FLOWMINDER.ORG

Haiti Mobility Data Platform

Indicators & metrics: description and methodology

Estimates version 3.0

Last updated: November 2024

Indicators and metrics: description and methodology

Estimates version 3.0

Last updated: 21 November 2024

List of content

General overview: about the data

All indicators and metrics made available on this platform have been derived from pseudonymised Call Detail Records (CDRs) from Digicel Haiti since January 2020, and adjusted for biases using survey data (except for the demo/experimental "presence" and "movements" estimates).

Temporal range of the indicators and metrics available on the platform as of publication date (November 2024): **January 2020 - October 2024**.

What are Call Detail Records?

CDRs are a type of information routinely recorded by mobile network operators (MNOs) about the use of the network by their subscribers, for billing purposes. Each time a subscriber is involved in a network event - whether that is making or receiving a call, sending or receiving an SMS message or using mobile data (**in Haiti, we only use calls**) - the operator records what type of network event it was, the time of the event, a phone number ID, and the cell tower ID which routed the event. Based on this information and the location of the cell towers, we can assess how people move within a country. Summing and analysing mobility of subscribers provides insights into the mobility of the population, while protecting the individual privacy of each subscriber (see Data protection and privacy).

For more detailed explanations about CDRs, please visit FlowGeek, our knowledge centre on CDR data analytics here.

Understanding the data

Resident and relocation categories

Turning phone usage data (CDR data) into population and mobility estimates

In order to extract usable mobility information which protects individual privacy and is relevant in the development and humanitarian sectors, CDR data need to go through several stages of processing and be combined with survey and population data. You can read more about the process here.

The platform is providing indicators and metrics (statistics) which address specific questions about the geographic distribution and mobility of the population in Haiti. This means that they do not contain any information about individual subscribers and represent the number of persons at a certain place and time.

Reducing biases & making the data representative of the population as a whole

The population and mobility information extracted from CDR only represent a subset of the subscribers of the participating MNO and not the whole population. Additionally, the frequency with which subscribers use their mobile devices also affect the accuracy and validity of

CDR-derived estimates, and there are risks that the data actually reflect coverage and phone usage changes rather than mobility changes (i.e. increased phone usage may result in the appearance of increased mobility).

The resident and relocation indicators and metrics (statistics) available on this platform are the results of years of research and method development, and have been corrected to:

- **Be representative of the whole population:** using survey data and existing population estimates, we scale our estimates so that they do not only represent numbers of Digicel IDs or subscribers. Our estimates represent the population as a whole, regardless of the operator used or whether a person uses a mobile phone.

and

- **Reduce the influence of phone usage behaviour:** we estimate the number of residents from observed mobility of subscribers (and existing population estimates) rather than from counting the number of subscribers residing in each location, to ensure that the data represent more actual mobility than varying phone usage.

The resident and relocation indicators and metrics presented on this platform are estimates of the overall population in Haiti, and not just Digicel subscriber counts. **They have been scaled and adjusted to be representative of the population as a whole - to measure actual mobility and the present population, reducing sensitivity to varying phone usage.** To do so, using advanced statistical models, we combined Digicel pseudonymised CDR data and mobility information extraction methods with recent field survey data and existing population estimates. We used population estimates from IHSI, WorldPop, and Meta/CIESIN. In an earlier release (July 2023), we had also included population change rate estimates from the UN published in 2018; however, as these estimates did not reflect the current crisis in Haiti, we have taken the decision to temporarily remove the change rate from our resident estimates.

The resident statistics are estimated for each month and each communal section. The relocation ones are estimates of mobility between pairs of communal sections each month in each direction. These estimates are calculated from the number of Digicel IDs (from Digicel subscribers) whose "home location" (their most visited location within a given month) changes between communal sections from one month to the next.

These estimates (in the resident and relocation category) are also adjusted to limit biases linked to the varying number of network events made by subscribers over time, SIM cards becoming active or inactive as well as the heterogeneous mobile phone penetration levels and market shares of Digicel across the country.

The estimates are fully anonymised, aggregated at communal section level (administrative level 3) per month, and are **our current monthly estimates of the number of people residing (staying) in, and relocating to, a specific area of Haiti.**

Demo 'experimental' estimates: Presence and Movements categories

Presence and movements estimates are currently a **demo "experimental' version - and not fully developed**; we recommend that you use these estimates for training and trialling purposes only, except for widespread events affecting the general population (such as the mobility restrictions that were in place during the COVID-19 pandemic), in which case, even though still in development, these estimates may be useful.

Presence estimates attempt to inform on the number of people who have been present in (visited) each communal section per day (people can be present in (visit) multiple communal sections in one day) and **movement** estimates attempt to capture the number of people travelling between pairs of communal sections per day.

However, these estimates have **not yet** been adjusted for two types of biases:

- **Representativity bias**: presence and movement (travel) estimates are more reflective of the distribution and mobility of the Digicel subscribers than of the general population
- **Phone usage influence:** estimates may be more influenced by changes in phone usage than by changes in presence and travel.

1. Representativity bias

While they are not direct counts of Digicel subscribers, the current presence and movement estimates do not *accurately* represent the mobility of the general population in Haiti.

We have not yet collected survey data on the differences of daily mobility between Digicel's subscribers, subscribers of other networks and non-phone users in Haiti. Without these data, we cannot estimate the error we make on general mobility when only looking at the mobility of Digicel subscribers, and therefore cannot adjust our estimates to the general population.

For example, these demo/experimental estimates may currently overestimate mobility in urban areas while underestimating it in rural areas.

2. Phone usage influence

The demo/experimental estimates are derived from **counts** of Digicel IDs (from Digicel subscribers) active each day, and the **communal sections** where they are active. Such counts can provide information on mobility but are also influenced by changes in phone use behaviour: where and when subscribers decide to make a call.

For example, the presence estimate may show a reduction every Sunday in most communal sections. While we expect presence to decrease in city centres on Sundays, we do not necessarily expect a decrease in residential areas. However, we know people make fewer calls on Sundays, leading to apparent reductions of 'presence' that is, in fact, a reduction in phone usage. Similarly, movement estimates may show a reduction of mobility on Sundays, also likely to result from a reduction in phone usage. Additionally, special events triggering increased phone use, or Digicel tariff changes, would also lead to changes in these demo/experimental estimates, which should then not be interpreted as changes in presence or travel.

As we have done for the resident and relocation estimates, our analysis team is working on survey data collection and further method development to attenuate both sources of bias, and provide estimates that can be used operationally.

To understand how we produce estimates from pseudonymised CDR data, please click here. You will be redirected to FlowGeek, our online knowledge centre on CDR data analytics.

Data protection & privacy

No personal data, such as an individual's identity, demographics, location, contacts or movements, is made available to the government or any other third party at any time. All of the results we produce and publish are aggregated (for example, the population number in a given municipality), which means that they do not contain any information about individual subscribers. The data is fully anonymised. This approach complies with the European Union's General Data Protection Regulation (EU GDPR 2016/679). Data are processed on a server installed behind Digicel Haiti's firewall in Haiti, and only aggregated data leave Digicel Haiti's premises.

1. Introduction

The Haiti Mobility Data Platform (haiti.mobility-dashboard.org) is a privacy-secure web-platform providing mobility and population estimates based on pseudonymised Call Detail Records (CDRs) from Digicel Haiti, survey data and further data sources, to approved third parties. These estimates aim to capture the distribution and movements of the Haiti population, and their dynamic changes over time. The platform allows users to visualise, interact and download the indicators and metrics for further analysis.

The indicators and metrics available on the platform, as of publication date (November 2024), are from **January 2020 to October 2024**.

To understand how we produce estimates from pseudonymised CDR data, please click here. You will be redirected to FlowGeek, our online knowledge centre on CDR data analytics.

