FLOWMINDER.ORG

Haiti Mobility Data Platform

Description of indicators

Last updated: 18 July 2023

Description of indicators

List of content

General overview: About the indicators

All indicators made available on this platform have been derived from pseudonymised Call Detail Records (CDRs) from Digicel Haiti, and corrected for representation biases using survey data (except for the experimental indicators).

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 subscriber ID, and the cell tower ID which routed the event. Based on this information and the location of the cell towers, we can estimate how people move within a country. Summing and analysing mobility across 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 our FlowGeek site.

Understanding the data

Residents and relocations indicator categories

Turning phone usage data (CDR data) into population distribution and mobility indicators

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, 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 volume of people at a certain place and time.

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

Often, the mobility indicators extracted from CDR only represent the subscribers of the participating MNO. Additionally, the frequency with which subscribers use their mobile devices also affect the accuracy of CDR-derived estimates, and there are risks that the data actually reflect phone usage changes rather than mobility changes (i.e. increased phone usage may result in the appearance of increased mobility).

The indicators available on this platform are the results of years of research and method development, and have been corrected to:

- **Make the indicators representative of the whole population:** using survey data and existing population estimates, we scaled our indicators so that they do not represent numbers of Digicel subscribers. Our estimates represent the population as a whole, regardless of their operator or whether they own 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 presented on this platform are robust estimated numbers of the overall population in Haiti, and not just Digicel subscribers counts. **They have been both scaled and adjusted to be representative of the population as a whole as well as to measure actual mobility, reducing sensitivity to varying phone usage.** To do so, using advanced statistical models, we combined Digicel CDR data and mobility information extraction methods with recent field survey data and existing population estimates. We used population estimates from IHSI, WorldPop, Meta/CIESIN as well as population growth estimates from the UN.

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

These two categories of indicators 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 territory.

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

EXPERIMENTAL indicators | Presence and Movement indicator categories

Presence and Movements indicators are at present **experimental**; we recommend that you use these indicators 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 pandemic), in which case, even though experimental, these indicators may be useful.

'Presence' indicators attempt to inform on the number of people present in each communal section each day and 'Movements' indicators attempt to capture the number of people travelling between each communal section each day.

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

- **Representativity bias**: Experimental presence and travel indicators are more reflective of the distribution and mobility of the Digicel subscribers than of the general population
- **Phone usage influence:** Experimental indicators 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 experimental indicators 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. 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.

Experimental indicators currently provided may for example overestimate mobility in urban areas while underestimate it in rural areas.

2. Phone usage influence

Experimental indicators are derived from counts of Digicel's 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 indicator 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 indicators 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 experimental indicators, which should not be interpreted as changes in presence or travel.

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

To understand how we produce indicators 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 results produced by the Flowminder Foundation are aggregated results (for example, subscriber density in a given municipality), which means that they do not contain any information about individual subscribers. This data is fully anonymised. This approach complies with the European Union's General Data Protection Regulation (EU GDPR 2016/679). Data is processed on a server installed behind Digicel Haiti's firewall in Haiti, and no personal data leaves Digicel Haiti's premises.

Description of indicators

Introduction

The Haiti Mobility Platform (haiti.mobility-dashboard.org) is a privacy-secure web-platform providing mobility and population estimates and indicators based on pseudonymised Call Detail Records (CDRs) from Digicel Haiti, survey data and further data sources, to approved third parties. These estimates and related indicators represent the distributions and movements of the Haiti population, and their dynamic changes over time. The platform allows users to visualise, interact and download the data for further analysis. To understand how Flowminder produces indicators from pseudonymised CDR data, please click here. You will be redirected to FlowGeek, our online knowledge centre on CDR data analytics.

This document is aimed at a technical/scientific audience and presents the estimates and indicators available on the Haiti Mobility Platform. In this document, you will learn more about the available indicators, what they mean and measure, and how we calculated them.

General remarks

- All estimates and indicators are calculated **per communal section** (residents, presence) or **pair of communal sections** (relocations, movements). Communal sections correspond to administrative units level 3 in Haiti.
- All estimates described below are **point estimates** (not interval estimates). A dataset also containing related confidence intervals of Residents and Relocations estimates is available upon request.
- We use the term **'subscriber'** to refer to pseudonymised MSISDNs (hashed phone numbers). This means that two 'subscribers' (phone numbers) may be the same individual, if that person uses more than one phone number, or that one 'subscriber' may represent several individuals, if a phone number is actively used by several people.
- **Baseline periods or months** differ by indicator, see the respective descriptions.

