FLOWMINDER.ORG

Flowminder Foundation (www.flowminder.org) West Africa human mobility models: Version 1.0, August 21st 2014.

A. Wesolowski[1,2,3], L. Bengtsson[1,4], C.O. Buckee[1,2,3], X. Lu[5,1,4], E. Wetter[1,6], A. J. Tatem[1,7,8]*

[1]Flowminder Foundation Research Group, Stockholm, Sweden.
[2]Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA.
[3] Center for Communicable Disease Dynamics, Harvard School of Public Health, Boston, MA, USA.
[4] Department of Public Health Sciences, Karolinska Institute, Stockholm, Sweden.
[5] College of Information System and Management, National University of Defense Technology, Changsha, China.
[6] Department of Management, Stockholm School of Economics, Stockholm, Sweden.
[7] Department of Geography and Environment, University of Southampton, Southampton, UK.
[8] Fogarty International Center, National Institutes of Health, Besthesda, MD, USA.

This document describes the production of a set of version 1 Flowminder human mobility models for West Africa, built on WorldPop population data (www.worldpop.org.uk), to support ongoing efforts to control the ebola outbreak. Readily available mobility data for the region are generally outdated and of poor resolution, so care should be taken in using these datasets. They represent pre-outbreak 'typical' mobility patterns and do not account for any travel restrictions that have been imposed in the region. Improved version 2 datasets are already under construction and will include larger and more recent mobility datasets, as well as different measurements of distance and settlement characterization. Please note that these data have been put together rapidly in response to requests for mobility measures, so please let us know of any problems/errors through the WorldPop contact page: www.worldpop.org.uk/contact.

We have analyzed a number of existing data sources from national census microdata samples, mobile phone call detail records (CDRs), and spatial population data in order to attempt to better understand intra and inter national mobility patterns in fifteen West African countries (Benin (BEN), Burkina Faso (BFA), Cote d'Ivoire (CIV), Cameroon (CMR), Ghana (GHA), GIN (Guinea), GMB (Gambia), Guinea-Bissau (GNB), Liberia (LBR), Mali (MLI), Niger (NER), NGA (Nigeria), Senegal (SEN), Sierra Leone (SLE), and Togo (TGO)).

Spatial Population Data

We obtained population estimates from the WorldPop Project (www.worldpop.org.uk). Settlement locations were obtained from the recently released WorldPop West Africa dataset (www.worldpop.org.uk/ebola). For the administrative unit-based models outlined below, the population totals for each administrative unit were extracted from the WorldPop layers. For the 'settlement'-based models, firstly thiessen polygons around each settlement location were constructed and the total population in each from WorldPop layers were extracted. This follows previous analyses for Kenya [1] and was undertaken to ensure that all populations in the region were captured, and that rural populations were not missed. Version 2 datasets will produce additional outputs based only on the estimated 2014 populations of settlements.

Mobility Data

Overall, movement data in the West Africa region is poor in terms of spatial and temporal resolution, and availability of recent data. In the majority of countries, the best, freely available source of movement data is census migration data that quantifies movement patterns in the form of change of residence over the course of a year and over large spatial areas. For two countries in the region (Senegal and Cote d'Ivoire), we analyzed travel patterns from mobile phone call detail records (CDRs). More information on CDRs and their application in disaster, disease and development contexts can be found through the publications on the Flowminder website: www.flowminder.org/research/. Using these CDRs, we were able to obtain more recent (Figure 1) and finer resolution depictions, both temporally and spatially, of movement within these countries, although in both instances the subset of individuals likely does not form a representative sample of the population.

