

Population and mobility estimates for Ghana

Ghana | Methodological report and description of datasets

Estimates version 2.2

Population and mobility estimates for Ghana

Ghana | Methodological report and description of datasets

Estimates version 2.2

Last updated : 18 October 2025

Table of content

1. Summary of methodology	2
1.1 Assigning monthly home locations (per ID)	2
1.2 Assigning monthly relocations (per ID)	2
1.3 Estimating monthly relocations (per district)	2
1.4 Estimating monthly residents (per district)	2
2. Residents estimates	4
2.1. General information	4
2.2. Estimated residents per district	4
2.3. Estimated inflows per district	6
2.4. Estimated outflows per district	6
3. Relocations estimates	7
3.1. General information	7
3.2. Estimated relocations between districts	7
4. Data sources	9
4.1. Call Detail Records (CDRs), Telecel Ghana	9
4.2. Annual Household Income and Expenditure Survey 2022, GSS	9
4.3. Population Projections 2021 - 2050, GSS	9
4.4. District shapefiles	9
5. Limitations	10
5.1. "Measurement" biases	10
5.2. Representativity and biases	10

1. Summary of methodology

From the raw CDR datasets and the dataset on cell locations provided by the mobile network operator (MNO), we can identify the approximate location of an ID, based on the cell's location. In combination with the time of the event included in the call detail records (CDR) dataset, we assume the ID to be close to the cell location at that time.

1.1 Assigning monthly home locations (per ID)

We determine each ID's **home location** per month as the **district** that contains the cell phone towers near which the ID was located the **majority of the month** (if such majority location exists), or the most visited location that matches the majority location of the previous month. This is determined in three steps, according to the rules and thresholds described below:

Step 1: For each day in the month we create a 7-day look-back window (day-6, ... , day), and then determine for each 7-day window and each ID the **daily stay location** as the district containing the cluster of cell towers which most frequently routed the **last event each day**.

Step 2: A monthly home location is determined as the district which was the daily stay location for **more than half of the calendar month**.

Step 3: If there is no district with an absolute majority of daily stay locations for the current month, but **more than a third of days in the current month** and **more than half of the days in the previous month**, then that district section is assigned it as the ID's home location for the current month.

If no district fulfils these criteria (step 2 and 3) then the location of the ID was not stable enough to determine a home location for the current month and it is categorised as **unlocatable**.

1.2 Assigning monthly relocations (per ID)

We define **relocations** as a **change in this home location** from one month to the next.

1.3 Estimating monthly relocations (per district)

For the demographic accounting approach we use, we first need to estimate relocations per month (inflows and outflows) in order to estimate residents per month.

To estimate the number of monthly relocations between all districts (the origin-destination matrix), we **weight** relocations between districts using a SIM-to-user parameter and a combination of the population coverage of CDR data in the origin district and the destination district (see [section 3.2](#)).

1.4 Estimating monthly residents (per district)

Estimated **residents per district per month** are conceptualised as the population that spent the majority of the month in that district (the "de facto" population). The calculation steps are:

- We sum up all weighted relocations to each district per month ([Total inflows](#)) and all weighted relocations from each district per month ([Total outflows](#)) to compute estimated **net relocations** for each district ([Net flows](#)). This is the difference in residents due to monthly mobility only.

- We then use population counts from the 2021 census as **baseline population estimates** for July 2021.
- For each month after the census month, we **add net relocations** (inflows minus outflows) for each district between that month and the next month to the baseline estimates. Only the mobility detected from CDRs is used, in combination with weights, to estimate residents - not the count of CDR-derived home locations. We multiply these monthly estimates by **district-specific monthly population change rates** derived from **Ghana Statistical Service's annual population projections** to account for those demographic components not covered in CDR data (births, deaths, immigration, emigration).
- For all months prior to the census month, we **subtract net relocations** (inflows minus outflows) for each district between that month and the previous month from the baseline estimates, and divide these estimates by the monthly population change rates.
- We **repeat** this process for each month available in CDR data, to estimate residents from the existing baseline population, internal mobility and other population changes that occurred since then.