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

Monthly population estimates (residents) and internal relocations:

- **Residents**
- **Relocations**

Experimental daily presence and movements (travel) estimates:

- **[Experimental] Presence**
- **[Experimental] Movements**

Monthly population estimates (residents) and internal relocations

Our monthly population estimates (residents) are derived from Digicel CDR data, survey data and existing population estimates for January 2020. **CDR data are used to infer subscribers' home locations each month and then changes in home location to estimate internal mobility (relocations between communal sections), then relocations and existing population estimates are used to estimate the number of residents of a communal section in a given month**. Therefore our estimates are based on counts of moving subscribers (subscribers who appear to relocate), and not on counts of residing subscribers (who appear to be resident of a given location), to reduce the effect of phone usage (and appearing and disappearing subscribers) in our estimates.

This documentation describes each indicator as it is given 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' categories we first need to compute 'relocations'.

To guide the reader of this documentation, we provided below a short overview of the computation of the Relocations and Residents indicators:

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

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

- We detect Digicel subscribers' home location each month (based on their pseudonymised phone number) as the communal section that contains the cell phone towers near which they spent the majority of the month. Then we detect relocations as a change in subscribers' home location.
- We adjust and scale relocations for pairs of communal sections in each direction using the MSNA 2022 survey. This is our method to correct for representativity biases, please see the section on the estimation of relocations for more information.
- We sum all adjusted and scaled relocations into each communal section each month (Total incoming) and subtract all adjusted and scaled relocations out (Total outgoing) to compute the net adjusted and scaled relocations into each communal section (Incoming minus outgoing). This corresponds to the difference in residents due to monthly mobility only (those who moved in minus those who moved out).

We then calculate residents' estimates from:

- Baseline population estimates derived from IHSI's 2015 and 2020 population estimates (with corrections for some communal sections to be detailed in our next documentation update) for January 2020
- We then add the number of net adjusted relocations into each communal section from January 2020 to February 2020 (Incoming minus outgoing), computed as described above from Digicel's CDRs and corrected with survey data, to account for changes in population

estimates due to internal mobility. Only the mobility detected in CDRs is used to estimate residents and not the count of subscribers' home locations. This is our method to reduce the influence of changes in phone use behaviour in our mobility and population estimates.

- We multiply the sum (baseline estimate + net relocations) by estimates of average monthly population change rates from UN data to account for other demographic components (births, deaths, immigration and emigration).
- We repeat this process each month up to the current month (or the month of interest), to estimate residents from the existing baseline population (January 2020), internal mobility and other population changes that occurred since then.

The next section on the estimation of residents provides more details on our methodology.

Residents

Residents

The estimated number of residents per communal section for the current month.

Calculation

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

The calculation of residents' estimates is based on:

- changes in Digicel subscribers' home communal section (subscribers' relocations)
- adjustment and scaling factors derived from the MSNA 2022 survey to correct for biases in the count of relocations between any two communal sections (Relocations)
- Estimates of the total incoming and total outgoing relocations for each section, resulting in estimated net relocations (incoming minus outgoing)
- estimates of average monthly population changerates from UN data
- baseline population estimates derived from IHSI's 2015 and 2020 population estimates

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 to that communal section between the two months (months *m-1* and *m*), multiplied by a communal section-specific rate of population change. The start month is January 2020 (*m=0*) for which the estimate of residents is based on existing population estimates.

The estimate of residents can be expressed as a system of recursive equations:

$$est\_residents_{a,m} = (est\_residents_{a,m-1} + est\_netreloc_{a,m-1,m}) * changerate_{a}$$

and

$$est\_residents_{a,m=0} = est\_residents_{a,base}$$

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*
$est\_netreloc_{a,m-1,m}$	is the estimated total net relocations for communal section *a* between months *m-1* and *m*
$changerate_{a,}$	is the estimated average monthly change rate for communal section *a*
$est\_residents_{a,m=0}$	is the population estimate for communal section *a* for *m=0* (January 2020), which is the baseline population estimate ($est\_residents_{a,base}$)

The net relocations estimate for communal section *a* between months *m-1* and *m* ($est\_netreloc_{a,m-1,m}$) is the sum of all estimated relocations to that communal section (total incoming: $est\_reloc\_to_{a,m-1,m}$) minus the sum of all estimated relocations from that communal section (total outgoing: $est\_reloc\_from_{a,m-1,m}$):

$est\_netreloc_{a,m-1,m} = est\_reloc\_to_{a,m-1,m} - est\_reloc\_from_{a,m-1,m}$

For the calculation of total estimated relocations to and from the communal section, see **Total incoming** and **Total outgoing**.