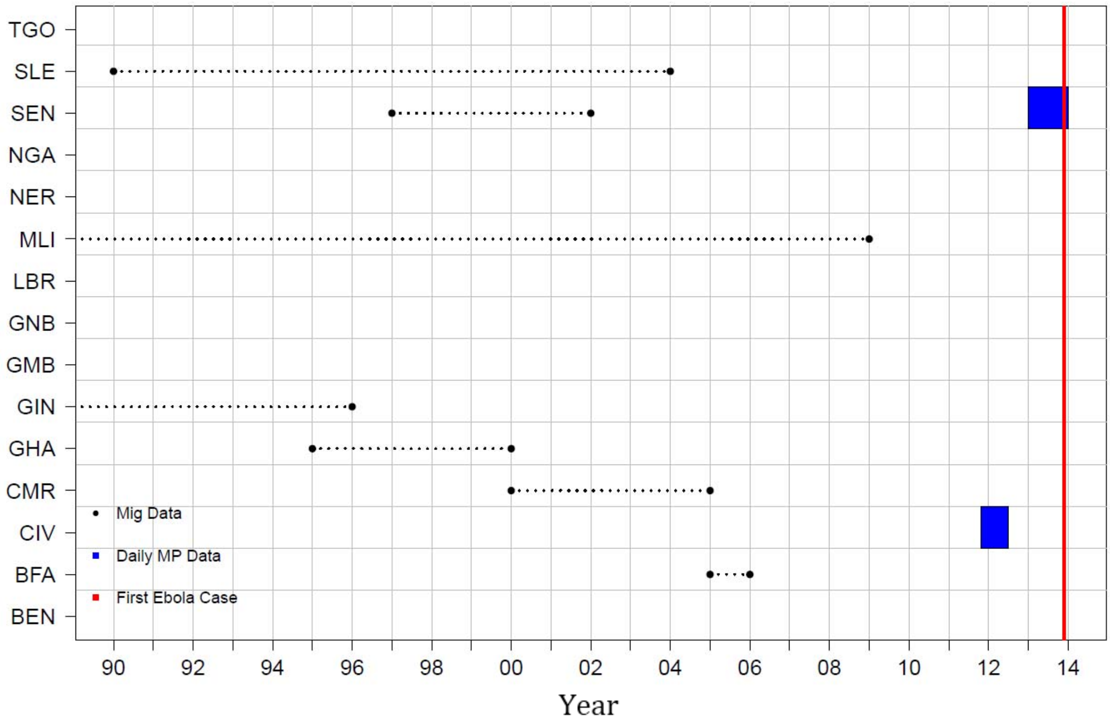

Figure 1. The time periods covered by the mobility datasets used to construct the version 1 Flowminder movement models. (Mig = Migration; MP = Mobile Phone).

Census microdata samples were obtained from the Integrated Public Use Microdata Series (IPUMS) International online repository (https://international.ipums.org/international/). Migration data (question phrased as: "*Where did you live last year/5 years ago/10 years ago?")* were available in a number of the microdata sets for West African countries and were aggregated either to administrative unit level 1 or level 2 (see Table 1 for a summary of the microcensus data used here [2]. These are being updated with more recent samples for version 2 data). For a number of countries (BFA, CMR, GIN, MLI, and SLE), international migration data were also available. Migration data can serve as a proxy for the relative connectivity between admin units and countries [3], however the type of travel (long term migration) is less relevant for the spread of infectious diseases than short-term movements (both temporally and spatially).

Table 1: Census microdata samples used in constructing the migration-based mobility models outlined here

Country/Year	Fraction of census in sample	Households	Persons	Census date	Smallest geography
Burkina Faso 2006	10	236,206	1,417,824	9/23-12-06	commune
Cameroon 2005	10	345,363	1,772,359	11/11/2005	arrondissement
Ghana 2000	10	397,097	1,894,133	26/03/2000	district
Guinea 1996	10	108,793	729,071	01/12/1996	prefecture
Mali 2009	10	235,834	1,451,856	14/04/2009	district
Senegal 2002	10	107,999	994,562	N/A	department
Sierra Leone 2004	10	82,518	494,298	04/12/2004	chiefdom

Two mobile phone CDR data sets were provided by Orange Telecom as part of the Data for Development Challenge (D4D). CDRs from a random sample of 500,000 anonymous mobile phone subscribers who were active from December 1, 2011 to April 28, 2012 were available for Cote d'Ivoire. The user's location was provided at the subprefecture level (255 total, out of which 237 had at least one mobile phone tower) of the routing mobile phone tower. A more detailed description of the data can be found in Lu et al [4]. Similarly, CDRs from Orange Telecom subscribers in Senegal were provided through an exceptional authorization in support of ebola control efforts. For one year, January 1 to December 31, 2013, coarse-grained (123 arrondissements) mobility data for 150,000 randomly sampled individuals were available (for a detailed description of the data, see: http://arxiv.org/abs/1407.4885). In the Senegal mobile dataset, anonymous phone users were included if they had at least one communication event during more than 75% of the days during 2013. It is possible that users who call more are on average also traveling more as both behaviors are often positively related to socioeconomic status. This could mean that mobility in general is overestimated. Relative connectivity between areas is likely to be less affected.