1.1. Audience

This document is aimed at a technical/scientific audience and presents the estimates available on the Haiti Mobility Data Platform.

In this document, you will learn more about the available estimates, what they mean and measure, and how we calculated them.

1.2. General remarks

- All estimates are experimental **in the statistical sense**: currently, there is no standard methodology and quality criteria for CDR-derived statistics, the statistics shown on the platform have not yet been officially endorsed or officially released by Haiti's Statistical Office (IHSI), and may still contain gaps, biases and errors.
- In order to produce a better assessment of the robustness of our estimates, we are continuously working towards triangulating them with other data sources, as well as seeking collaboration and reviews by other experts, outside of our organisation.
- The CDR-derived estimates provided on this platform are useful to understand monthly mobility across the country, and are particularly informative in Haiti for humanitarian need assessments and other disaster preparedness activities. Compared to existing population and mobility data in Haiti, our estimates do quantify monthly changes in the population (population trends) in a consistent manner since 2020. This shows, for example, a decrease in the Port-au-Prince metropolitan area population, dipping under 3 million in early 2024 for the first time since 2020 - and notably lower than during the COVID-19 mobility restrictions - while in Les Cayes, for example, the population has been increasing (apart from a few temporary drops such as the 2021 earthquake) [see our reports 'Impact of the Haiti crisis on population mobility' on our website].
- Events triggering important population mobility (disasters, violence, restriction, holiday periods) are consistently observable in the CDR-derived estimates where there is coverage - while this is not a validation of the magnitude nor trend of the estimates, it is a

demonstration that our estimates provide useful timely information on mobility and population trends.

- At the time of this documentation the statistics are provided up to October2024 - our objective is to continue to update them on a monthly basis going forward.
- All estimates are calculated **per communal section** (residents, presence) or **pair of communal sections** (relocations, movements). Communal sections correspond to administrative units level 3 (admin3) in Haiti.
- All estimates described below are **point estimates** (not interval estimates). A dataset also containing related confidence intervals of resident and relocation estimates is available upon request.
- We sometimes use the term **'subscriber'** or **'Digicel ID'** to refer to pseudonymised MSISDNs (hashed phone numbers). All our CDR aggregates are based on pseudonymised **MSISDNs**, i.e. IDs of phone numbers. Two IDs may belong to the same individual, if that person uses more than one phone number. Conversely, one ID may represent several individuals, if a phone number is actively used by several people.
- **Baseline periods or months** differ by estimate, see the respective descriptions.

The documentation below lists all estimates (indicators and metrics) available on the Haiti Mobility Data Platform, grouped into the **four categories** of available estimates:

- **Residents**, per month
- **Relocations**, month-to-month
- **[Demo/Experimental] Presence**, per day
- **[Demo/Experimental] Movements**, per day

2. Estimates of residents and relocations

Our resident population estimates per month (from January 2020 to October 2024) are derived from Digicel pseudonymised CDR data, survey data and existing population estimates for January 2020. **CDR data are used to infer subscribers' home locations each month; changes in home location are used to estimate internal mobility (relocations between communal sections). Relocations and existing population estimates are used to estimate the number of de-facto residents of a communal section in a given month**. Therefore our estimates are based on scaled counts of moving IDs (from subscribers who seem to relocate), and not on counts of "residing" IDs (from subscribers who seem to be "resident" of a given location), to reduce the effect of phone usage (which leads to IDs appearing and disappearing in the dataset) in our estimates.

This documentation describes each estimate (indicator and metric) in the order they are presented on the platform ('Resident' category first, then 'Relocation'). However, it is important to note that in order to compute 'Residents' and other indicators of the 'Resident' category, we first need to compute 'Relocations'.

To guide the reader of this documentation, we provide below a short overview of the computation of the Relocation and Resident estimates:

Residents per communal section per month are conceptualised as the population that spent the majority of the month in that communal section (the "de facto" population).

We first estimate relocations between communal sections, from each month to the next:

- First, we assign each ID's **"home" location** per month as the communal section that contains the cell phone towers near which the ID was located the majority of the month (if no majority communal section then no location is assigned).
- Then, we compute **relocations** as a change in this home location from one month to the next.
- We **weight** relocations for pairs of communal sections in each direction using survey data from the MSNA 2022 and 2024 surveys and Flowminder's 2023 phone survey, as well as additional data we derive from the CDRs. This is our current method to correct for representativity biases (please see the section on the estimation of relocations for more information).
- We sum up all weighted relocations **into** each communal section per month (Total inflows) and all weighted relocations **from** each communal section per month (Total outflows) to compute the estimated **net relocations** for each communal section (Net flows). This corresponds to the difference in residents due to monthly mobility only (those who moved in minus those who moved out).

We then calculate **resident estimates for each communal section each month**:

- We calculate residents estimates per month from **baseline population estimates** derived from IHSI's 2015 and 2020 population estimates for **January 2020** (incl. some corrections, see annex). IHSI datasets are not available online at the time of writing.

- Then, we **sum the net relocations** (inflows minus outflows, 'incoming minus outgoing') for each communal section from January 2020 to February 2020, computed as described above from Digicel's CDRs and weighted with parameters derived from survey and additional CDR-derived data, to estimate changes in population due to internal mobility. Only the mobility detected in CDRs is used to estimate residents, but the count of home locations is used as part of the weighting methodology. **This is our method to reduce the influence of changes in phone use behaviour in our CDR-derived mobility and population estimates.**
- We **repeat this process** each month up to the current month, to estimate residents from the existing baseline population and internal mobility.

2.1. Residents (monthly)

2.1.1. Filters and redactions

Of the 570 communal sections in Haiti, **the resident estimates of this release** cover **393 sections** - but baseline estimates for January 2020 (not derived from CDRs) are given for all 570 communal sections. Among all 570 communal sections, 167 did not have any CDR-derived home location aggregates for any month of the reporting period. Of the remaining 403 communal sections, those sections for which the population coverage was below 1% were dropped (10 communal sections) for statistical reasons.

In addition, estimates **for any month and section** in which the population coverage was below 1% are being **redacted to missing** for statistical reasons, in an ongoing manner.

2.1.2. Residents

The estimated number of residents represents the **de facto population** per communal section per month. Residents are conceptualised as the population that spent the majority of the month in that communal section. These estimates are currently **unadjusted for population change due to births, deaths, immigration and emigration**, and are only considering internal mobility in Haiti.

Calculation

The calculation of resident estimates is based on:

- Changes in "home" communal section of a Digicel ID (assumed relocations)
- Weights derived from the MSNA 2022 and 2024 surveys, the Flowminder Phone Survey 2023, and other measures from Digicel CDRs to adjust for biases in the count of relocations between any two communal sections (relocations)
- Estimates of the total inflows and total outflows (relocations) for each section, resulting in estimated net relocations (net flows)
- Baseline population estimates derived from IHSI's 2015 and 2020 population estimates (please see Annex 1 for corrections of these from triangulation with other datasets, including with building footprints derived from high-resolution satellite imagery).

The estimate of residents in a communal section *a* for month *m* ($est\_residents_{a,m}$) is calculated as the sum of the population for that communal section in the previous month *m-1* ($est\_residents_{a,m-1}$) and the net relocations for that communal section between the two months (months *m-1* and *m*).

The baseline month is January 2020 ($m=0$) for which the estimate of residents is based on existing population estimates.

