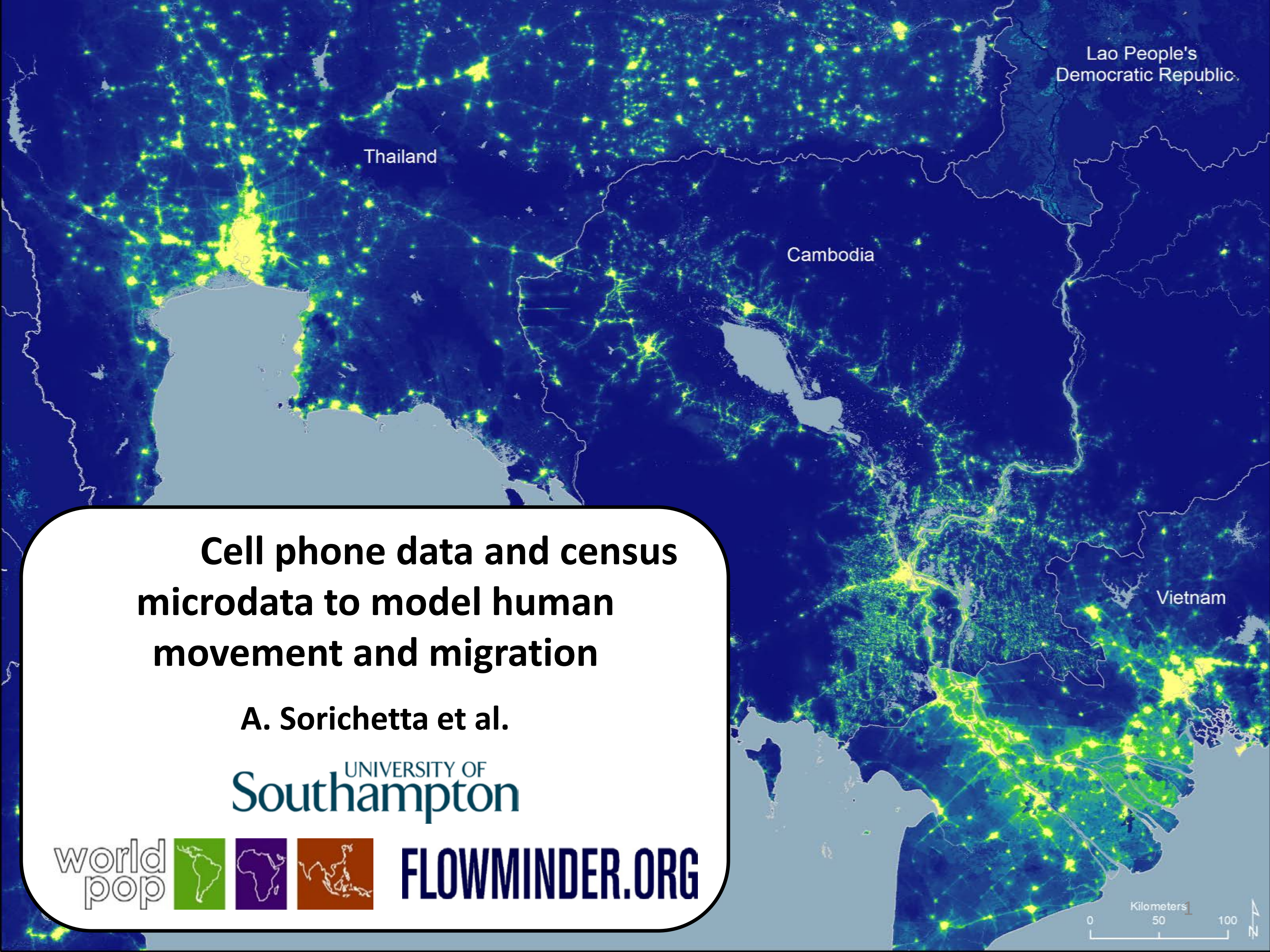


Cell phone data and census microdata to model human movement and migration

A. Sorichetta et al.

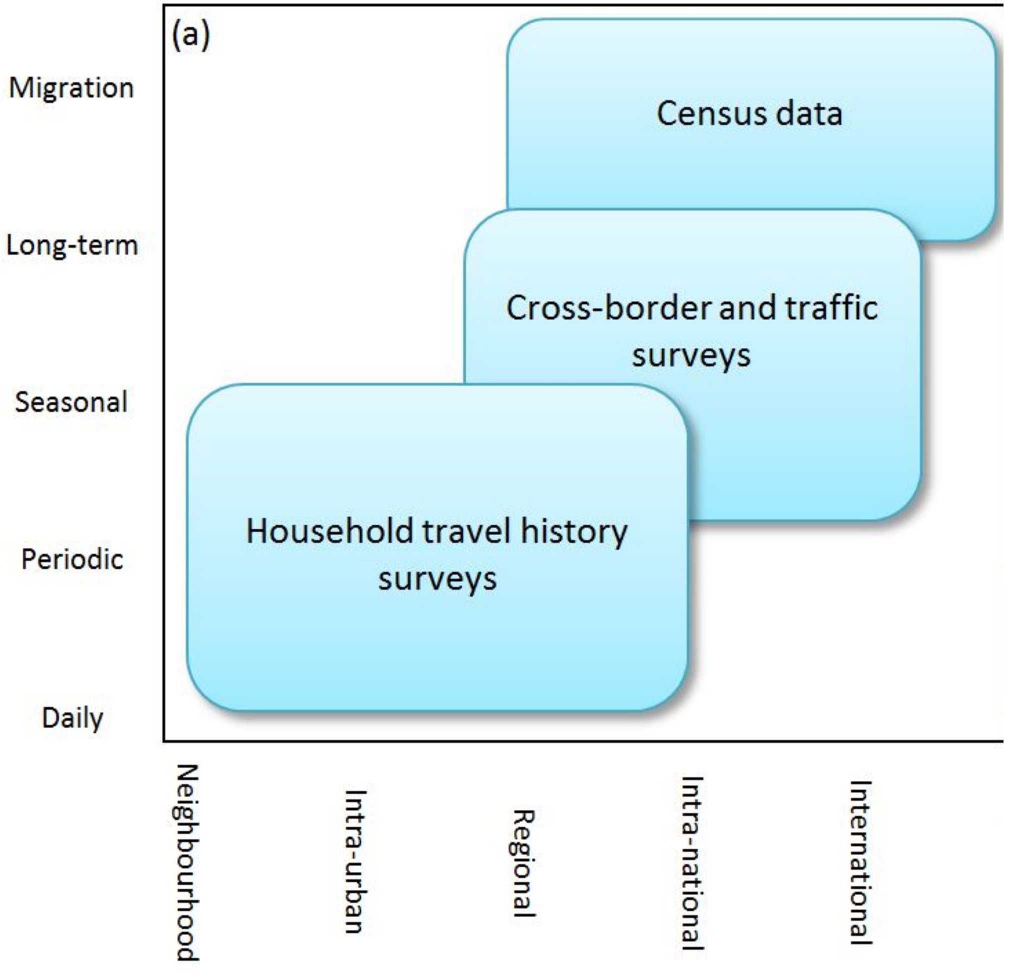
UNIVERSITY OF Southampton

world pop

FLOWMINDER.ORG

Modeling Human Mobility in Space and Time

"Traditional"



Tatem, 2014 (International Health)

Cellphone Call Data Records (CDRs)

