources

Producing monthly internal migration estimates for 3 countries

FLOWMINDER.ORG

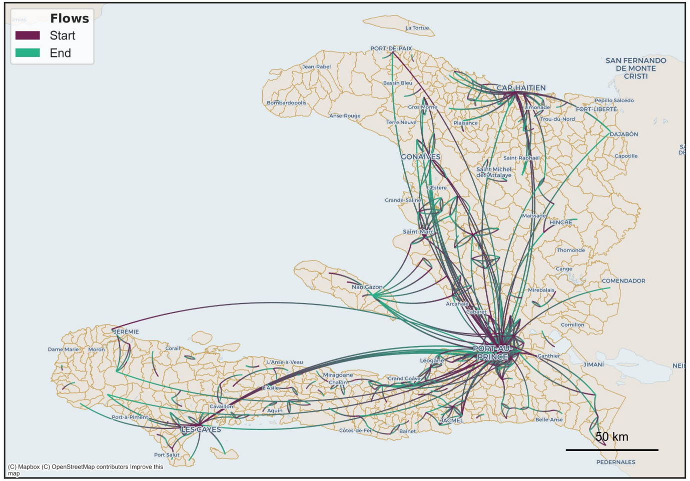
Monthly relocations between sub-regions

DRC



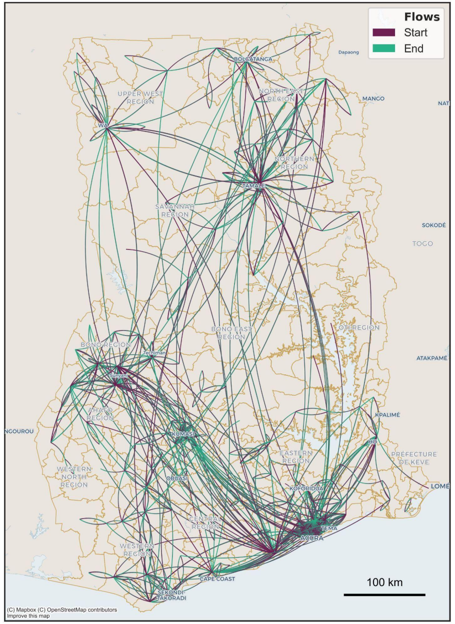
Note: estimated top 1,000 flows between health zones, median, Nov 2021 - Dec 2022

Haiti



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

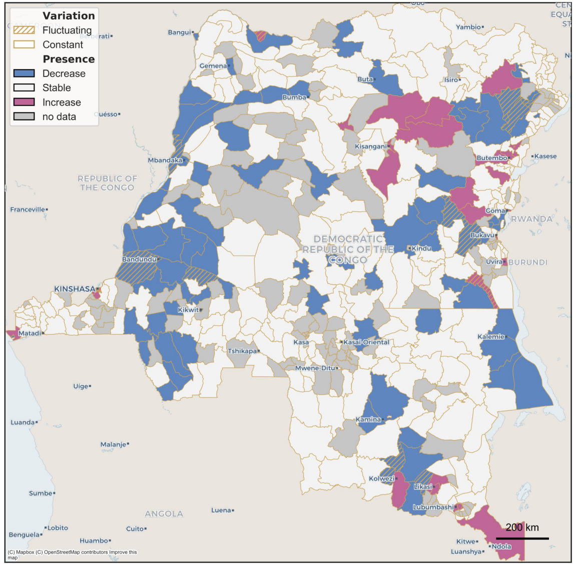
Ghana



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

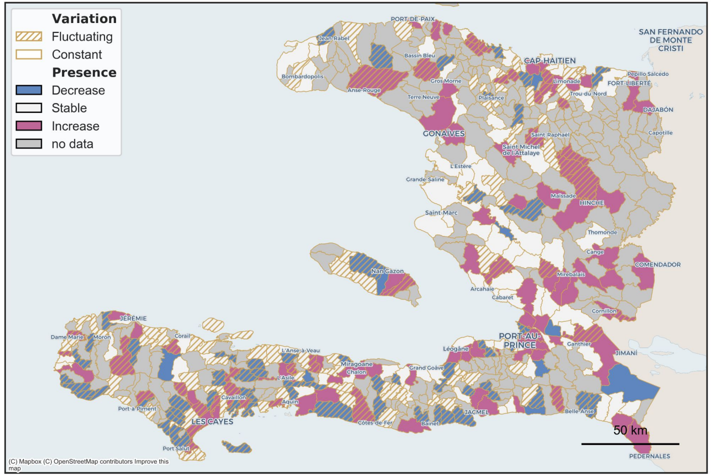
Monthly population change per sub-region

DRC



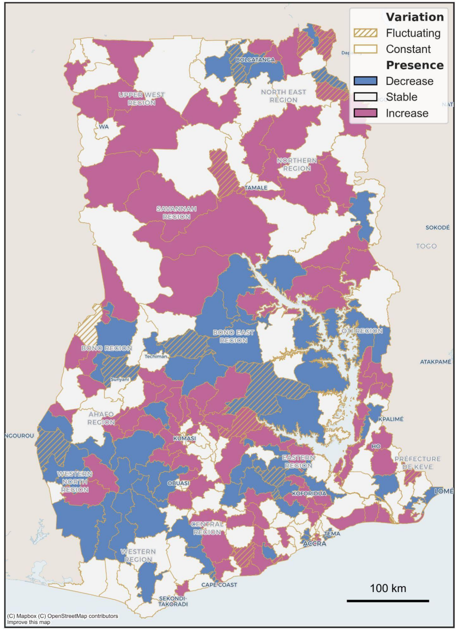
Note: average monthly variation in residents by health zone, Nov 2021 - Dec 2022, relative to baseline residents in each area

Haiti



Note: average monthly variations in residents by communal section, Feb 2020 - Feb 2022, relative to baseline residents in each area

Ghana



Note: average monthly variations in residents by district, Jan - July 2021, relative to baseline residents in each area

FLOWMINDER.ORG

Thank you!

FLOWMINDER.ORG

FLOWMINDER.ORG

Galina Veres
Senior Data Scientist
galina.veres@flowminder.org

Roland Hosner
Survey Statistician
roland.hosner@flowminder.org

🌐 www.flowminder.org

✉ info@flowminder.org

@Flowminder