2. Residents estimates

2.1. General information

2.1.1. Description

A dataset containing monthly estimates of the de-facto population per district per month, since Jan 2020.

2.1.2. Temporal units and coverage

Each month in the data covers a **calendar month**. Estimates for individual districts may include gaps or shorter temporal coverage of estimates.

2.1.3. Geographic units and coverage

The estimates cover **251 of the 261 administrative districts** of Ghana. For four districts, no CDR data are available.

- Pru West (Bono East)
- Krachi West Municipal (Oti)
- Nanumba South (Northern)
- Lambussie Karni (Upper West)

Further six districts had CDR data but were discarded, since the population coverage (share of CDR-derived home location counts in the general population) was never above 1% in any month.

- Nkwanta South Municipal (Oti)
- Nkwanta North/Kpassa (Oti)
- Tatale Sanguli (Northern)
- Mion (Northern)
- Yunyoo Nasuan (North East)
- Tempene (Upper East)

District boundaries from the 2021 Population and Housing Census are used.

2.1.4. Baseline population estimates

For the resident estimates the **population counts of the 2021 Population and Housing Census, compiled by Ghana Statistical Service**, were used as baseline estimates.

2.2. Estimated residents per district

In general, the estimate of de facto residents in district a for month n (est_pop_{an}) is calculated as the sum of the population for that district in the previous month m (est_pop_{am}) and the net relocations for that district between the two months ($est_netflows_{amn}$), multiplied by a population change factor ($1 + change_rate_a$).

2.2.1. Calculation

The baseline month is July 2021 for which the estimate of residents is equal to the census count. The estimate of residents can be expressed as a system of recursive equations.

For all months after the census month July 2021, the sum of all net arrivals per district ($est_netflows_{amn}$) is added, and this sum is multiplied by district-specific monthly rate of population change ($1+change_rate_a$). For all months prior to the census month July 2021, the sum of all net arrivals per district ($est_netflows_{amn}$) is subtracted, and this sum is divided by the district-specific rate of population growth ($1+change_rate_a$).

$$\begin{aligned} est_pop_{an} &= (est_pop_{an} - est_netflows_{amn}) / (1 + change_rate_a) && \text{for all months } m \text{ before July 2021} \\ est_pop_{am} &= census_pop_a && \text{for } m = \text{July 2021} \\ est_pop_{an} &= (est_pop_{am} + est_netflows_{amn}) * (1 + change_rate_a) && \text{for all months } n \text{ after July 2021} \end{aligned}$$

Where:

est_pop_{an} is the residents estimate for district a for the current month n
 est_pop_{am} is the residents estimate for district a for the previous month m
 $est_netflows_{amn}$ is the estimated total net relocations for district a between months m and n
 $change_rate_a$ is the estimated average monthly population change rate for district a in percent
 $census_pop_a$ is the census population count for district a for July 2021

The net relocations estimate for district a between months m and n ($est_netflows_{amn}$) is the sum of all estimated relocations to that district ($est_inflows_{amn}$) minus the sum of all estimated relocations from that district ($est_outflows_{amn}$):

$$est_netflows_{amn} = est_inflows_{amn} - est_outflows_{amn}$$

See also [estimated inflows](#) and [estimated outflows](#).

2.2.2. Filters and redactions

Values for any month per district for which the population coverage (share of CDR-derived home location counts in the general population) was below 1% were redacted to missing.

2.2.3. Upper and lower bound estimates

$est_pop_LB_{an}$ is the residents estimate for district a for month n , lower bound
 $est_pop_UB_{an}$ is the residents estimate for district a for month n , upper bound

2.2.4. Monthly change rates

Estimates of monthly population change rates for the years 2021 to 2025, specific for each district, were derived from Ghana Statistical Service's [population projections](#) from 2021 to 2050.