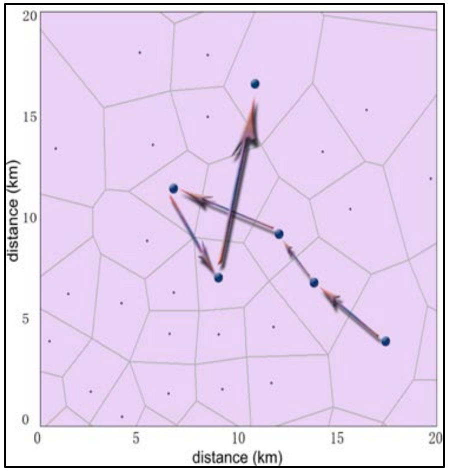
User makes a call from location X

Call routed through nearest tower

Network operator records time and tower of call for billing

X

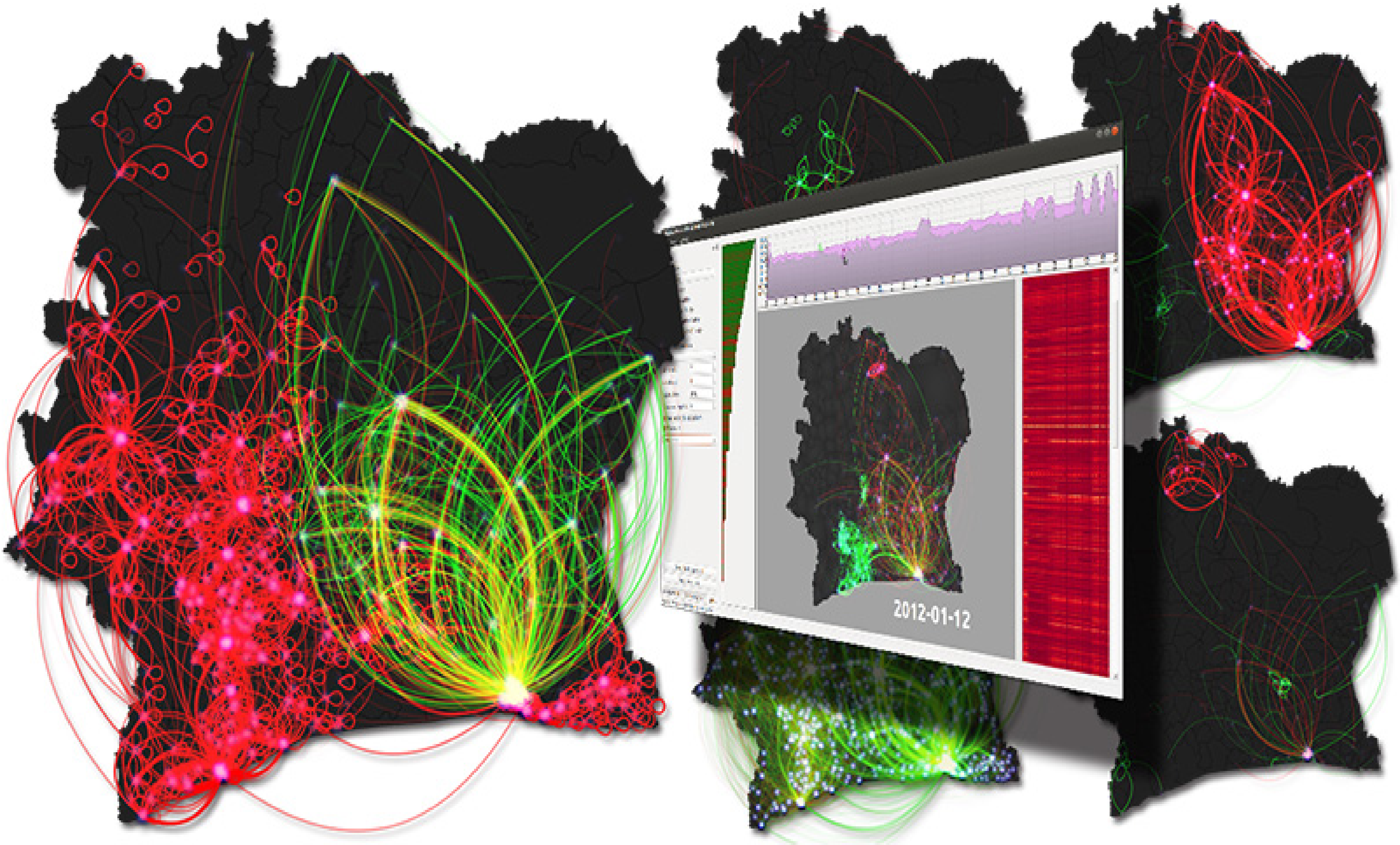
Y



User travels to Y and makes a call

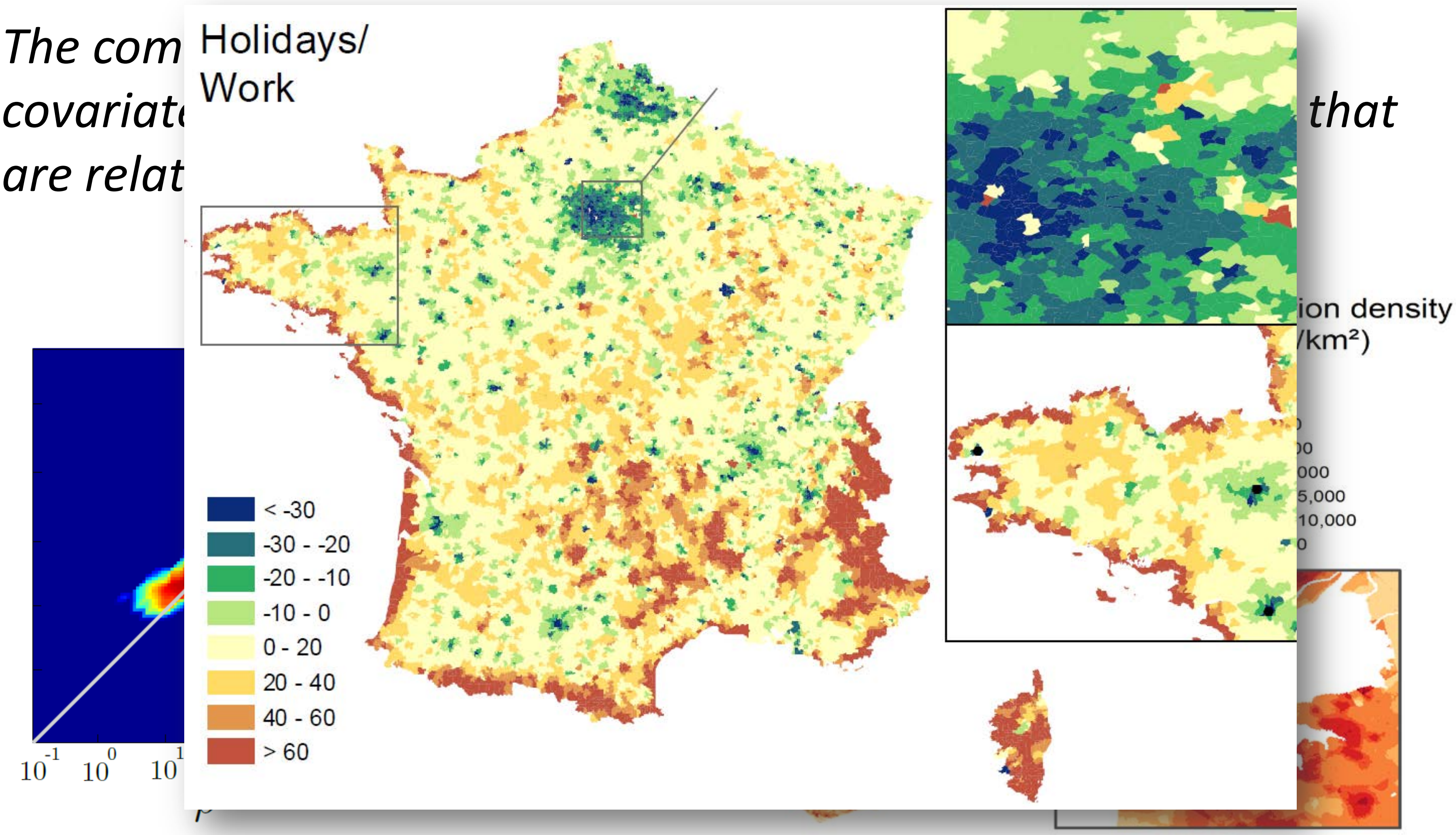
Call routed through nearest tower

Cellphone Call Data Records (CDRs)