The **estimate of residents (unadjusted for population change due to births, deaths, immigration and emigration)** can be expressed as a system of recursive equations:

$$est\_residents_{a,m} = est\_residents_{a,m-1} + est\_netflow_{a,m-1,m}$$
$$est\_residents_{a,m=0} = est\_residents_{a,base}$$

Where:

$est\_residents_{a,m}$ is the population estimate for communal section a for the current month
$est\_residents_{a,m-1}$ is the population estimate for communal section a for the previous month $m-1$
$est\_netflow_{a,m-1,m}$ is the estimated total net relocations for communal section a between months $m-1$ and m
$est\_residents_{a,m=0}$ is the population estimate for communal section a for $m=0$ (January 2020) i.e. the baseline population estimate ($est\_residents_{a,base}$)

The net relocations estimate for communal section a between months $m-1$ and m ($est\_netflow_{a,m-1,m}$) is the sum of all estimated relocations to that communal section (total inflows: $est\_inflow_{a,m-1,m}$) minus the sum of all estimated relocations from that communal section (total outflows: $est\_outflow_{a,m-1,m}$):

$$est\_netflow_{a,m-1,m} = est\_inflow_{a,m-1,m} - est\_outflow_{a,m-1,m}$$

For the calculation of total estimated relocations to and from the communal section, see Total incoming (inflows) and Total outgoing (outflows).

Relocations estimates, in turn, are based on CDR-derived relocations, i.e. detected changes of home locations of an ID (from subscribers). A home location is determined as the communal section containing those cell towers which most frequently (and in at least 3 separate weeks) routed the last call of the day of an ID over a calendar month. If there is no call on at least 1 day per week in 3 separate weeks per month or there is no majority location, then no home location is assigned for that month and the subscriber is not considered a resident. Home locations are updated monthly. For each ID, relocations are then detected as a change in the communal section of the home location from one month to the next.

NOTE: The estimate of residents is computed using the system of recursive equations given above. These equations can also be written in an iterative manner to better understand the terms that make up the estimate of residents for a given month $m=n$ for communal section a ($est\_residents_{a,m=n}$):

$$est\_residents_{a,m=n} = est\_residents_{a,m=0} + est\_netflow_{a,m=0,m=1} + est\_netflow_{a,m=1,m=2} + \dots + est\_netflow_{a,m-1,m} +$$

... +
$est\_netflow_{a,m=n-1,m=n}$

This equation can be expressed concisely as:

$$est\_residents_{a,m=n} = est\_residents_{a,base} + \sum_{m=1}^{n} (est\_netflow_{a,m-1,m})$$

I.e. the cumulative net flows over time, added to the baseline population estimate.

All values are **rounded** to the nearest 100.

For adjustments from IDs to individuals, see relocations.

Confidence intervals

The additional variables showing the confidence intervals for the estimated number of residents (*est_residents_lb* and *est_residents_ub*) are available on request.

NOTE: Adjusting for births, deaths, immigration and emigration

A previous methodology [link: documentation from previous release, July 2023] included estimates of the population change rate derived from the UN World Urbanization Prospects 2018. However, as these did not reflect the current crisis in Haiti we have taken the decision to temporarily remove the change rate from our estimates.

2.1.3. Residents per square km

Population density, measured as de-facto residents per km^2 per month, is calculated from the number of estimated residents per communal section for the current month, divided by the area of that communal section.

Calculation

$$est\_res\_per\_km2_a = est\_residents_{a,m} / adm3\_km2_a$$

Where:

$est\_residents_{a,m}$	is the number estimated residents of communal section *a* in month *m*
$adm3\_km2_a$	is the area of communal section *a* in km^2

All values are **rounded** to the nearest 10.

2.1.4. Change in residents

The absolute change in estimated residents per communal section between January 2020 and the current month.

Calculation

$est\_res\_chg_{a,m=n} = est\_residents_{a,m=n} - est\_residents_{a,m=0}$

Where:

$est\_residents_{a,m=n}$	is estimated residents for communal section *a* in month *m=n*
$est\_residents_{a,m=0}$	is estimated residents for communal section *a* in the baseline month (*m=0*, Jan 2020)

NOTE: The estimated change in residents compared to the baseline estimate corresponds to the cumulative sum of estimated net relocations, as we do not account for other parameters of population change in this version (births and deaths, immigration and emigration)

All values are **rounded** to the nearest 10.

2.1.5. Relative change in residents (%)

The change in estimated residents per communal section, between January 2020 and the current month, expressed as a percentage of each communal section's baseline population estimate in January 2020.

Calculation

$est\_res\_chg\_pct_{a,m=n} = (est\_res\_chg_{a,m=n} / est\_residents_{a,m=0})*100$

And replacing $est\_res\_chg_{a,m=n}$ by its equation as above we obtain:

$est\_res\_chg\_pct_{a,m=n} = ((est\_residents_{m=n} - est\_residents_{a,m=0}) / est\_residents_{a,m=0})*100$

Where:

$est\_residents_{a,m=n}$	is the estimated number of residents of communal section *a* in month *m=n*
$est\_residents_{a,m=0}$	is the estimated number of residents of communal section *a* in the baseline month (*m=0*, Jan 2020)

2.1.6. Abnormality score

The abnormality score (which can also be termed 'change outlier score') indicates how different the estimated absolute change of residents between the current month and the previous month is, compared to the median absolute change of residents during the 12 preceding months (the reference period), measured in median absolute deviations.

Calculation

The abnormality (outlier) score for the monthly change of residents per communal section is calculated as a modified z-score, that is the difference between the change of residents from the previous month *m-1* to the current month *m*, standardising it by dividing through the median monthly change in residents during the reference period.

The score is calculated as:

$$z\_res\_chg_{a,m-1,m} = ((est\_residents_{a,m} - est\_residents_{a,m-1}) - median\_chg\_ref(est\_residents_{a})) / (1.486 * mad\_chg\_ref(est\_residents_{a}))$$

if the median absolute deviation (MAD) of the reference time series is not equal to 0. If the MAD is equal to 0, the mean absolute deviation is used in the denominator instead:

$$z\_res\_chg_{a,m-1,m} = ((est\_residents_{a,m} - est\_residents_{a,m-1}) - median\_chg\_ref(est\_residents_{a})) / (1.253 * meanad\_chg\_ref(est\_residents_{a}))$$

Where:

$z\_res\_chg_{a,m-1,m}$	is the abnormality (outlier) score for estimated change of residents for communal section *a*, between months *m-1* and *m*
$est\_residents_{a,m}$	is estimated residents for communal section *a* in month *m*
$median\_chg\_ref(est\_residents_{a})$	is the median estimated change of residents per month for communal section *a* during the reference period
$mad\_chg\_ref(est\_residents_{a})$	is the median absolute deviation of estimated residents per month for communal section *a* during the reference period
$meanad\_chg\_ref(est\_residents_{a})$	is the mean absolute deviation of estimated residents per month for communal section *a* during the reference period

The reference period is defined as the 12 months prior to the current month (with at least 3 available values), not including the current month.

NOTE: As the data start in January 2020, there are fewer than 12 months available as prior reference data for some 2020 months. In this case, a shorter reference period is used, with a minimum of 3 months as the shortest reference period (corresponding to 2 changes, i.e. the minimum to compute a reference). Dates for which there are fewer than 3 months of available reference data will not have an abnormality (change outlier) score, so **abnormality (change outlier) scores start with April 2020.**

Abnormality (change outlier) scores are helpful in identifying unusual changes, which can correspond to data issues or important real-world events that might be impacting people's place of residence. However, note that in case of unusual resident values over several consecutive months, the abnormality (or change outlier) score will only be large on the first month and the last month of the unusual period, i.e. **indicating transitions between usual and unusual time periods**, and not indicating unusual resident values - **only unusual change**.

For this score, a positive value greater than 3 indicates an unusual increase in the number of residents in the communal section; a value less than -3 indicates an unusual decrease in residents. Values between 3 and -3 are within the bounds of normal variation based on the reference period. Scores above 6 in absolute value are more likely to correspond to data issues, particularly in the absence of known disrupting events (disruption of mobility and/or phone usage).