Relocations estimates, in turn, are based on CDR-derived relocations, i.e. detected changes of home locations of 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 a subscriber over a calendar month. If there is no call on at least 1 day in 3 separate weeks in the 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 subscriber, 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. 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* ($est\_residents_{a,m=n}$):

$$
\begin{aligned}
est\_residents_{a,m=n} = {} & est\_residents_{a,m=0} * changerate_a^{\,n} + \\
& est\_netreloc_{a,m=0,m=1} * changerate_a^{\,n} + \\
& est\_netreloc_{a,m=1,m=2} * changerate_a^{\,n-1} + \\
& \ldots + \\
& est\_netreloc_{a,m-1,m} * changerate_a^{\,n-m+1} + \\
& \ldots + \\
& est\_netreloc_{a,m=n-1,m=n} * changerate_a
\end{aligned}
$$

Which can be expressed concisely as:

$$est\_residents_{a,m=n} = est\_residents_{a,m=0} * changerate_a^{\,n} + \sum_{m=1}^{n} (est\_netreloc_{a,m-1,m} * changerate_a^{\,n-m+1})$$

NOTE: We use the term 'subscriber' to refer to pseudonymised MSISDNs (hashed phone numbers). This means that two 'subscribers' (phone numbers) may be the same individual, if that person uses more than one phone number, or one 'subscriber' may represent several individuals, if a phone number is actively used by several people.

For adjustments from pseudonymised MSISDNs to individuals, see Relocations.

Filters and redactions

Residents estimates are not provided for all communal sections for the following reasons:

- Some communal sections do not have Digicel cell tower coverage
- Some communal sections have small counts of CDR-derived home location counts

Of the 570 communal sections in Haiti, the estimates of this release cover 370 of them. Among all 570 communal sections, 167 did not have any CDR-derived resident aggregates for any month of

the reporting period. Of the remaining 403 communal sections with resident aggregates, the ones where the median number of CDR-derived home location counts across all available months was smaller than 200 were dropped (33 communal sections).

Values are rounded to the nearest 100.

Residents per square km

Residents per km^2 (per month) is the number of estimated residents per communal section for the current month, divided by the area of that communal section, giving the spatial density of residents.

Calculation

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

Where:

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

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 that, as we account for population change (births and deaths, immigration and emigration), the change in residents compared to the baseline estimate incorporates both the cumulative sum of net relocations and the monthly population change rate.

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 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)

Abnormality score

The abnormality score indicates how different the last monthly change of residents per communal section (between the current month and the previous month) is, compared to the median monthly change of residents estimated during the 12 preceding months (the baseline period), measured in median absolute deviations. It describes how unusual the last monthly change in the number of residents in a communal section is, compared to the monthly changes estimated during the 12 preceding months.

Calculation

The abnormality 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*, and the median monthly change in residents during the baseline period, and standardising this difference by the median absolute deviation of monthly resident change during the baseline period.

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

$$abn\_est\_res\_chg_{a,m-1,m} = ((est\_residents_{a,m} - est\_residents_{a,m-1}) - median\_chg\_base(est\_residents_a)) / (1.486 * mad\_chg\_base(est\_residents_a))$$

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

$$abn\_est\_res\_chg_{a,m-1,m} = ((est\_residents_{a,m} - est\_residents_{a,m-1}) - median\_chg\_base(est\_residents_a)) / (1.253 * meanad\_chg\_base(est\_residents_a))$$

Where:

$abn\_est\_res\_chg_{a,m-1,m}$	is the abnormality 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*
$median\_chg\_base(est\_residents_a)$	is the median estimated change of residents per month for

communal section *a* during the baseline period

mad_chg_base(est_residents$_a$) is the median absolute deviation of estimated residents per month for communal section *a* during the baseline period

meanad_chg_base(est_residents$_a$) is the mean absolute deviation of estimated residents per month for communal section *a* during the baseline period

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

Abnormality 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 score will only be large on the first month and the last month of the unusual period, i.e. **indicating transitions between normal and unusual time periods - and not indicating abnormal resident values - only abnormal change).**

For this indicator, a positive value greater than 3 indicates an abnormal increase in the number of residents in the communal section (a statistical outlier); a value less than -3 indicates an abnormal decrease in residents. Values between 3 and -3 are within the bounds of normal variation based on the reference period. Abnormality 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 score if less than 3 data points (months) are available. Values are rounded to 3 decimal points.