Additionally an anonymized comprehensive set of CDRs from June 2008 – June 2009 (excluding February 2009) was provided by the leading mobile phone operator in Kenya (92% market share) for individual subscribers (14,816,521) with locations identified at the mobile phone tower level (12,502 in total). This dataset enabled us to produce finer resolution mobility estimates than from the aggregated datasets outlined above, with parameters showing little differences from the Cote d'Ivoire and Senegal dataset models (see below). We aggregated our data to quantify human travel patterns over the course of the year between 69 Kenyan districts and 692 mapped settlements.

Models of movement

We developed multiple movement models for within country and between country travel patterns using the available data sources described above and pre-existing models. Computing code for implementing these models in the open statistical package R (www.r-project.org) will soon be provided. The gravity model is the simplest spatial interaction model, where the amount of travel (N_{ij}) between two locations (i,j) is dependent on their populations (pop_i, pop_j) and the physical distance separating them (d(i,j)):

$$N_{ij} = k\frac{\left(pop_i^{\alpha} + pop_j^{\beta}\right)}{dist(i,j)^{\gamma}}$$

where the parameters α, β, γ, k are fit based on a Poisson distribution.

Table 2: The parameter estimates from fitted gravity models.

Locations	**Intercept (k)**	**Pop From (α)**	**Pop To (β)**	**Euclidean Distance (γ)**	**% Reduction in Deviance**
Cote d'Ivoire (***civ***)	-13.83	0.86	0.78	-1.52	73.28
Senegal (***sen***)	-3.93	0.47	0.46	-1.78	89.73
Kenya – district (***kenya***)	-20.61	1.22	1.22	-2.05	80.06
Kenya - settlement	-6.00	0.66	0.61	-0.67	47.31
Entire IPUMS migration data set (***ipums***)	-23.51	1.13	1.11	-0.95	95.30
IPUMS – BEN (***ipums_country***)	-13.86	0.82	0.79	-0.95	91.29
IPUMS – BFA (***ipums_country***)	-25.32	1.07	1.09	-1.03	60.89
IPUMS – CIV (***ipums_country***)	-15.72	0.90	0.86	-1.18	95.88
IPUMS – CMR (***ipums_country***)	-29.85	1.10	1.51	-0.93	68.33

IPUMS – GHA (***ipums_country***)	-12.68	0.29	1.04	-0.94	66.99
IPUMS – GIN (***ipums_country***)	-29.45	0.99	1.64	-0.69	69.40
IPUMS – GMB (***ipums_country***)	-20.59	1.05	1.01	-1.31	78.00
IPUMS – GNB (***ipums_country***)	-16.98	0.94	0.91	-0.98	91.24
IPUMS – LBR (***ipums_country***)	-16.04	0.90	0.86	-1.04	95.55
IPUMS – MLI (***ipums_country***)	-27.45	1.03	1.25	-0.59	62.05
IPUMS – NER (***ipums_country***)	-6.13	0.56	0.54	-1.09	90.59
IPUMS – NGA (***ipums_country***)	-17.90	0.96	0.92	-0.99	93.57
IPUMS – SEN (***ipums_country***)	-15.20	0.42	1.07	-1.09	68.07
IPUMS – SLE (***ipums_country***)	-54.58	1.67	2.89	-0.51	60.10
IPUMS – TGO (***ipums_country***)	-16.58	0.93	0.89	-1.33	98.61

Previously, we fit a number of gravity models to a more comprehensive mobile phone data set in Kenya. We used these existing models to also provide estimates on the average number of trips per week between settlements in each country. We fit gravity models to all the sets of mobile phone CDRs, the entire set of census migration data, and each country's individual census migration data. For countries missing from each source of data, we used the estimated parameters from these models to estimate amounts of travel.

Data and Model Outputs Available

Models

We have produced the following sets of models, and Table 2 provides the parameter estimates from the fitted gravity models.

1. ***Ipums (MicrocensusModel.rda)***
   Gravity model fit to the entire census microdata set
   $$N_{ij} = k\frac{\left(pop_i^{\alpha} + pop_j^{\beta}\right)}{dist(i,j)^{\gamma}}$$
2. ***ipums_country***

Gravity model fit to each country's census microdata set

3. ***civ (CIVModel.rda)***

   Gravity model fit to mobility between subprefectures in Cote d'Ivoire from mobile phone CDRs

4. ***kenya (KenyaModel.rda)***

   Gravity model fit to mobility between districts in Kenya from mobile phone CDRs

5. ***sen (SenModel.rda)***

   Gravity model fit to mobility between arrondissements in Senegal from mobile phone CDRs

Spatial Data

The spatial data section contains ESRI shapefiles of:

1. IPUMS sublocations (admin 1 or 2, depending on the spatial resolution of the census microdata).
2. Locations of mapped settlements.