NOTE: As per the equation we are currently using (excluding a population change rate), the estimated absolute change of residents between two consecutive months corresponds to the total net flow, therefore the change outlier score for residents currently corresponds to the net flow outlier score.

Filters and redactions

No values are calculated for the outlier score if less than 3 data points (months) are available. Values are rounded to 3 decimal points.

2.1.7. Total incoming (inflows)

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

Calculation

The sum of inflows to communal section *a* in month *m* is calculated as the sum of estimated relocations to section *a* from all other sections *b* between months *m-1* and *m*:

$$est\_inflow_{a,m\text{-}1,m} = \sum_{b=1}^{k} est\_flow_{b,a,m\text{-}1,m}$$

Where:

$est\_flow_{b,a,m\text{-}1,m}$ is the number of estimated relocations to communal section *a* from all sections *b*, between months *m-1* and *m*

For the calculation of estimated bilateral relocations ($est\_flow_{b,a,m\text{-}1,m}$), see **Relocations**.

Filters and redactions

Values are rounded to the nearest 10.

2.1.8. Total outgoing (outflows)

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

Calculation

The sum of outflows from communal section *a* in month *m* is calculated as the sum of estimated relocations from section *a* to all other sections *b* between months *m-1* and *m*:

$$est\_outflow_{a,m\text{-}1,m} = \sum_{b=1}^{k} est\_flow_{a,b,m\text{-}1,m}$$

Where:

$est\_flow_{a,b,m-1,m}$	is the number of estimated relocations from communal section *a* to all sections *b,* between months *m-1* and *m*

For the calculation of estimated bilateral relocations ($est\_flow_{a,b,m-1,m}$), see **Relocations**.

Filters and redactions

Values are rounded to the nearest 10.

2.1.9. Incoming minus outgoing (Net flows)

The **difference** between the number of people moving **into** (incoming, inflows) and **out of** (outgoing, outflows) a communal section between the previous and the current month. It describes the net change in the number of people residing in a communal section between two months that is due to internal mobility. This estimate can also be referred to as **'net relocations'** or **'net flows'.**

Calculation

The net flows estimate for communal section *a* between months *m-1* and *m* is the sum of all estimated relocations to that communal section (inflows) minus the sum of all estimated relocations from that communal section (outflows):

$$est\_netflow_{a,m-1,m} = est\_inflow_{a,m-1,m} - est\_outflow_{a,m-1,m}$$

Where:

$est\_netflow_{a,m-1,m}$	is estimated total net relocations for communal section a between months *m-1* and *m*
$est\_inflow_{a,m1-,m,}$	is estimated total relocations to communal section a between months *m-1* and *m*
$est\_outflow_{a,m-1,m}$	is estimated total relocations from communal section a between months *m-1* and *m*

Positive values represent net inflows, negative values represent net outflows.

Note that by design **the estimate of net flows is equivalent to the change in residents between two consecutive months**. As per the recursive equation given in the section 2.1.2. Residents and below, the estimated net flow is added to the estimated residents in the previous month to obtain the estimated residents for the current month:

$$est\_residents_{a,m} = est\_residents_{a,m-1} + est\_netflow_{a,m-1,m}$$

That is, no population change other than internal mobility (net flows) is included in this version of our estimates (population change due to births, deaths, immigration and emigration are not included).

Filters and redactions

Values are rounded to the nearest 10.

2.2. Relocations (monthly)

2.2.1. Filters and redactions

Only relocation corridors (relocations from communal section *a* to *b*) with a maximum time-series value ($cdr\_flow_{a,b,m-1,m}$) equal to or above 15 are shown on the platform. Corridors with maximum values below 15 have been excluded. Also, only values based on $cdr\_flow_{a,b,m-1,m}$ equal to or above 15 are shown, smaller values are redacted to missing, i.e. at least 15 MSISDNs (subscribers) need to have been observed to relocate for the derived estimate to be shown.

2.2.2. Relocations

The estimated number of persons relocating (i.e. changing their estimated monthly home location (as defined below)) from communal section *a* to another communal section *b* between the current and the previous month.

Calculation

Relocations from communal section *a* to communal section *b* between months *m-1* and *m* are estimated based on CDR aggregates of relocations (the number of IDs changing their home locations from communal section *a* to communal section *b*) between those months. A home location is determined as the communal section containing those cell towers which most frequently (and in at least 3 separate weeks) routed the last call of the day of an ID over a calendar month. For each ID, relocations are then detected as a change in the communal section of the home location from one month to the next.

Then CDR aggregates of relocations ($cdr\_flow_{a,b,m-1,m}$) from communal section *a* to communal section *b* between months *m-1* and *m* are adjusted accounting for the estimated SIM-to-user ratio ($geom(sims_{am-1},sims_{bm-1})^{-1}$) and for the population coverage of CDR data ($median\_6m(geom(pop\_coverage_{am-1},pop\_coverage_{bm-1})^{-1})$. *The latter term is downscaled by* attenuation factor (*att*):

$$est\_flow_{a,b,m-1,m} = cdr\_flow_{a,b,m-1,m} * geom(sims_{am-1},sims_{bm-1})^{-1} * median\_6m(geom(pop\_coverage_{am-1},pop\_coverage_{bm-1})^{-1})^{att}$$

Where:

$est\_flow_{a,b,m-1,m}$	is estimated relocations from communal section *a* to communal section *b* between months *m-1* and *m*
$cdr\_flow_{a,b,m-1,m}$	is CDR-derived relocations from communal section *a* to communal section *b* between months *m-1* and *m*
geom()	is the geometric mean
$sims_{am-1}$	is the nr of SIMs per user in section a for month m-1
median_6m()	is 6-month median

$pop\_coverage_{am-1}$	is the population coverage of CDR aggregates in section a for month m-1
att	*is the attenuation factor for weights (power shrinkage)*

The SIM adjustment factors used for relocations can be estimated from several survey datasets: the MSNA 2022 and 2024 surveys and Flowminder's 2023 phone survey).

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

Filters and redactions

Values are rounded to the nearest 10.

2.2.3. Change in relocations

The absolute difference in the number of estimated relocations from communal section *a* to communal section *b* for the months *m-1* and *m* in comparison to the reference months *basem0* and *basem1* (generally relocations from January 2020 to February 2020).

NOTE: This estimate will be updated to use a longer reference in our next release.

Calculation

$$est\_flow\_chg_{a,b,m,base} = est\_flow_{a,b,m-1,m} - est\_flow_{a,b,basem0,basem1}$$

Where:

$est\_flow\_chg_{a,b,m,base}$	is the absolute difference in estimated relocations from communal section *a* to communal section *b* for the months *m-1* and *m* in comparison to the reference months *basem0* and *basem1*
$est\_flow_{a,b,m-1,m}$	is estimated relocations from communal section *a* to communal section *b* for the months *m-1* and *m*
$est\_flow_{a,b,basem0,basem1}$	is estimated relocations from communal section *a* to *b* for the reference months *basem0* and *basem1*, the first available relocation estimate.

The reference estimate is usually relocations from January to February 2020. If no relocations estimate is available for these months, the reference relocations estimate is the first available estimate in the time series.

Filters and redactions

Values are rounded to the nearest 10.

2.2.4. Relative change in relocations (%)

 FLOWMINDER.ORG

The relative change of estimated relocations between communal section *a* and communal section *b* for the months *m-1* and *m* in comparison to the reference months *basem0* and *basem1*. It is expressed as a percentage of the number of relocations estimated for the reference months *basem0* and *basem1* (generally January 2020 to February 2020).

NOTE: This estimate will be updated to use a longer reference in our next release.