Total incoming

The estimated number of people who relocated into (i.e. moved into) a communal section *a* (from all other communal sections) between the previous and the current month.

This sum is the total of incoming relocations for the current month.

Calculation

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

$$est\_reloc\_to_{a,m\text{-}1,m} = \sum_{b=1}^{k} est\_reloc_{b,a,m\text{-}1,m}$$

Where:

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

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

Filters and redactions

Values are rounded to the nearest 10.

Total outgoing

The estimated number of people who relocated (moved home) out of a communal section *a* (to all other communal sections) between the previous and the current month.

This sum is the total of outgoing relocations for the current month.

Calculation

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

$$est\_reloc\_from_{a,m\text{-}1,m} = \sum_{b=1}^{k} est\_reloc_{a,b,m\text{-}1,m}$$

Where:

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

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

Filters and redactions

Values are rounded to the nearest 10.

Incoming minus outgoing

The difference between the number of people relocating (moving) into (total incoming) and out of (total outgoing) 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, due to internal migration. This indicator can also be referred to as 'net relocations'.

Calculation

The incoming minus outgoing indicator (net estimated relocations) for communal section *a* between months *m-1* and *m* is the sum of all estimated relocations to that communal section (incoming) minus the sum of all estimated relocations from that communal section:

$est\_netreloc_{a,m-1,m} = est\_reloc\_to_{a,m-1,m} - est\_reloc\_from_{a,m-1,m}$

Where:

$est\_netreloc_{a,m-1,m}$	is estimated total net relocations to/from communal section a between months *m-1* and *m*
$est\_reloc\_to_{a,m1-,m,}$	is estimated total relocations to communal section a between months *m-1* and *m*
$est\_reloc\_from_{a,m-1,m}$	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 this indicator is not equivalent to the change in residents between two consecutive months. As per the recursive equation given in section Residents and below, the sum of estimated residents in the previous month and the net relocations is multiplied by the population change rate to obtain the estimated residents for the current month:

$est\_residents_{a,m} = (est\_residents_{a,m-1} + est\_netreloc_{a,m-1,m}) * changerate_a$

Therefore the change in residents from one month to the next is derived from both the net relocations (internal mobility) and the population change rate.

Filters and redactions

Values are rounded to the nearest 10.

Relocations

Relocations

The estimated number of persons relocating (i.e. changing their home location) 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 ($cdr\_reloc_{a,b,m-1,m}$), the number of subscribers 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 a subscriber over a calendar month. For each subscriber, 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 ($cdr\_reloc_{a,b,m-1,m}$) from communal section *a* to communal section *b* between months *m-1* and *m* are multiplied by survey-derived adjustment factors ($adj_{a,b}$) and scaling factors ($sf_{a,b}$):

$$est\_reloc_{a,b,m-1,m} = cdr\_reloc_{a,b,m-1,m} * adj_{a,b} * sf_{a,b}$$

Where:

$est\_reloc_{a,b,m-1,m}$	is estimated relocations from communal section *a* to communal section *b* between months *m-1* and *m*
$cdr\_reloc_{a,b,m-1,m}$	is CDR-derived relocations from communal section *a* to communal section *b* between months *m-1* and *m*
$adj_{a,b}$	is a survey-derived adjustment factor for relocations from communal section *a* to communal section *b*
$sf_{a,b}$	is a survey-derived scaling factor for relocations from communal section *a* to communal section *b*

For Haiti, the adjustment factors and scaling factors for relocations can currently only be estimated from cross-sectional survey data (the MSNA 2022 survey), but not their variation over time. These time-invariant parameters have to be used until other or new data (such as new survey data, census data or regulator data) become available to estimate their variation over time.

NOTE: We use the term 'subscriber' to refer to pseudonymised MSISDNs (hashed phone numbers). This means that two 'subscribers' (phone numbers) may be the same individual, if that person uses more than one phone number, or one 'subscriber' may represent several individuals, if a phone number is actively used by several people.