The IDs in the tables of each of these datasets enables matching to the model prediction outputs (see below).

Mobility Data and Model Predictions

The text and tables below describes each variable in the various mobility data and model predictions. Figure 2 shows the predicted ranges of within-country mobility for four of the models. Each model has different features and benefits that should be considered when choosing between them. In terms of an overall 'best' model, the features of the Senegal CDRs and the fact that the data are the most recent makes this our currently preferred model.

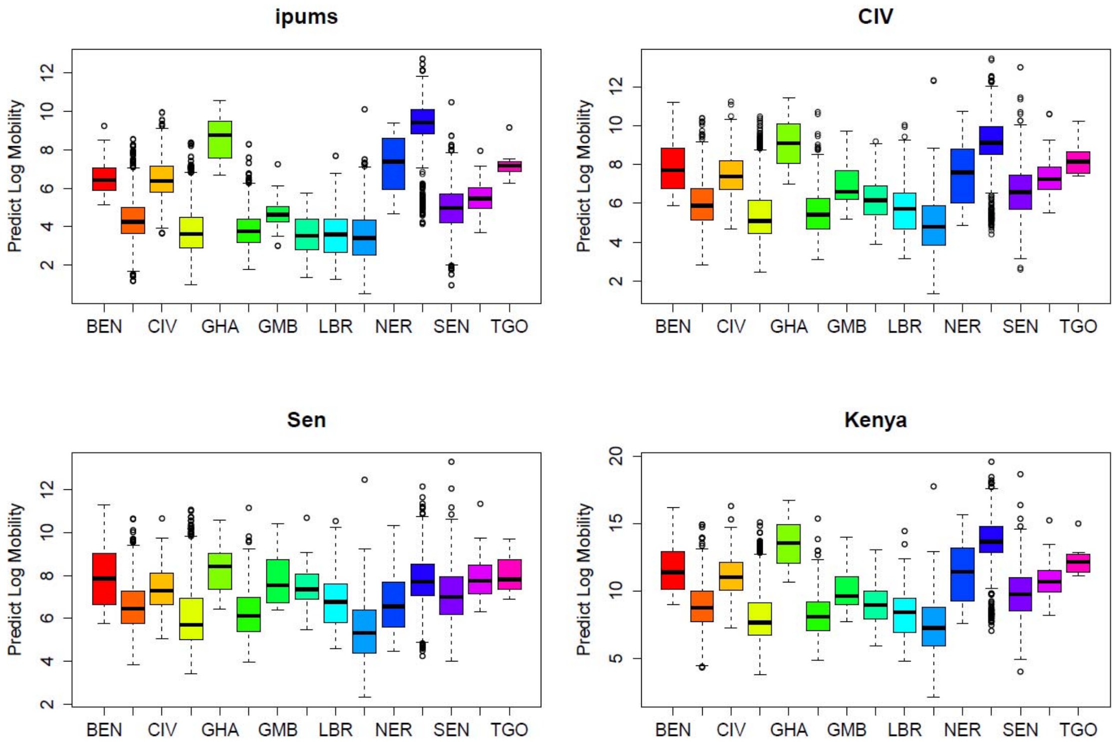

Figure 2. Predicted ranges of within-country mobility using the models parameterized on census microdata migration data (ipums), Cote d'Ivoire CDRs (CIV), Senegal CDRs (Sen) and Kenya CDRs (Kenya).

The following datasets are available:

1. **AdminUnits_Within.csv**
   a. All pairs of within country census microdata sublocations
   b. Number of trips from the census microdata
   c. Population estimates
   d. Euclidean distance between sublocation centroids
   e. Model predictions from ***ipums***, ***ipums_country***, ***civ***, ***senegal***, and ***Kenya***

Table 3: AdminUnits_Within.csv variable descriptions.