Calculation

$$est\_flow\_chg\_pct_{a,b,m,base} = (est\_flow_{a,b,m-1,m} - est\_flow_{a,b,basem0,basem1})/est\_flow_{a,b,basem0,basem1}$$

Where:

$est\_flow\_chg\_pct_{a,b,m,base}$	is the relative change in estimated relocations between communal sections *a* and *b* for the months *m-1* and *m* in comparison to the reference months *basem0* and *basem1*
$est\_flow_{a,b,m-1,m}$	is estimated relocations between communal sections *a* and *b* for the months *m-1* and *m*
$est\_flow_{a,b,basem0,basem1}$	is estimated relocations between communal sections *a* and *b* for the reference months *basem0* and *basem1*, the first available relocation estimate.

2.2.5. Abnormality score

The abnormality score (which can also be termed 'outlier score'), indicates how different the last estimated number of monthly relocations is (from communal section *a* to a communal section *b*), compared to the median monthly number of relocations estimated during the 12 preceding months (reference period), measured in median absolute deviations.

Calculation

The score is calculated as:

$$z\_est\_flow_{a,b,m-1,m} = ((est\_flow_{a,b,m-1,m}) - median\_ref(est\_flow_{a,b})) / (1.486 * mad\_ref(est\_flow_{a,b})$$

if the median absolute deviation (MAD) of the reference time series change is not equal to 0. If the MAD is equal to 0, the mean absolute deviation (meanAD) is used instead:

$$z\_est\_flow_{a,b,m-1,m} = ((est\_flow_{a,b,m-1,m}) - median\_ref(est\_flow_{a,b})) / (1.253 * meanad\_ref(est\_flow_{a,b})$$

Where:

$z\_est\_flow_{a,b,m-1,m}$	is the outlier score for the number of estimated relocations from communal section *a* to section *b*, between months *m-1* and *m*
$est\_flow_{a,b,m-1,m}$	is estimated relocations from communal section a to communal

	section b, between months *m-1* and *m*
median_ref(est_flow$_{a,b}$)	is the median monthly number of estimated relocations from communal section *a* to communal section *b* during the reference period
mad_ref(est_flow$_{a,b}$)	is the median absolute deviation of the monthly number of estimated relocations from communal section *a* to communal section *b* during the reference period
meanad_ref(est_flow$_{a,b}$)	is the mean absolute deviation of the monthly number of estimated relocations from communal section *a* to communal section *b* during the reference period

The reference period is defined as the 12 months prior to the current month (with at least 3 available values), not including current month.

A positive value for this estimate greater than 3 indicates an unusually large number of relocations (a statistical outlier); a value less than -3 indicates an unusually small number of relocations. Scores above 6 in absolute value are more likely to correspond to technical issues, particularly in the absence of known disrupting events (disruption of mobility and/or phone usage).

Filters and redactions

No values are calculated for the abnormality (outlier) score if less than 3 data points (months) are available. Values are rounded to 3 decimal points.

3. Demo (experimental) daily presence and movement categories

Unlike our resident and relocation estimates, our presence and movement (travel) estimates are at present a **demo/experimental version**; we recommend that you use these estimates for training and trialling purposes only, **except for widespread events affecting the general population** (such as the mobility restrictions that were in place during the COVID-19 pandemic), **in which case, even though still in development, these indicators and metrics may be useful.**

Presence estimates attempt to inform on the number of people that have been present in each communal section each day (within 24 hours). **Movement** estimates attempt to capture the number of people travelling between a pair of communal sections each day.

However, these estimates have not yet been adjusted for two types of biases and errors:

- **Representativity bias**: Demo/experimental presence and travel estimates are more reflective of the distribution and mobility of the Digicel subscribers than of the general population
- **Phone usage influence:** these estimates may be more influenced by changes in phone usage than by changes in presence and travel.

However, they do not provide direct counts of Digicel subscribers - they have been scaled with specific scaling factors, but not in a representative manner. They do not use the same robust adjustment and scaling method as our Resident and Relocation estimates.

Additionally, the extraction of mobility information from CDRs is not as robust to changes in phone usage as it is for our Resident and Relocation estimates. For example, changes in presence are not derived from movements, we do not exclude occasional and infrequent subscribers from the dataset, and we do not use a detection of meaningful locations such as home and work. This is why we recommend our demo/experimental Presence and Movement estimates for trialling purposes at this time; however, we note they may be of interest in case of very widespread and unusual events impacting the general population (such as mobility restrictions in place during the pandemic).

3.1. [Demo/experimental] Presence (daily)

3.1.1. Presence

The estimated **number of people that have been present** in each communal section per day (within 24 hours). It includes residents present in their home communal section as well as visitors and travellers. People can be **present and counted in several communal sections** during the same day.

Calculation

Presence in a communal section *a,* per day, is derived from the count of Digicel IDs (from Digicel subscribers) whose call(s) were routed by a cell tower in that communal section on that day ($cdr\_presence_{a,d}$).

We then compute the relative change in presence counts in a communal section between the current day ($cdr\_presence_{a,d}$) and the median presence counts during the baseline period (med($cdr\_presence_{a,base}$)), expressed as a proportion of the median presence counts during the baseline period. The baseline period used here is from August 2020 to September 2021 included (this baseline was chosen to avoid the unusual period of COVID-19 mobility restrictions from March to July 2020, and to have a long enough baseline).

We scale this relative change to the population by first multiplying it by a factor x_a, a scaling factor calculated for each communal section *a* as described below (see annex: scaling factors). If, for example, x_a = 0.5, then a presence increase of 20% 'subscribers' would lead to an estimated presence increase of 10% people. Then, we apply this scaled proportional change to the existing estimate of the population for each communal section *a* ($est\_residents_{a,m=0}$) to obtain the change in the number of people present on day *d* and the baseline, then add it to the existing population estimate to obtain the number of people present on day *d* communal section *a*.

$$est\_presence_{a,d} = est\_residents_{a,m=0} + est\_residents_{a,m=0} * x_a * ((cdr\_presence_{a,d} - med(cdr\_presence_{a,base})) / med(cdr\_presence_{a,base}))$$

Where:

$est\_presence_{a,d}$ is the estimated presence of people in communal section *a* on day *d*
$est\_residents_{a,m=0}$ is the baseline population estimate derived from IHSI's 2015 and 2020 population estimates

x_a	is the scaling factor for section *a* (see annex scaling factors)
$cdr\_presence_{a,d}$	is the count of Digicel IDs (MSISDNs) whose call(s) were routed by a cell tower in communal section *a* on day *d*
$med(cdr\_presence_{a,base})$	is the median presence count during the baseline period (August 2020 to September 2021)

If we define $est\_pres\_chg\_rel_{a,\,d,base}$ as the relative change in presence compared to the baseline period:

$$est\_pres\_chg\_rel_{a,\,d,base} = (cdr\_presence_{a,d} - med(cdr\_presence_{a,base})) \,/\, med(cdr\_presence_{a,base})$$

we can then simplify the presence equation as:

$$est\_presence_{a,d} = est\_residents_{a,m=0} + est\_residents_{a,m=0} * x_a * est\_pres\_chg\_rel_{a,d,base}$$

to emphasise that we scale the relative change in presence, then add it to existing population estimates.

Filters and redactions

Values based on $cdr\_presence_{a,d}$ under 15 IDs (from subscribers (pseudonymised MSISDNs)) have been redacted to missing for privacy purposes. However, to improve robustness of the estimates and avoid estimating changes in presence from a very small sample of subscribers, in a later version of this dataset, we will update this redaction threshold to a larger value (e.g. redacting any value under 200 pseudonymised MSISDNs, or any communal section with $med(cdr\_presence_{a,base})$ below 200).

Values of $est\_presence_{a,d}$ are rounded to the nearest 100.