More Accurate and Dynamic Assessments of Population Distributions

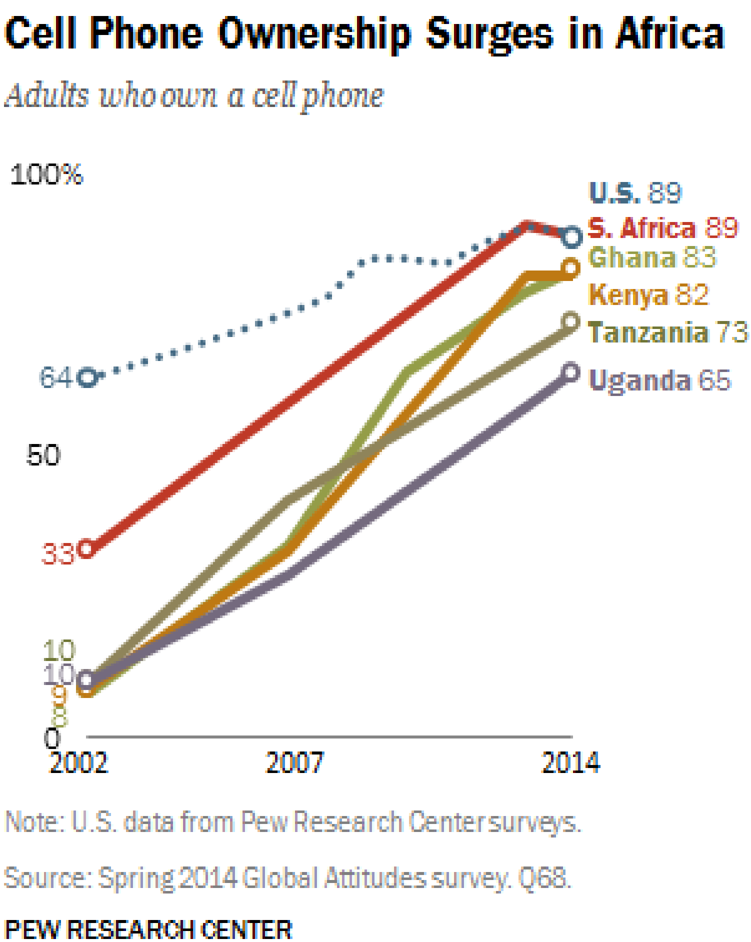
The com
covariat *that*
are relat

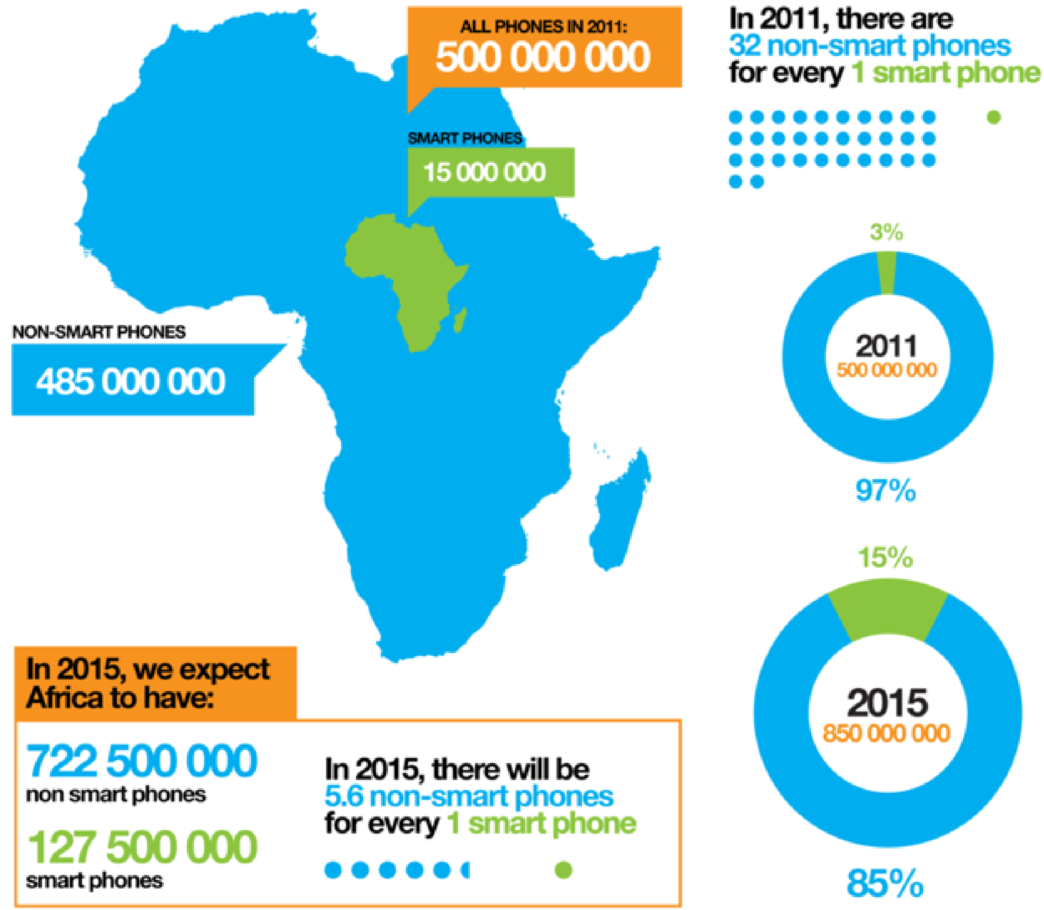


Deville et al., 2014 (PNAS)

Model built on phone usage and satellite data

Cell Phone Ownership

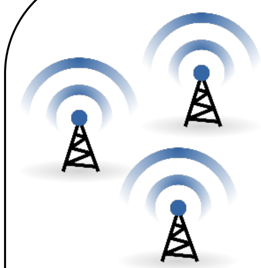


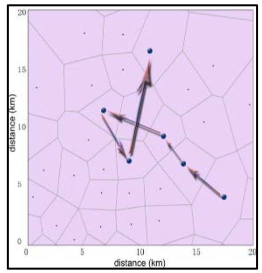


Mobile Phone Data Access

PARTNERS

Preserving Confidentiality!



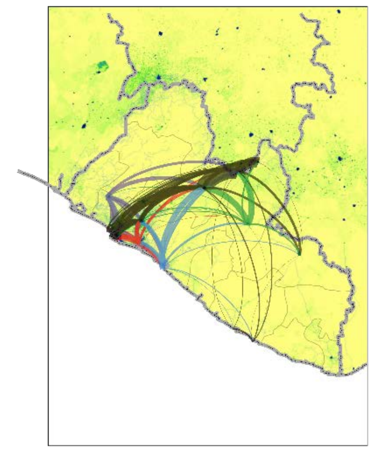


Call Detail Records' (CDRs) including low-resolution location data (nearest tower location) anonymized on separate server hosted by operator.

Analyses are conducted under operator supervision always behind operator firewall