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

Filters and redactions

Only relocation corridors (relocations from communal section *a* to *b*) with a median time-series value ($cdr\_reloc_{a,b,m-1,m}$) equal to or above 50 are included in the data. Corridors with median values below 50 have been excluded from the data, as estimated relocations below a value of 50 may have low reliability.

Only values based on $cdr\_reloc_{a,b,m-1,m}$ equal to or above 15 are shown, smaller values are redacted to missing, i.e. at least 15 subscribers need to have been observed to relocate for the derived estimate to be shown.

Values are rounded to the nearest 10.

Reliability

Estimated relocations below a value of 50 may have low reliability.

Change in relocations

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

NOTE: This indicator will be updated to use a longer baseline in our next release.

Calculation

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

Where:

$est\_reloc\_chg_{a,b,m,base}$	is the absolute change in estimated relocations from communal section *a* to communal section *b* for the months *m-1* and *m* in comparison to the baseline months *basem0* and *basem1*
$est\_reloc_{a,b,m-1,m}$	is estimated relocations from communal section *a* to communal section *b* for the months *m-1* and *m*
$est\_reloc_{a,b,basem0,basem1}$	is estimated relocations from communal section *a* to *b* for the baseline months *basem0* and *basem1*, the first available relocations' estimate

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

Filters and redactions

Values are rounded to the nearest 10.

Reliability

Estimated relocations below a value of 50 and related changes in relocations may have low reliability.

Relative change in relocations (%)

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

NOTE: This indicator will be updated to use a longer baseline in our next release.

Calculation

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

Where:

$est\_reloc\_chg\_pct_{a,b,m,base}$	is the relative change in estimated relocations between communal sections a and b for the months m-1 and m in comparison to the baseline months basem0 and basem1
$est\_reloc_{a,b,m-1,m}$	is estimated relocations between communal sections a and b for the months m-1 and m
$est\_reloc_{a,b,basem0,basem1}$	is estimated relocations between communal sections a and b for the baseline months basem0 and basem1, the first available relocations' estimate

Reliability

Estimated relocations below a value of 50 and related relative changes in relocations may have low reliability.

Abnormality score

The abnormality score indicates how different the last monthly change of relocations (from communal section *a* to a communal section *b*) is, compared to the median monthly change of relocations estimated during the 12 preceding months (baseline period), measured in median absolute deviations. It describes how unusual the last monthly change in the number of relocations from a communal section *a* to a communal section *b* is, compared to the monthly changes estimated during the 12 preceding months.

Calculation

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

$$abn\_est\_reloc_{a,b,m-1,m} = ((est\_reloc_{a,b,m-1,m} - est\_reloc_{a,b,m-2,m-1}) - median\_chg\_base(est\_reloc_{a,b})) / (1.486 * mad\_chg\_base(est\_reloc_{a,b}))$$

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

$$abn\_est\_reloc_{a,b,m-1,m} = ((est\_reloc_{a,b,m-1,m} - est\_reloc_{a,b,m-2,m-1}) - median\_chg\_base(est\_reloc_{a,b})) / (1.253 * meanad\_chg\_base(est\_reloc_{a,b}))$$

Where:

$abn\_est\_reloc_{a,b,m-1,m}$	is the abnormality score for the change in estimated relocations from communal section *a* to communal section *b*, between months *m-1* and *m*
$est\_reloc_{a,b,m-1,m}$	is estimated relocations from communal section a to communal section b, between months *m-1* and *m*
$median\_chg\_base(est\_reloc_{a,b})$	is the median monthly change of estimated relocations from communal section *a* to communal section *b* during the baseline period

mad_chg_base(est_reloc$_{a,b}$)	is the median absolute deviation of monthly change of estimated relocations from communal section *a* to communal section *b* during the baseline period
meanad_chg_base(est_reloc$_{a,b}$)	is the mean absolute deviation of monthly change of estimated relocations from communal section *a* to communal section *b* during the baseline period

The baseline 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 indicator greater than 3 indicates an abnormal increase in the number of relocations (a statistical outlier); a value less than -3 indicates an abnormal decrease in relocations. Abnormality 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 score if less than 3 data points (months) are available. Values are rounded to 3 decimal points.

Experimental daily presence and movement (travel) indicators

Unlike our Residents and Relocation indicators, our Presence and Movement (travel) indicators are at present **experimental**; we recommend that you use these indicators for training and trialling purposes only.

'Presence' indicators attempt to inform on the number of people present in each communal section each day and 'Movements' indicators attempt to capture the number of people travelling between each communal section each day.