3.1.2. Presence per km^2

The estimated **number of people present** per square kilometre in the communal section on the current day.

Calculation

We estimate presence per km^2 by dividing the estimated number of people present in the communal section *a* for day *d* by the area of that communal section, giving the spatial density of presence.

$$est\_presence\_per\_km2_{a,d} = est\_presence_{a,d} \,/\, adm3\_km2_a$$

Where:

$adm3\_km2_a$ is the area in km^2 of communal section *a*

Filters and redactions

Values are rounded to the nearest 10.

3.1.3. Change in presence

Difference in the estimated number of people present in the communal section, on the current day, and the estimated number of people present on the first day of available data.

Calculation

The change in presence is calculated as the difference in the number of people present in the communal section *a* on the current day *d* with the first day of available data.

$$est\_pres\_chg_{a,d} = est\_presence_{a,d} - est\_presence_{a,\ based0}$$

Where:

$est\_presence_{a,d}$	is the estimated presence of people in communal section *a* on day *d*
$est\_presence_{a,\ based0}$	is the estimated presence on the first day for which we have a presence estimate for communal section *a*.

This is a demo/experimental estimate. It will soon be replaced by an estimate comparing the number of people present in the communal section on the current day, with the median number of people present during a relevant baseline period (such as the preceding calendar year). This will help contextualise and interpret the presence value on a given day.

3.1.4. Relative change in presence (%)

The percentage change in the estimated number of people present in the communal section, on the current day, compared to the estimated number of people present on the first day of available data.

Calculation

$$est\_pres\_chg\_pct_{a,d} = ((est\_presence_{a,d} - est\_presence_{a,\ based0}) / est\_presence_{a,\ based0}) * 100$$

Where:

$est\_presence_{a,d}$	is the estimated presence of people in section *a* on day *d*
$est\_presence_{a,\ based0}$	is the estimated presence on the first day for which we have a presence estimate for section *a*

This measure standardises comparisons, allowing for meaningful comparisons across different scales. For instance, a change from 100 to 200 people represents a much larger relative increase than a change from 1,000 to 1,100 people, despite both being an absolute increase of 100 people. However, comparing current presence with a single day in 2020 may not be relevant.

This is a demo/experimental estimate. It will soon be replaced by an estimate comparing the number of people present in the communal section on the current day, with the median number of people present during a relevant baseline period (such as the preceding calendar year). This will help contextualise and interpret the presence value on a given day.

3.1.5. Abnormality score

The abnormality score, which can also be termed 'outlier score', indicates how different the last estimated presence per communal section is, compared to the median daily presence estimated during the 365 preceding days (the baseline period), measured in median absolute deviations. It describes how unusual the last presence estimate in a communal section is, compared to the daily presence estimated during the 365 preceding days.

Calculation

The abnormality/outlier score for the daily estimated presence per communal section is calculated as a modified z-score as follows:

1. Calculate the estimated presence on the current day d $est\_presence_{a,d}$.
2. Compute presence estimates for the previous 365 days, the baseline period, and calculate the median daily presence during that period ($median\_ref(est\_presence_a)$), the median absolute deviation of presence ($mad\_ref(est\_presence_a)$) and the mean absolute deviation of presence ($meanad\_chg\_ref(est\_presence_a)$).
3. Calculate the modified z-score for the current day using this 365 day period as the baseline period for 'expected presence'.

If the median absolute deviation (MAD) of the baseline time series is not equal to 0, the score is calculated as:

$$z\_est\_pres_{a,d} = (est\_presence_{a,d} - median\_ref(est\_presence_a)) / (1.486 * mad\_ref(est\_presence_a))$$

If the MAD is equal to 0, the mean absolute deviation (meanAD) is used instead:

$$z\_est\_pres_{a,d} = (est\_presence_{a,d} - median\_chg\_ref(est\_presence_a)) / (1.253 * meanad\_chg\_ref(est\_presence_a))$$

Where:

$est\_presence_{a,d}$	is the estimated presence of people in communal section a on day d
$median\_ref(est\_presence_a)$	is the median daily change of estimated presence in communal section a during the baseline period
$mad\_ref(est\_presence_a)$	is the median absolute deviation of estimated presence in communal section a during the baseline period
$meanad\_chg\_ref(est\_presence_a)$	is the mean absolute deviation of estimated presence in communal section a during the baseline period

The baseline period is defined as the 365 days prior to the current day (with at least 90 available values), not including the current day.

NOTE: As our data start in January 2020, there are fewer than 365 days of available prior baseline data for some months in 2020. In this case, a shorter baseline is used for the total number of days available, with a limit of 90 days as the shortest reference period (for reliability purposes). Dates for which there are fewer than 90 days of available baseline data do not have an outlier score, so **abnormality/outlier scores start in April 2020.**

These scores are helpful in identifying unusual presence, which can correspond to data issues or important real-world events that might be impacting people's presence in a communal section. Note that in case of unusual presence values over several consecutive days, the score will continue to be large if the unusual period is short, but the score will decrease over time if the presence remains at a similar level every day (i.e. the novel presence level becomes more usual).

A positive value for this estimate greater than 3 indicates an unusually large presence (a statistical outlier); a value less than -3 indicates an unusually small presence. Scores above 6 in absolute value are more likely to correspond to technical issues, particularly in the absence of known disrupting events (disruption of mobility and/or phone usage).

3.1.6. Total incoming (inflows)

The estimated number of inflows of travellers (incoming) to a communal section *a,* from any other communal section, per day.

Calculation

The estimated number of travellers to a communal section *a*, per day, is derived from the count of Digicel IDs (pseudonymised MSISDNs) who made a call from a different communal section (that is not *a*) and made an immediately subsequent call from the communal section *a* during the day *d* ($cdr\_travellers\_in_{a,d}$). This is then scaled (however not in a representative manner) using the adjustment term: $x_a * (est\_residents_{a,m=0} / med(cdr\_presence_{a,base}))$, where x_a is a scaling factor calculated for each communal section (see annex: scaling factors), $est\_residents_{a,m=0}$ is the existing estimate of the population for each communal section *a*, and $med(cdr\_presence_{a,base})$ is the median count of presence per day in the communal section during the baseline period (August 2020 to September 2021 included). For more explanation on the adjustment term and the baseline period, please see the section on the estimation of presence.

$$est\_travellers\_in_{a,d} = \sum_{b=1}^{k} cdr\_travellers\_in_{b,a,d} * x_a * (est\_residents_{a,m=0} / med(cdr\_presence_{a,base}))$$

Where:

x_a	is the scaling factor for section *a* (see annex scaling factors)
$est\_residents_{a,m=0}$	is the existing population estimate derived from IHSI's 2015 and 2020 population estimates
$med(cdr\_presence_{a,base})$	is the median estimated presence during the baseline period (August 2020 to September 2021)
$cdr\_travellers\_in_{a,b,d}$	is the count of Digicel IDs (MSISDNs) that made a call from a communal section section *b* and a subsequent call from communal section *a* during the day *d*

NOTE: Travellers to a communal section per day may be counted several times if they have entered the communal section from several other communal sections. However, travellers doing multiple trips to the communal section from a single communal section are only counted once. This is another issue with this type of demo/experimental estimate which we are working to resolve.

Filters and redactions

Values of $cdr\_travellers\_in_{a,d}$ under 15 IDs (MSISDNs) have been redacted to missing for privacy purposes. Values of $est\_travellers\_in_{a,d}$ (travelling population) under 50 are also redacted to missing as they originate from a sample of travelling IDs that is too small to provide a robust estimate. Values are rounded to the nearest 10.

3.1.7. Total outgoing (outflows)

The estimated number of outflows travellers (outgoing) from communal section *a* to any other communal section, during the current day.