Mobile operator firewall

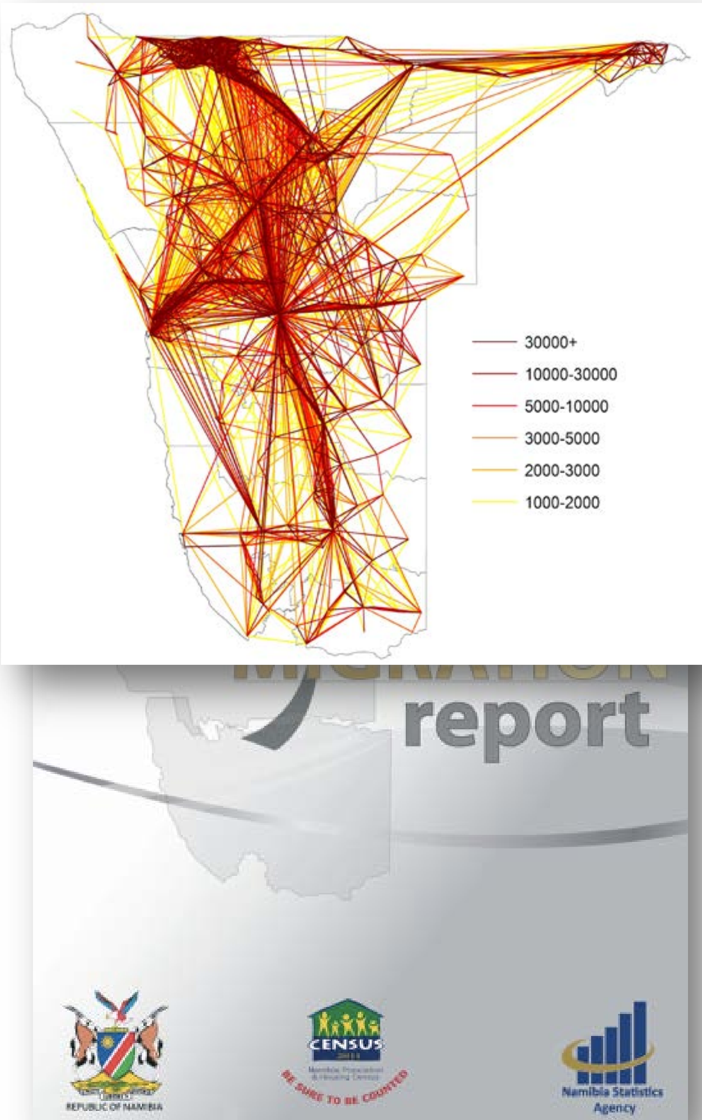
FLOWMINDER.ORG

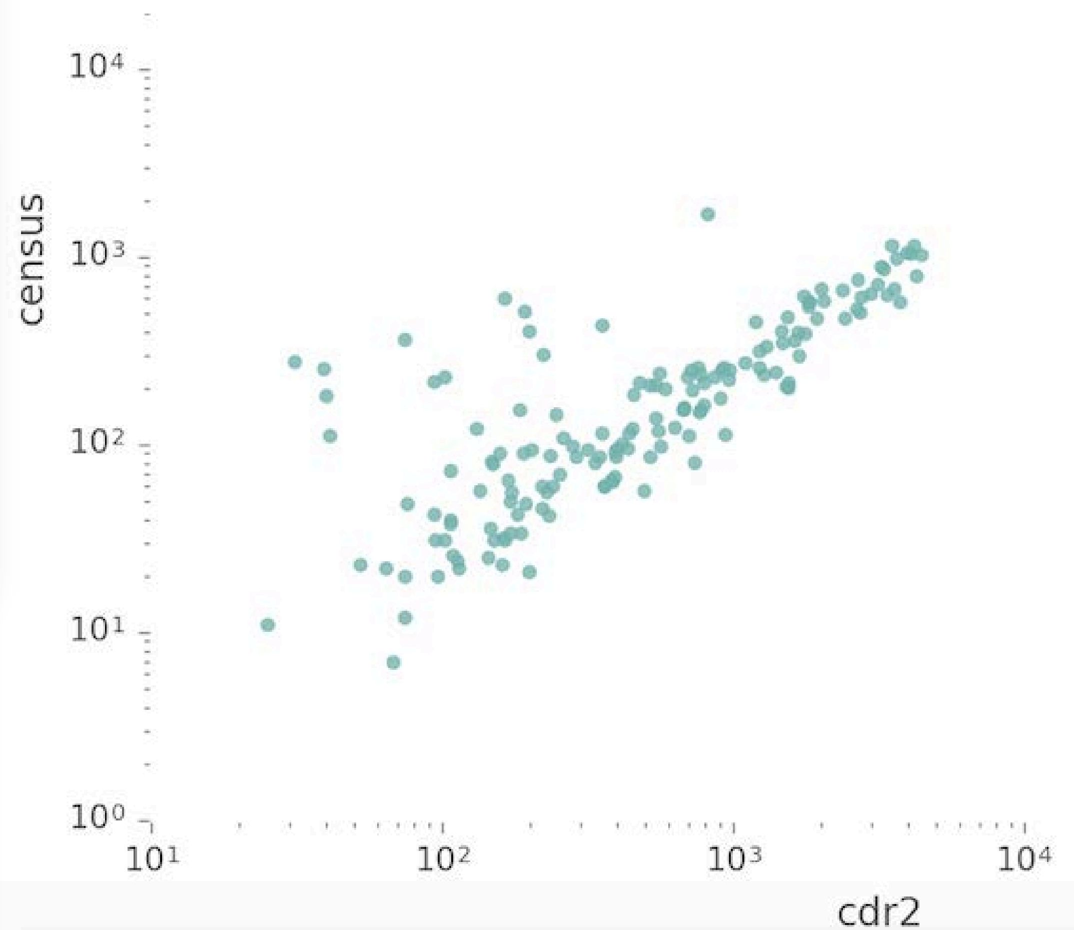


Aggregated mobility estimates are exported and made open access for being potentially used with other