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

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

However, our experimental estimates do not provide counts of Digicel subscribers - they have been scaled to a number of people, using the X factor method but not in a representative manner. They do not use the same robust adjustment and scaling method as our Residents and Relocations estimates, as we have not yet collected the necessary survey data (this data collection exercise is however planned in the months to come).

Additionally, the extraction of mobility information from the CDRs is not robust to changes in phone usage as for our Residents and Relocations 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 experimental Presence and Movement indicators for trialling purposes only 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).

[Experimental] Presence

Presence

The estimated number of people present in each communal section on the current day. It includes residents present in their own communal section as well as travellers, and people can be present in several communal sections during the same day.

Calculation

Presence of people in a communal section *a,* per day, is derived from the count of Digicel subscribers whose call(s) were routed by a cell tower in this communal section on this 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 ($median(cdr\_presence_{a,base})$), 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 (however not in a representative manner) by first multiplying it by a factor x_a, where x_a is a scaling factor calculated for each communal section *a* as described below in Regional scaling factors x. 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.*

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

Where:

x_a	is the scaling factor for communal section *a* (described in section Regional scaling factors X)
$est\_presence_{a,d}$	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 during the baseline period (August 2020 to September 2021)

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

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

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.

NOTE: We use the term 'subscriber' to refer to pseudonymised MSISDNs (hashed phone numbers). This means that two 'subscribers' (phone numbers) may be the same individual, if that person uses more than one phone number, or one 'subscriber' may represent several individuals, if a phone number is actively used by several people.

Filters and redactions

Values based on $cdr\_presence_{a,d}$ under 15 subscribers (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 MSISDNs, or any communal section with $median(cdr\_presence_{a,base})$ below 200).

Values of $est\_presence_{a,d}$ (after scaling) are rounded to the nearest 100 as we cannot provide a higher precision for this estimate.

Presence per km^2

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

Calculation

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

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

Where:

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

Filters and redactions

Values are rounded to the nearest 10.

Change in presence

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

This is a trial metric. It will soon be updated with a metric on the change between presence on the current day and the median presence during a baseline period (such as the preceding calendar year).

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 trial metric. It will soon be replaced by a metric 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.

Relative change in presence (%)

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

This is a trial metric. It will soon be updated with a metric on the percent change between presence on the current day and the median presence during a baseline period (such as the preceding calendar year).

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 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 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 trial metric. It will soon be replaced by a metric 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.

Abnormality score

The abnormality score indicates how different the last daily change in presence per communal section (between the current day and the previous day) is, compared to the median daily change of presence estimated during the 365 preceding days (the baseline period), measured in median absolute deviations. It describes how unusual the last daily change in presence in a communal section is, compared to the daily changes estimated during the 365 preceding days.

Calculation

The abnormality score for the daily change in presence per communal section is calculated as a modified z-score as follows:

1. Calculate the difference in estimated presence between the current day *d* and the previous day *d-1* ($est\_presence_{a,d}$ - $est\_presence_{a,d-1}$). This gives us the 'change' for that day.
2. Compute this day-to-day change for the previous 365 days, the baseline period, and calculate the median daily change during that period ($median\_chg\_base(est\_presence_a)$), the median absolute deviation of change ($mad\_chg\_base(est\_presence_a)$) and the mean absolute deviation of change ($meanad\_chg\_base(est\_presence_a)$).
3. Calculate the modified z-score for the current day using this 365 day period as the baseline period for 'expected changes in presence'.

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

$$abn\_est\_pres\_chg_{a,d} = ((est\_presence_{a,d} - est\_presence_{a,d-1}) - median\_chg\_base(est\_presence_a)) / (1.486 * mad\_chg\_base(est\_presence_a))$$

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

$$abn\_est\_pres\_chg_{a,d} = ((est\_presence_{a,d} - est\_presence_{a,d-1}) - median\_chg\_base(est\_presence_a)) / (1.253 * meanad\_chg\_base(est\_presence_a))$$

Where:

$est\_presence_{a,d}$	is the estimated presence of people in communal section *a* on day *d*
$est\_presence_{a,d-1}$	is the estimated presence of people in communal section *a* on day *d-1*
$median\_chg\_base(est\_presence_a)$	is the median daily change of estimated presence in

	communal section *a* during the baseline period
mad_chg_base(est_presence$_a$)	is the median absolute deviation of daily change of estimated presence in communal section *a* during the baseline period
meanad_chg_base(est_presence$_a$)	is the mean absolute deviation of daily change 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 will not have an abnormality score, and so **abnormality scores start in April 2020.**