Calculation

The estimated number of travellers from a communal section *a,* per day, is derived from the count of Digicel IDs (peudonymised MSISDNs) who made a call from the communal section *a* and made an immediately subsequent call from another communal section*,* during the day *d* ($cdr\_travellers\_out_{a,d}$).

This is then scaled using the adjustment term $x_a*(est\_residents_{a,m=0}/med(cdr\_presence_{a,base}))$, where X_a is a scaling factor calculated for each communal section as described below (see annex scaling factors), $est\_residents_{a,m=0}$ is the existing estimate of the present population per communal section *a*, and $med(cdr\_presence_{a,base})$ is the median count of IDs present in the communal section during the baseline period (August 2020 to September 2021 included). For more explanation on the adjustment term and the baseline period, please see the section on the estimation of present population.

$$est\_travellers\_out_{a,d} = \sum_{b=1}^{k} cdr\_travellers\_out_{a,b,d} * x_a * (est\_residents_{a,m=0}/med(cdr\_presence_{a,base}))$$

Where:

$cdr\_travellers\_out_{a,d}$	is the count of Digicel IDs (MSISDNs) that made a call from communal section *a* and made an immediately subsequent call from another section b during the day *d*
x_a	is the scaling factor for section *a* (see annex scaling factors)
$est\_residents_{a,m=0}$	is the existing population estimate derived from IHSI's 2015 and 2020 population estimates
$med(cdr\_presence_{a,base})$	is the median estimated presence during the baseline period (August 2020 to September 2021)

NOTE: Travellers from a communal section per day may be counted several times if they have left the communal section to go to several other communal sections. However, travellers doing multiple trips from the communal section to a single communal section are only counted once. This is another issue with this type of demo/experimental estimate which we are working to resolve.

The difference 'total inflows - total outflows' corresponds to a net number of travellers to a communal section *a* (no double counting) as travellers going in and out multiple communal sections to/from the communal section *a* cancel out.

Filters and redactions

Values of $cdr\_travellers\_out_{a,d}$ under 15 IDs (MSISDNs) have been redacted to missing for privacy purposes. Values of $est\_travellers\_out_{a,d}$ (travelling population) under 50 are also redacted to missing as they originate from a sample of travelling IDs that is too small to provide a robust estimate. Values are rounded to the nearest 10.

3.2. [Demo/experimental] Movements (daily)

3.2.1. Travellers

The estimated number of people who travelled between communal section *a* and section *b* on the same day.

Calculation

The estimated number of people travelling from *a* to *b* per day ($est\_travellers_{a,b,d}$), is derived from the count of Digicel IDs (pseudonymised MSISDNs of subscribers) who made a call from communal section *a* and made an immediately subsequent call from communal section *b* during the day *d* ($cdr\_travellers_{a,b,d}$).

This count is scaled by destination using a scaling factor x_b calculated for each communal section (see annex scaling factors). $est\_residents_{b,m=0}$ is the baseline resident population in section b and $med(cdr\_presence_{b,base})$ is the median count of presence in section b during the baseline period:

$$est\_travellers_{a,b,d} = cdr\_travellers_{a,b,d} * x_b * (est\_residents_{b,m=0} / med(cdr\_presence_{b,base}))$$

Where:

$est\_travellers_{a,b,d}$	is estimated travellers from communal section *a* to communal section *b* during the day *d*
$cdr\_travellers_{a,b,d}$	is the count of Digicel IDs (MSISDNs) that showed a call from the communal section *a* and an immediately subsequent call from the section a n *b*, during the day *d*
x_a	is the scaling factor for section *a* (see annex scaling factors)
$est\_residents_{b,m=0}$	is estimated residents for communal section *b* in the baseline month (*m=0*, Jan 2020)
$med(cdr\_presence_{b,base})$	is the median presence during the baseline period (August 2020 to September 2021)

NOTE: An ID travelling from A to B to C during the same day will be counted as a traveller from A to B and from B to C but not from A to C. Therefore, longer travels may not be captured by this estimate (if calls are made along the way).

Filters and redactions

Several filters are applied to travellers between communal section:

- Pairs of communal sections with more than 60% of available days missing were redacted to missing
- Values of less than estimated 50 travellers on the current day are redacted to missing as they originate from a sample of travelling subscribers that is too small to provide a robust estimate
- Values corresponding to fewer than 15 subscribers are redacted for privacy purposes

Values are rounded to the nearest 10.

3.2.2. Change in travellers

Difference in the number of people who travelled from communal section *a* to communal section *b* during the current day and the number of travellers from *a* to *b* during the first day of available data.

Calculation

The change in travellers is calculated as the difference in the number of travellers from communal section *a* to communal section *b* during the current day *d* with the number of travellers during the first day of available data (*'based0'*).

$est\_trvlr\_chg_{a,b,d} = est\_travellers_{a,b,d} - est\_travellers_{a,b,based0}$

Where:

$est\_travellers_{a,b,d}$	is estimated travellers from communal section *a* to communal section *b* during the day *d*
$est\_travellers_{a,b,based0}$	is estimated travellers from communal section *a* to communal section *b* during the day *based0* (first day of available data)

This is a demo/experimental estimate. It will soon be replaced by an estimate comparing the number of travellers on the current day, with the median number of travellers during a relevant baseline period (such as the preceding calendar year). This will help contextualise and interpret the value on a given day.

3.2.3. Relative change in travellers (%)

Percent change in the number of people who travelled from communal section *a* to communal section *b* during the current day and the number of travellers from *a* to *b* during the first day of available data.

Calculation

$est\_trvlr\_chg\_pct_{a,bd} = ((est\_travellers_{a,b,d} - est\_travellers_{a,b,based0}) / est\_travellers_{a,b,based0}) * 100$

Where:

$est\_travellers_{a,b,d}$	is estimated travellers from communal section *a* to communal section *b* during the day *d*
$est\_travellers_{a,b,based0}$	is estimated travellers from communal section *a* to communal section *b* during the day *based0* (first day of available data)

This is a demo/experimental estimate. It will soon be replaced by an estimate on the difference in the number of travellers on the selected day, and the median number of travellers during a relevant baseline period (such as the preceding calendar year). This will help contextualise and interpret the value on a given day.

Filters and redactions

Values are rounded to the nearest 2 decimal points.

3.2.4. Abnormality score

The abnormality score, which can also be termed 'outlier score', indicates how different the last estimated number of travellers from communal section *a* to communal section *b* is, compared to the median number of travellers from *a* to *b* estimated during the 365 preceding days (the baseline period), measured in median absolute deviations. It describes how unusual the last estimate in travellers from communal section *a* to communal section *b* is, compared to the number of travellers estimated during the 365 preceding days.

Calculation

The abnormality (outlier) score for the estimate in travellers from communal section *a* to communal section *b* is calculated as a modified z-score as follows:

1. Calculate the estimated travellers from a to b during the current day *d* ($est\_travellers_{a,b,d}$).
2. Compute the number of travellers for the previous 365 days, the baseline period, and calculate the median number during that period ($median\_ref(est\_travellers_{a,b})$), the median absolute deviation of travellers ($mad\_ref(est\_travellers_{a,b})$) and the mean absolute deviation of travellers ($meanad\_ref(est\_travellers_{a,b})$)
3. Calculate the modified z-score for the current day using this 365 day period as the baseline period for 'expected numbers of travellers'.