mobility estimates and epidemiological data

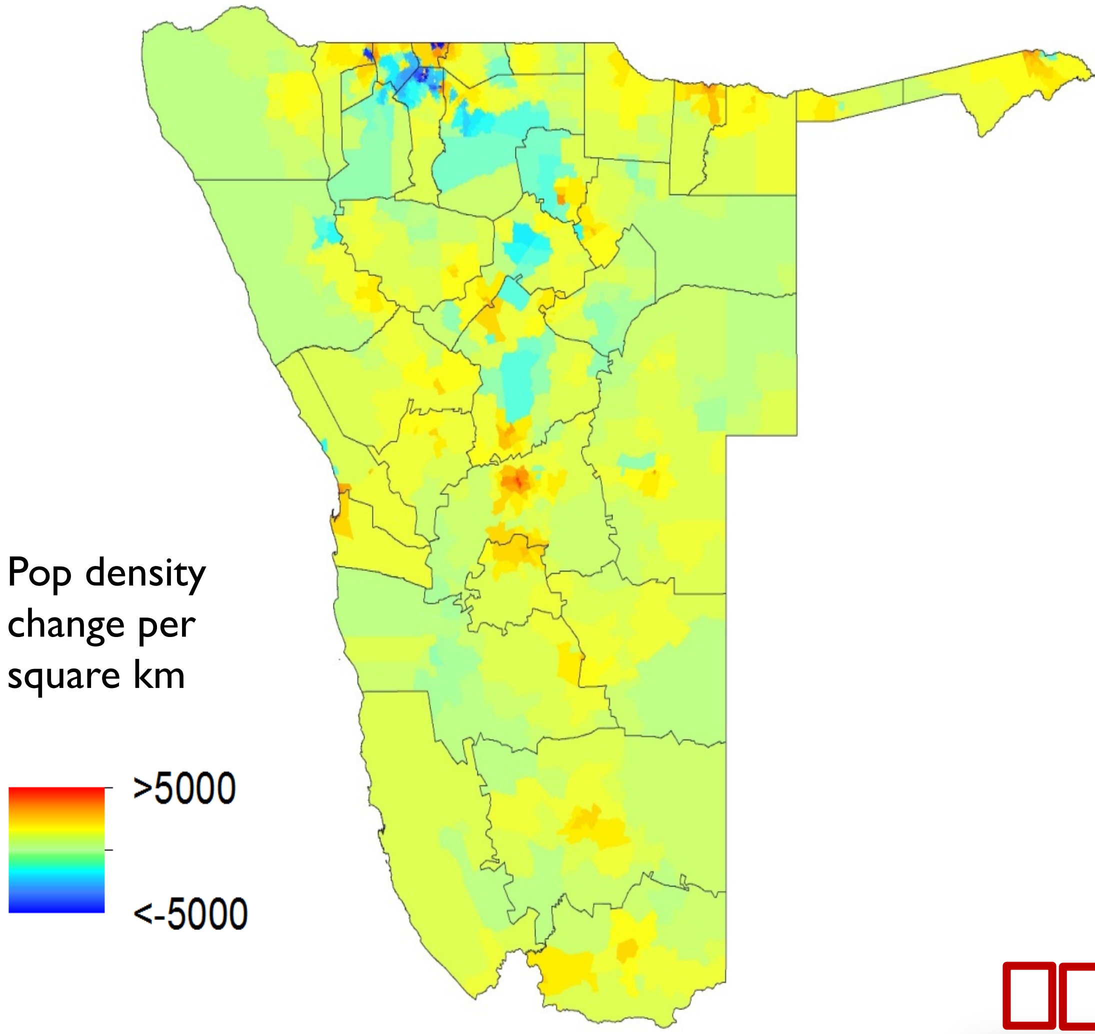
Compliance with GSMA data integrity guidelines: Data never leaves mobile operator's system to avoid any privacy and/or commercial concerns