The projected annual change rates for Ghana from 2021 to 2025 are shown in table 1, as well as monthly change rates derived from annual change rates. Monthly change rates for 2020 were projected backward from annual change rates between 2021 and 2023.

Table 1: District-specific annual and monthly population change rates in %

Annual change rates	2020/2021	2021/2022	2022/2023	2023/2024	2024/2025	2025/2026
Min	-	+1.157	+1.163	+1.155	+1.147	+0.983
Mean	-	+2.213	+2.192	+2.152	+2.109	+1.785
Max	-	+3.846	+3.763	+3.649	+3.533	+2.956
Monthly change rates	2020	2021	2022	2023	2024	2025
Min	+0.094	+0.095	+0.096	+0.096	+0.096	+0.088
Mean	+0.185	+0.183	+0.182	+0.179	+0.176	+0.161
Max	+0.325	+0.318	+0.312	+0.304	+0.294	+0.266

Source: Population projections 2021-2050, GSS, own calculations

2.3. Estimated inflows per district

The estimated number of people who relocated (i.e. moved) to a district (from all other districts) between the previous and the current month. In other words, this is the total of all inflows to a district.

2.3.1. Calculation

The sum of inflows to district a in month n is calculated as the sum of estimated relocations to district a from all other districts b between months m and n :

$$est_inflows_{amn} = \sum_{b=1}^k est_flows_{bamn}$$

Where:

est_flows_{bamn} is the number of estimated relocations to district a from all districts b ($b \neq a$) between months m and n

For the calculation of estimated bilateral relocations (est_flow_{bamn}), see [estimated relocations](#).

2.3.2. Filters and redactions

None.

2.4. Estimated outflows per district

The estimated number of people who relocated from a district (to all other districts) between the previous and the current month. In other words, this is the total of all outflows from a district.

2.4.1. Calculation

The sum of outflows from district a in month n is calculated as the sum of estimated relocations from district a to all other districts b between months m and n :

$$est_outflows_{amn} = \sum_{b=1}^k est_flows_{abmn}$$

Where:

est_flows_{abmn} is the number of estimated relocations from district a to all districts b ($b \neq a$), between months m and n

For the calculation of estimated bilateral relocations (est_flows_{abmn}), see [estimated relocations](#).

2.4.2. Filters and redactions

None.

3. Relocations estimates

3.1. General information

3.1.1. Description

A dataset containing estimates of month-to-month relocations of population between Ghana districts since Jan 2020.

3.1.2. Temporal units and coverage

Relocation estimates, month-to-month, are operationalised as changes of estimated stay locations ("home locations") between the previous month and the reference month. For example, relocation estimates for March 2024 cover changes in the stay locations between February and March 2024.

Individual districts may include gaps or shorter temporal coverage of estimates.

3.1.3. Geographic units and coverage

The estimates cover **directional population flows** between Ghana districts across all regions. District boundaries from the 2021 Population and Housing Census are being used.

3.1.4. Redactions

None.

3.2. Estimated relocations between districts

The estimated number of persons relocating from one district to another district between the current and the previous month.

3.2.1. Calculation

Relocations from district a to district b between months m and n are estimated based on CDR aggregates of relocations, i.e. the number of IDs changing their home locations from district a to district b between those months.

A home location is determined as the district containing those cell towers which most frequently (and in at least 3 separate weeks) routed the last event of the day of an ID over a calendar month. For each ID, relocations are then detected as a change in the district of the home location from one month to the next.