If the median absolute deviation (MAD) of the baseline time series is not equal to 0, the score is calculated as:

$$z\_est\_trvlr_{a,b,d} = (est\_travellers_{a,b,d} - median\_ref(est\_travellers_{a,b})) / (1.486 * mad\_ref(est\_travellers_{a,b}))$$

If the MAD is equal to 0, the mean absolute deviation (meanAD) is used instead:

$$z\_est\_trvlr_{a,d} = (est\_travellers_{a,b,d} - median\_ref(est\_travellers_{a,b})) / (1.253 * meanad\_ref(est\_travellers_{a,b}))$$

Where:

$est\_travellers_{a,b,d}$	is estimated travellers from communal section *a* to communal section *b* during the day *d*
$median\_ref(est\_travellers_{a,b})$	is the median number of estimated travellers from communal section *a* to section *b* during the baseline period
$mad\_ref(est\_travellers_{a,b})$	is the median absolute deviation of estimated travellers from communal section *a* to communal section *b* during the baseline period
$meanad\_ref(est\_travellers_{a,b})$	is the mean absolute deviation of estimated travellers from communal section *a* to communal section *b* during the baseline period.

The baseline period is defined as the 365 days prior to the current day (with at least 90 available values), not including the current day.

NOTE: As our data start in January 2020, there are fewer than 365 days of available prior baseline data, for some months in 2020. In this case, a shorter baseline is used for the total number of days available, with a limit of 90 days as the shortest reference period (for reliability purposes). Dates for which there are fewer than 90 days of available baseline data will not have an outlier score, and so **abnormality/outlier scores start in April 2020.**

Abnormality/Outlier scores are helpful in identifying unusual values, which can correspond to data issues or important real-world events that might be impacting people's travel to a communal section. Note that in case of unusual traveller values over several consecutive days, the score will continue to be large if the unusual period is short, but the score will decrease over time if the number of travellers remains at a similar level every day (i.e. the novel level becomes more usual).

A positive value for this estimate greater than 3 indicates an unusually large number of travellers (a statistical outlier); a value less than -3 indicates an unusually small number of travellers. Scores above 6 in absolute value are more likely to correspond to technical issues, particularly in the absence of known disrupting events (disruption of mobility and/or phone usage).

Filters and redactions

Values are rounded to the nearest 3 decimal points.

4. Annexes

Annex 1: Baseline Population: corrections

After comparison of IHSI's 2015 & 2020 estimates, 2003 census data, 2020 HRSL estimates, 2020 WorldPop estimates,Microsoft building footprints, and CDR-derived location counts for January 2020, further corrections were made to the initial baseline estimates to adjust **highly implausible values** (including where CDR counts exceeded other estimates). These changes were applied to for following 31 sections:

- 1re Section Morne Chandelle, Ouest
- 4e Section Bellevue la Montagne, Ouest
- 1re Section Morne à Bateau, Ouest
- 3e Section Petit Boucan, Ouest
- 1re Section des Varreux (HT0117-02), Ouest
- 3e Section Grande Rivère, Ouest
- 10e Section Fond d'Oie, Ouest
- 2e Section Delatre, Ouest
- 6e Section Trou Canari, Ouest
- 7e Section des Platons, Ouest
- 8e Section des Platons, Ouest
- 9e Section des Palmes, Ouest
- 10e Section des Palmes, Ouest
- 1re Section des Varreux (HT0131-01), Ouest
- 4e Section Petit Bois, Ouest
- 5e Section Petit Bois, Ouest
- 9e Section des Crochus, Ouest
- 4e Section des Crochus, Ouest
- 1re Section Galette Chambon, Ouest
- 1re Section Boucassin, Ouest
- 2e Section Boucassin, Ouest
- 3e Section Source Matelas, Ouest
- 1re Section Palma, Ouest
- 2e Section Petite Source, Ouest
- 3e Section Grande Source, Ouest
- 2e Section Chabotte, Nord
- 3e Section Roche Plate, Nord-Est

- 2e Section Fonfrède, Sud
- 5e Section Laroque, Sud
- 1re Section Basse Voldrogue, Grande'Anse
- 3e Section Jean Bellune, Grande'Anse

Annex 2: Scaling factors for presence and movement

The following scaling factors are applied to **scale the variation of presence and movement** from a count of subscribers (MSISDNs) to an estimate of people. However, unlike for our estimates of residents and relocations, we do not currently have survey data enabling us to measure - and correct for - differences in this type of mobility (daily presence and movements) between different groups of people (Digicel phone users, other mobile phone users, and non phone users). As a result, these scaling factors do not correct for representation biases, they only capture the proportion of Digicel subscribers within each communal section.

The factor x_a is used in the equations for

- Presence
- Travellers
- Total inflows of travellers
- Total outflows of travellers

Calculation

Calculations are shown below for the presence estimates, but the same methodology is applied for all estimates derived from presence and movements (travellers).

The x_a factor is used, for example, in the presence equation, to scale a relative variation in subscribers present to a variation in people present in a communal section *a* (e.g. if scaling change by 0.5, a 20% change in subscriber present would correspond to a 10% change in people present):

$$est\_presence_{a,d} = est\_residents_{a,m=0} + x_a * (cdr\_presence_{a,d} - median(cdr\_presence_{a,base}))$$

With:

x_a = phone_user_rate$_k$ * $est\_residents_{a,m=0} / median(cdr\_presence_{a,base})$,
IF $est\_residents_{a,m=0} / med(cdr\_presence_{a,base}) <= 25$

OR

x_a = phone_user_rate$_k$ * median_*k* ($est\_residents_{a,m=0} / median(cdr\_presence_{a,base})$),
IF $est\_residents_{a,m=0} / med(cdr\_presence_{a,base}) > 25$

Where:

$est\_presence_{a,d}$	is the estimated presence of people in communal section *a* on day *d*
$est\_residents_{a,m=0}$	is the existing population estimate derived from IHSI's 2015 and 2020 population estimates
$cdr\_presence_{a,d}$	is the count of Digicel subscribers (MSISDNs) whose call(s) were routed by a cell tower in communal section *a* on day *d*
$median(cdr\_presence_{a,base})$	is the median presence count over days during the baseline period (August 2020 to September 2021)

$phone\_user\_rate_k$ is the mobile phone user rate in department *k* (static over time).

Please see the equation below.

median_*k* (*est_residents*$_{a,m=0}$ / *median(cdr_presence*$_{a,base}$) is the median over all communal sections *a* within the department *k*, of the ratio of the population estimate and the median number of present subscribers in each section.

Note that the metropolitan area of Port-au-Prince is considered a separate department for this purpose.

The $phone\_user\_rate_k$ for each unit *k* (each department and the Port-au-Prince metropolitan area) is expressed as:

$$phone\_user\_rate_k = (D_k + N_k)/P_k$$

Where:

D_k is the number of Digicel subscribers in the department k
N_k is the number of Natcom subscribers in the department k
P_k is the HRSL population estimate in the department k

The following data sources and parameters are used to calculate the phone user penetration rate:

- HRSL 2020 population layer where pixel values are adjusted uniformly so that the total population equals the 2021 IHSI national population total
- Digicel's June 2022 market share report

Description of scaling factors for presence and movement

In communal sections where the population is lower than 25 times the median number of present subscribers in baseline, the scaling factor is equal to the mobile phone user rate multiplied by the ratio of the population over the median number of present subscribers in base, for the communal section.

On the opposite, in communal sections where the population is larger than 25 times the median number of present subscribers in baseline, instead of multiplying the mobile phone user rate by the population to subscriber ratio for the communal section, we use the median ratio for the department. This means that for those communal sections with very high scaling factors, we replace the scaling factor of individual communal sections by the median value for the department, thus suppressing the scale up of the CDR observations, since they have higher uncertainty (smaller subscriber to people ratio).

Contact us

For queries or information about the Haiti Mobility Data Platform, the methods presented in this document or on mobile data analytics in general, please contact us at **haiti.mobility-dashboard@flowminder.org**

General enquiries info@flowminder.org

Website www.flowminder.org

Twitter @Flowminder

LinkedIn www.linkedin.com/company/flowminder

FlowGeek www.flowgeek.org