Variable Name	**Description**
from_loc	Origin location admin unit 1 or 2
to_loc	Destination location admin unit 1 or 2
amt	Amount of migration reported in the census microdata or modeled amount from [Ref]
country	Country ISO code
from_pop	Origin population (www.worldpop.org.uk)
from_x	Origin centroid, x

from_y	Origin centroid, y
to_pop	Destination population (www.worldpop.org.uk)
to_x	Destination centroid, x
to_y	Destination centroid, y
euc_dist	Euclidean distance between polygon centroids
predict_ipums	Predicted amount of travel from microcensus model (***ipums***)
predict_ipums_country	Predicted amount of travel from microcensus model per country (***ipums_country***)
predict_civ	Predicted amount of travel from CIV model (***civ***)
predict_kenya	Predicted amount of travel from Kenya model (***kenya***)
predict_sen	Predicted amount of travel from Senegal model (***sen***)

2. **AdmUnits_WBtwn.csv**
    a. All pairs of sublocations (including international pairs) from the census microdata
    b. Number of trips from the census microdata
    c. Population estimates
    d. Euclidean distance between sublocation centroids
    e. Model predictions from ***ipums***, ***ipums_country***, ***civ***, ***senegal***, and ***kenya***

Table 4: AdmUnits_WBtwn.csv variable descriptions.

Variable Name	Description
from_loc	Origin location admin unit 1 or 2
to_loc	Destination location admin unit 1 or 2
from_pop	Origin population (www.worldpop.org.uk)
from_x	Origin centroid, x
from_y	Origin centroid, y
from_loc_adm_id	Origin location ID (matches labels in AdminUnits_Within.csv)
from_loc_country	Origin location country (matches labels in AdminUnits_Within.csv)
to_pop	Destination population (www.worldpop.org.uk)
to_x	Destination centroid, x
to_y	Destination centroid, y
to_loc_adm_id	Destination location ID (matches labels in AdminUnits_Within.csv)
to_loc_country	Destination location country (matches labels in AdminUnits_Within.csv)
euc_dist	Euclidean distance between polygon centroids
predict_ipums	Predicted amount of travel from microcensus model (***ipums***)

predict_civ	Predicted amount of travel from CIV model (***civ***)
predict_kenya	Predicted amount of travel from Kenya model (***kenya***)
predict_sen	Predicted amount of travel from Senegal model (***sen***)

3. **MigrationBtwnCountries.csv**
    a. Migration from Burkina Faso, Cameroon, Guinea, Mali, and Sierra Leone to other countries from census microdata

Table 5: MigrationBtwnCountries.csv variable descriptions.

Variable Name	**Description**
from_loc	Origin country
to_loc	Destination country
amt	Amount of migration reported in the census microdata
from_x	Origin centroid, x
from_y	Origin centroid, y
to_x	Destination centroid, x
to_y	Destination centroid, y

4. **CIV_GModel.csv**
    a. Predictions from the gravity model (***civ***) of movement between subprefectures based on mobile phone data from Cote d'Ivoire.
5. **Sen_GModel.csv**
    a. Predictions from the gravity model (***sen***) of movement between arrondissements based on mobile phone data from Senegal.
6. **Kenya_GModel.csv**
    a. Predictions from the gravity model (***kenya***) of movement between districts based on mobile phone data from Kenya.

Table 6: CIV_GModel.csv, Sen_GModel.csv, and Kenya_GModel.csv variable descriptions.

Variable Name	**Description**
from_loc	Origin location admin unit 1 or 2
to_loc	Destination location admin unit 1 or 2
from_pop	Origin population (www.worldpop.org.uk)
from_x	Origin centroid, x
from_y	Origin centroid, y
to_pop	Destination population (www.worldpop.org.uk)
to_x	Destination centroid, x
to_y	Destination centroid, y
euc_dist	Euclidean distance between polygon centroids

predict_model	Predicted amount of travel from country, mobile phone data based gravity model

References

1. A. Wesolowski, N. Eagle, A.J. Tatem, D.L. Smith, A.M. Noor, R.W. Snow, C.O. Buckee (2012) Quantifying the impact of human mobility on malaria. *Science* 338:267-270.

2. A.J. Garcia, D.K. Pindolia, K.K. Lopiano, A.J. Tatem (2014) Modeling internal migration flows in sub-Saharan Africa using census microdata. *Migration Studies* 1-22.

3. A. Wesolowski, C.O. Buckee, D. Pindolia, N. Eagle, D.L. Smith, A. Garcia, A.J. Tatem (2013) The use of census migration data to approximate human movement patterns across temporal scales, *PLoS One* 8(1):e52971

4. X. Lu, E. Wetter, N. Bharti, A.J. Tatem, L. Bengtsson (2013) Approaching the limit of predictability in human mobility, *Sci. Rep* 3, 2923