Abnormality scores are helpful in identifying unusual changes, which can correspond to data issues or important real-world events that might be impacting people's presence in a communal section. However, note that in case of unusual presence values over several consecutive days, the abnormality score will only be large on the first day and the last day of the unusual period, i.e. **indicating transitions between normal and unusual time periods.**

A positive value for this indicator greater than 3 indicates an abnormal increase in presence (a statistical outlier); a value less than -3 indicates an abnormal decrease in presence. Abnormality 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).

Total incoming

The estimated number of incoming travellers to a communal section *a,* from any other communal section, during the current day.

Calculation

The estimated number of travellers to a communal section *a,* per day, is derived from the count of Digicel subscribers (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*$_{a,m=0}$ / *median(cdr_presence*$_{a,base}$*)*), where x_a is a regional scaling factor calculated for each communal section as described below in Regional scaling factors X, *est_residents*$_{a,m=0}$ is the existing estimate of the population for each communal section *a*, and *median(cdr_presence*$_{a,base}$*)* is the median count of subscribers' presence 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$_{a,d}$ = x_a * (*est_residents*$_{a,m=0}$ / *median(cdr_presence*$_{a,base}$*)) * cdr_travellers_in*$_{a,d}$

Where:

x_a	is the scaling factor for communal section *a* (described in section Regional scaling factors X)
$est\_residents_{a,m=0}$	is the existing population estimate derived from IHSI's 2015 and 2020 population estimates
$median(cdr\_presence_{a,base})$	is the median estimated presence during the baseline period (August 2020 to September 2021)
$cdr\_travellers\_in_{a,d}$	is the count of Digicel subscribers (MSISDNs) who made a call from a communal section section that is not *a* and made a subsequent call from the 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 experimental indicator which we are working to resolve.

Filters and redactions

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

Total outgoing

The estimated number of outgoing travellers 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 subscribers (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 (however not in a representative manner) using the adjustment term $x_a*(est\_residents_{a,m=0}/median(cdr\_presence_{a,base}))$, where X_a is a regional scaling factor calculated for each communal section as described below in Regional scaling factors X, $est\_residents_{a,m=0}$ is the existing estimate of the population for each communal section *a*, and $median(cdr\_presence_{a,base})$ is the median count of subscribers' presence 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.

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

Where:

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

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 experimental indicator which we are working to resolve.

NOTE: The difference 'total incoming - total outgoing' 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 $cdr\_travellers\_in_{a,d}$ under 15 subscribers (MSISDNs) have been redacted out for privacy purposes. Values of $est\_travellers\_in_{a,d}$ (travelling subscribers after scaling) under 50 are also redacted out as they originate from a sample of travelling subscribers that is too small to provide a robust estimate. Values are rounded to the nearest 10.

[Experimental:] Movements

Travellers

An estimated number of people who visited communal section *a* then communal section *b* within 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 subscribers (MSISDNs) who made a call from the communal section *a* and made an immediately subsequent call from the communal section *b*, during the day *d* ($cdr\_travellers_{a,b,d}$).

This is scaled (however not in a representative manner) by destination using the adjustment term ($x_b * (est\_residents_{b,m=0}/median(cdr\_presence_{b,base})$), where x_b is a regional scaling factor calculated for each communal section as described below in Regional scaling factors X, $est\_residents_{b,m=0}$ is the static population in destination and $median(cdr\_presence_{b,base})$ is median counts of subscribers' presence in destination during the baseline period:

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

Where:

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

NOTE: A subscriber travelling (and calling) 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 long travels may not be captured by this estimate (if calls are made along the way).

NOTE: We use the term 'subscriber' to refer to pseudonymised MSISDNs (hashed phone numbers). This means that two 'subscribers' (phone numbers) may be the same individual, if that person uses more than one phone number, or one 'subscriber' may represent several individuals, if a phone number is actively used by several people.

Filters and redactions

Several filters are applied to travellers per communal section:

- Pairs of communal sections with more than 60% of available days missing were redacted out
- Values of less than 50 travellers (after scaling) on the current day are redacted out 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.

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.

This is a trial metric - it will soon be updated with a metric on the change between travellers during the current day and the median travellers during a baseline period (such as the preceding calendar year).