Then CDR aggregates of relocations (cdr_flows_{abmn}) from district a to district b between months m and n are multiplied by weights (w_{ab}):

$$\begin{aligned} est_flows_{abmn} &= cdr_flows_{abmn} * w_{abmn} \\ w_{abmn} &= 1/geom(sims_o, sims_d) * (1/median_6m(geom(cdr_to_est_{am}, cdr_to_est_{bm})))^{att} \end{aligned}$$

Where:

est_flows_{abmn} is estimated relocations from district a to district b between months m and n

cdr_flows_{abmn}	are CDR-derived relocations from district a to district b between months m and n
w_{abmn}	are weights for relocations from district a to district b between months m and n
$geom()$	is the geometric mean
$sims_o$	is the average number of Telecel SIM cards per Telecel user per origin province
$sims_d$	is the average number of Telecel SIM cards per Telecel user per destination province
$median_6m()$	is the 6-months median for months $m, m-1, \dots, m-5$
$cdr_to_est_{am}$	is the proportion of CDR residents (home locations) out of the estimated population for the origin district a in month m
$cdr_to_est_{bm}$	is the proportion of CDR residents (home locations) out of the estimated population for the destination district b in month m
att	is the attenuation factor for weights (power shrinkage)

The parameters $sims_o$ and $sims_d$ are calculated from the **Annual Household Income and Expenditure Surveys (AHIES)** and capture the average number of Telecel SIMs per Telecel user. It is calculated as the geometric mean of the parameters $sims_per_user$ for the origin province and the destination province. The term $1/geom(sims_o, sims_d)$ downscales the CDR relocation aggregates to account for multiple CDR records of the same users. For the years 2020 to 2022, the parameters are calculated from the AHIES 2022 survey, for the years 2023 to 2025, the parameters are calculated from the AHIES 2024 survey.

The second component of the weights - the term $(1/median_6m(geom(cdr_to_est_{am}, cdr_to_est_{bm})))^{att}$ - accounts for the varying shares of CDR home locations in the total population per district per month. Analyses have shown that flow sizes depend both on origin and destination population sizes, so this component of the weights is calculated as the inversion of the geometric mean of CDR-to-population shares in the origin and destination districts a and b , for month m . To avoid large fluctuations in weights due to low home location counts in some months, the 6-month median of the geometric mean is used.

This term is then downscaled by taking the root (exponent att), to reduce the variance of the weights.

$$w_{abmn} = 1/geom(sims_o, sims_d) * (1/median_6m(geom(cdr_to_est_{am}, cdr_to_est_{bm})))^{att}$$

For the **upper bound estimates**, this downsampling is implemented by using a higher value for att .

For the **lower bound estimates**, this downsampling is implemented by using a lower value for att .

NOTE: Relocations refer to directional bilateral relocations, from district a to district b . These are not usually equal to the number of relocations from b to a .

3.2.2. Filters and redactions

None

4. Data sources

4.1. Call Detail Records (CDRs), Telecel Ghana

For the current version of the estimates, monthly stay locations and month-to-month relocations derived from **Call Detail Records** from **Telecel Ghana** are used. For billing purposes, MNOs keep a record of an ID's activities in a database. These records are generated each time an ID makes or receives a call, sends or receives an SMS, or uses mobile data. These are called Call Detail Records (CDRs). CDRs contain information about the sending and receiving ID of a call or text, and duration of a call or data volume of data session, as well as the ID of the cell routing the call.

4.2. Annual Household Income and Expenditure Survey 2022, GSS

The Annual Household Income and Expenditure Surveys (AHIES) are designed and implemented by Ghana Statistical Service. Flowminder submitted further questions on mobile phone use and mobility to GSS, which were added to the questionnaire in 2022 and 2024. Data were collected via CAPI interviews in multiple rounds. Flowminder derived the survey parameters from the datasets for Q2 and Q3 2022. Sample sizes of households and household members are shown in table 2.

Table 2. Sample sizes of the AHIES datasets used

	2022			2024			
	Q1	Q2	Q3	Q1	Q2	Q3	Q4
Households	6,199	10,334	10,387	9,890	9,877	9,875	9,783
Household members	30,172	52,919	53,135	48,063	49,384	56,180	53,073

Source: AHIES 2022, 2024

4.3. Population Projections 2021 - 2050, GSS

The population projections published by GSS in June 2024 are the source used to derive projected population change rates per district. Annual change rates were derived from projected total population counts per district per year. Monthly change rates were derived from annual change rates.