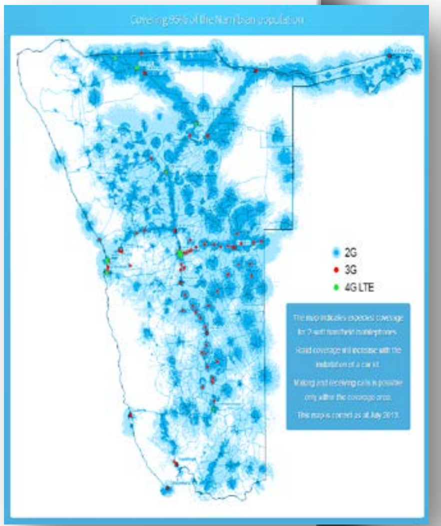
Measuring migration



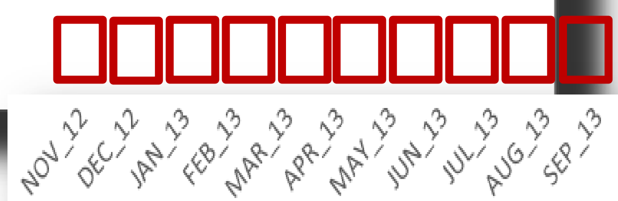


Seasonal Population Mapping



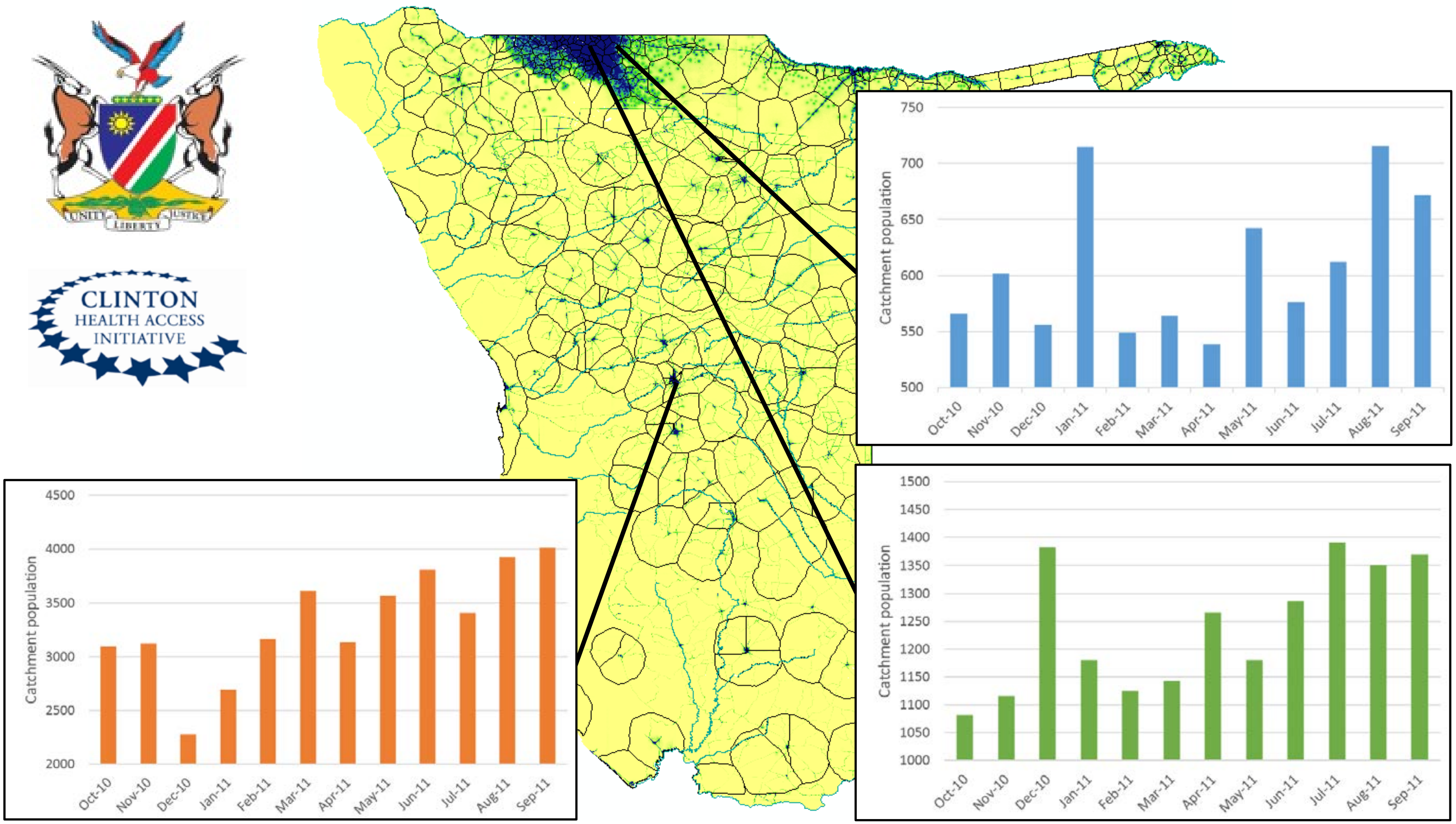


Namibia Pop: 2.3 mill
MTC active subscriptions: 2.1 mill