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 trial metric. It will soon be replaced by a metric 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.

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.

This is a trial metric - it will soon be updated with a metric on the change between travellers during the current day and the median travellers during a baseline period (such as the preceding calendar year).

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 trial metric. It will soon be replaced by a metric 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.

Abnormality score

The abnormality score indicates how different the last daily change in travellers from communal section *a* to communal section *b* (comparing number of travellers during the current day to numbers of travellers during the previous day) is, compared to the median daily change 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 daily change in travellers from communal section *a* to communal section *b* is, compared to the daily changes estimated during the 365 preceding days.

Calculation

The abnormality score for the daily change in travellers from communal section *a* to communal section *b* is calculated as a modified z-score as follows:

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

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

$$abn\_est\_trvlr\_chg_{a,b,d} = ((est\_travellers_{a,b,d} - est\_travellers_{a,b,d-1}) - median\_chg\_base(est\_travellers_{a,b})) / (1.486 * mad\_chg\_base(est\_travellers_{a,b}))$$

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

$$abn\_est\_trvlr\_chg_{a,d} = ((est\_travellers_{a,b,d} - est\_travellers_{a,b,d-1}) - median\_chg\_base(est\_travellers_{a,b})) / (1.253 * meanad\_chg\_base(est\_travellers_{a,b}))$$

Where:

$est\_travellers_{a,b,d}$	is estimated travellers from communal section *a* to communal section *b* during the day *d*
$est\_travellers_{a,b,d-1}$	is estimated travellers from communal section *a* to communal section *b* during the day *d-1*
$median\_chg\_base(est\_travellers_{a,b})$	is the median daily change of estimated travellers from communal section *a* to communal section *b* during the baseline period
$mad\_chg\_base(est\_travellers_{a,b})$	is the median absolute deviation of daily change of estimated travellers from communal section *a* to communal section *b* during the baseline period

$meanad\_chg\_base(est\_travellers_{a,b})$	is the mean absolute deviation of daily change 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 abnormality score, and so **abnormality scores start in April 2020.**

Abnormality scores are helpful in identifying unusual changes, which can correspond to data issues or important real-world events that might be impacting people's presence in a communal section. However, note that in case of unusual traveller values over several consecutive days, the abnormality score will only be large on the first day and the last day of the unusual period, i.e. **indicating transitions between normal and unusual time periods.**

A positive value for this indicator greater than 3 indicates an abnormal increase in the number of travellers (a statistical outlier); a value less than -3 indicates an abnormal decrease in travellers. Abnormality 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.

Regional scaling factors X

The following regional 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.

Calculation

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

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

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 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 during the baseline period (August 2020 to September 2021)

The x_a factor is computed as:

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

THEN $x_a = phone\_user\_rate_k$

ELSEIF $est\_residents_{a,m=0} / median(cdr\_presence_{a,base}) > 25$

THEN $x_a = (median(cdr\_presence_{a,base}) / est\_residents_{a,m=0}) * median ((est\_residents_{a,m=0} * phone\_user\_rate_k) / median(cdr\_presence_{a,study}))$

Where:

$phone\_user\_rate_k$	is the mobile phone user rate in department *k* (static over time)
$median(cdr\_presence_{a,study}))$	is the the median presence count during the study period (January 2020 to August 2022)

We suppress very large scaling factors (over 25) as the communal sections where this occurs have higher uncertainty (fewer subscribers).

x = 1, corresponds to the assumption that the variation of the indicator calculated from the CDR subset is representative of the variation in the general population (i.e. a 20% in subscribers translates to a 20% change in population).

If *X* is the phone-using penetration rate, then the assumption is that the variation of the indicator calculated from the CDR subset is representative of the variation in the phone-using population.

The following data sources and parameters are used to calculate *X* for each section:

- HRSL 2020 population layer where pixel values are adjusted uniformly so that the total population equals the 2021 IHSI national population total
- Department (admin 1 units + Port-au-Prince metropolitan section) phone user penetration rate calculated from Digicel's June 2022 market share report with the denominator calculated from HRSL population layer i.e.

  pen_department = (cdr_digicel_department + cdr_natcom_department) / hrsl_department_population

The phone user penetration rate for each admin 1 unit *j* (department + PaP metro) is calculated:

$p_j = (D_j + N_j)/P_j$

Where:

D_j is the number of Digicel subscribers
N_j is the number of Natcom subscribers
P_j is the HRSL population estimate

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