Monthly residents estimates were also compared to annual population projections for cross-validation.

4.4. District shapefiles

The aggregates and estimates are produced based on a district boundary shapefile provided by GSS. The relevant shapefile represents the district boundaries used for the 2021 Population and Housing Census and covers 261 administrative distinct districts in Ghana, with Metropolitan Areas as single districts (no sub-districts).

5. Limitations

5.1. “Measurement” biases

The accuracy, precision and overall validity of CDR-derived statistics depends on both “measurement” and representation errors or biases.

In general, the **density of cell towers** in an area affects the **precision** of the **location** estimates. The density of cell towers is usually higher in urban areas and lower in rural areas.

The extent to which CDR-derived statistics are a reliable proxy for population mobility depends on the correspondence between **activities observable** from available CDR data and users’ actual activities. Travel episodes which take place unpunctuated by communications (e.g. when SIM cards are not in use or are outside of covered areas) are non-retrievable.

Conversely, an increased **frequency of phone usage** either at the individual or at the aggregate level may correspond to increased measured mobility, as movements previously occurring in-between communications are then captured. Also, CDRs recorded for each ID are not distributed uniformly over time and depend on the frequency with which users initiate and/or receive calls, send text messages and/or use data sessions for internet connectivity on their phones. We only ‘see’ a subscriber when they use their phone. If they don’t use their phone on a particular day, we can’t confidently say where they are on that day. Subscribers with low use frequency tend to be dropped from analysis - as missing episodes from individual trajectories cannot be inferred, or as they contribute a negligible amount of information towards group aggregates. Lower frequency use could, however, be associated with different mobility patterns and typically lower mobility.

Another source of bias is the extent to which the CDRs corresponding to an individual ID actually correspond to a **single individual**. Individuals may use multiple SIM cards concurrently, or frequently change the SIM card they use, and a SIM card may be shared among multiple individuals. Ownership may also be transferred to a different user or groups of users.

Measurement biases complicate the extraction of information from CDRs (on mobility and on locations of meaningful places such as the home location) and complicate inference on the basis of behavioural traces in CDRs.

5.2. Representation biases

Biases in terms of representation can be categorised into three main categories. The main **coverage and selection errors** of CDR-derived statistics are linked to

- the **share of mobile phone users** in a particular area’s population
- the **market share of particular MNOs** in the area, and
- the **subset of sufficiently active users** in the area that is used for analyses

Not everyone has access to a mobile phone. In Lower and Middle Income Countries, mobile phone ownership and (to a lesser extent) access has been shown to be disproportionately skewed towards individuals who are male, educated, urban-resident, wealthy and working-age, whose mobility differs from that of the general population.

Where CDR data are only available from a single MNO, this can add an undercoverage error, as it is not uncommon for the characteristics of a particular operators' subscriber base to be skewed towards relatively wealthier or relatively poorer individuals.

That is, CDRs capture only a **non-random subset of the population**, which may also change over time due to changes in phone and SIM prices, extensions in signal coverage, marketing and pricing strategies of MNOs, and demographic and societal changes. This may result in increased phone use for previous low-use groups, and lead to changes in population representation. These undercoverage errors and their temporal variations can distort the interpretation and use of observations derived from CDRs.

MNO's market shares and coverage areas change over time, which will have an impact on the analysis of CDR data. Furthermore, the error rate in MNO's data management and network issues is generally high, rendering analysis of their data more difficult due to data gaps in time and space as well as inconsistencies. These issues are monitored by Flowminder before the analysis stage, and mitigated where possible.

Contact us

For queries or information about the Ghana estimates, the methods presented in this document or on mobile data analytics, please contact us at info@flowminder.org.