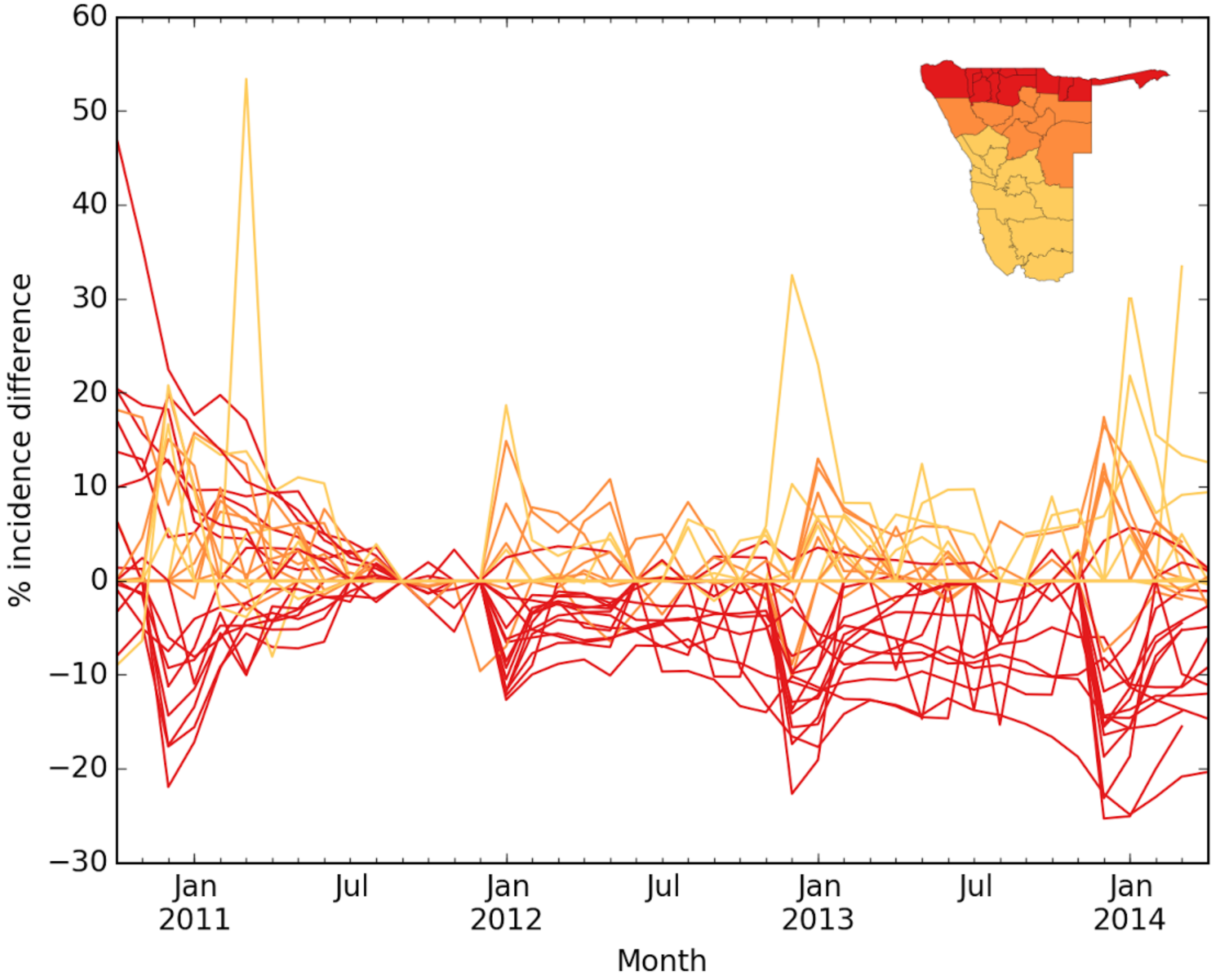
Erbach-Schoenberg et al., 2016 (Population Health Metrics)

Dynamic facility catchment populations



Erbach-Schoenberg et al., 2016 (Population Health Metrics)

Namibia closer to elimination than previously assumed?

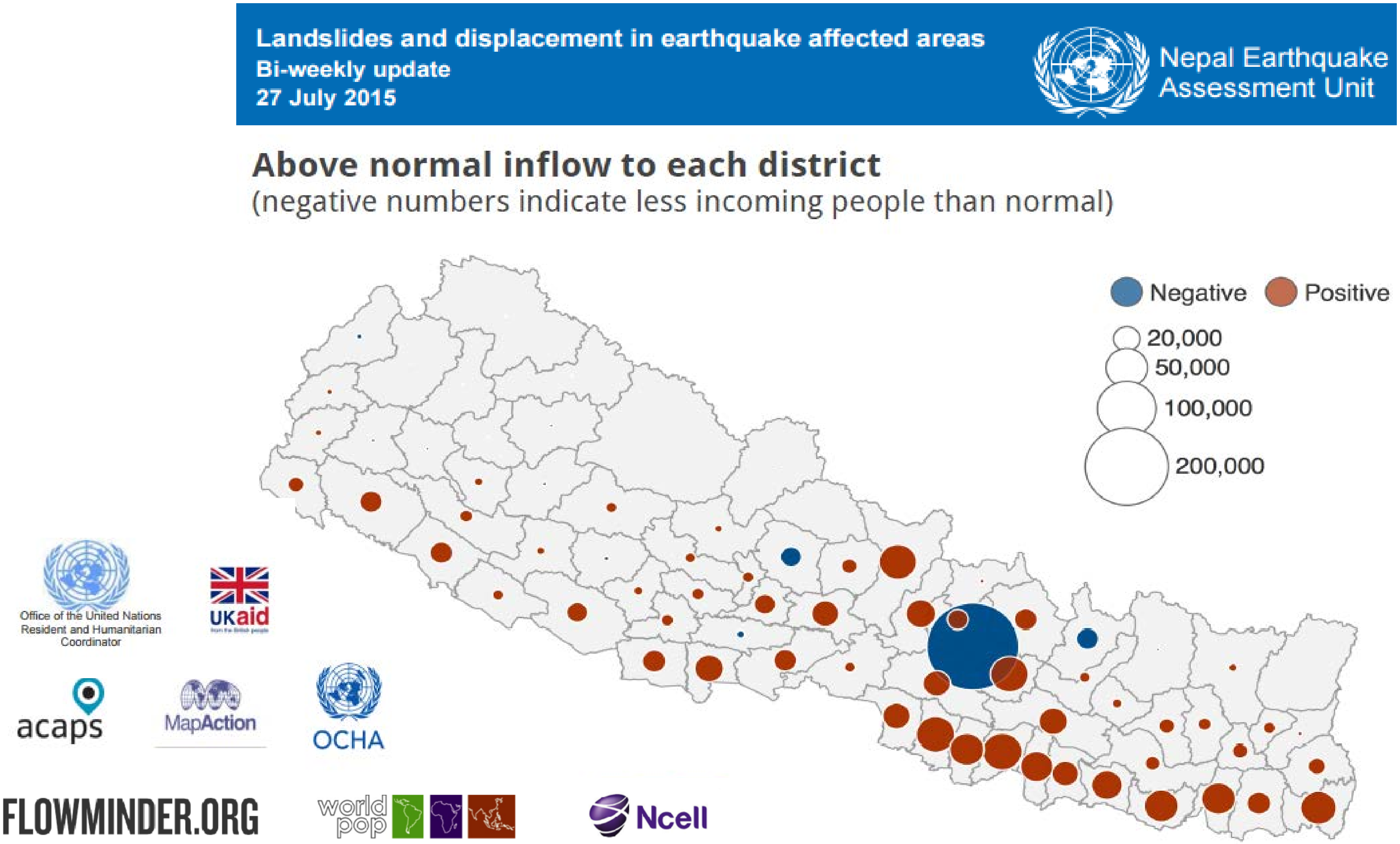


% change in health district incidence through the year after accounting for dynamic catchment populations

Areas in red may have lower incidence than currently assumed using static catchment denominators

Erbach-Schoenberg et al., 2016 (Population Health Metrics)

Understanding shot term mobility



Wilson et al., 2016 (PLoSCurr)

Very large flows from Kathmandu to other districts immediately after the earthquake…

Nepal Population Estimates as of 10th June 2015

Pre-earthquake population	Population outflow (above normal)	Population inflow (above normal)
2.8m	**+180,000** (110,000 ~ 250,000)	**-55,000** (-33,000 ~ -77,000)

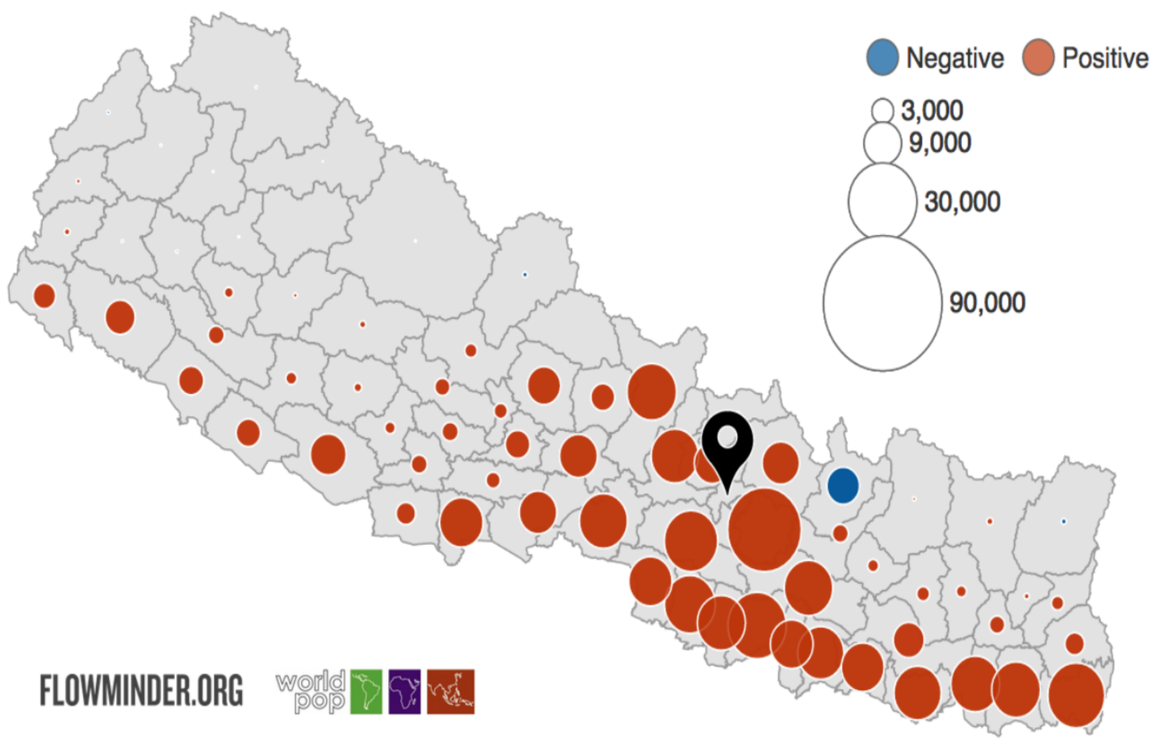
2. Kathmandu Valley

Kathmandu Valley is here defined as the districts Kathmandu, Bhaktapur and Lalitpur. Kathmandu Valley is one of the most densely populated areas in Nepal and home to ca 2.8 m people [1].

Key findings:

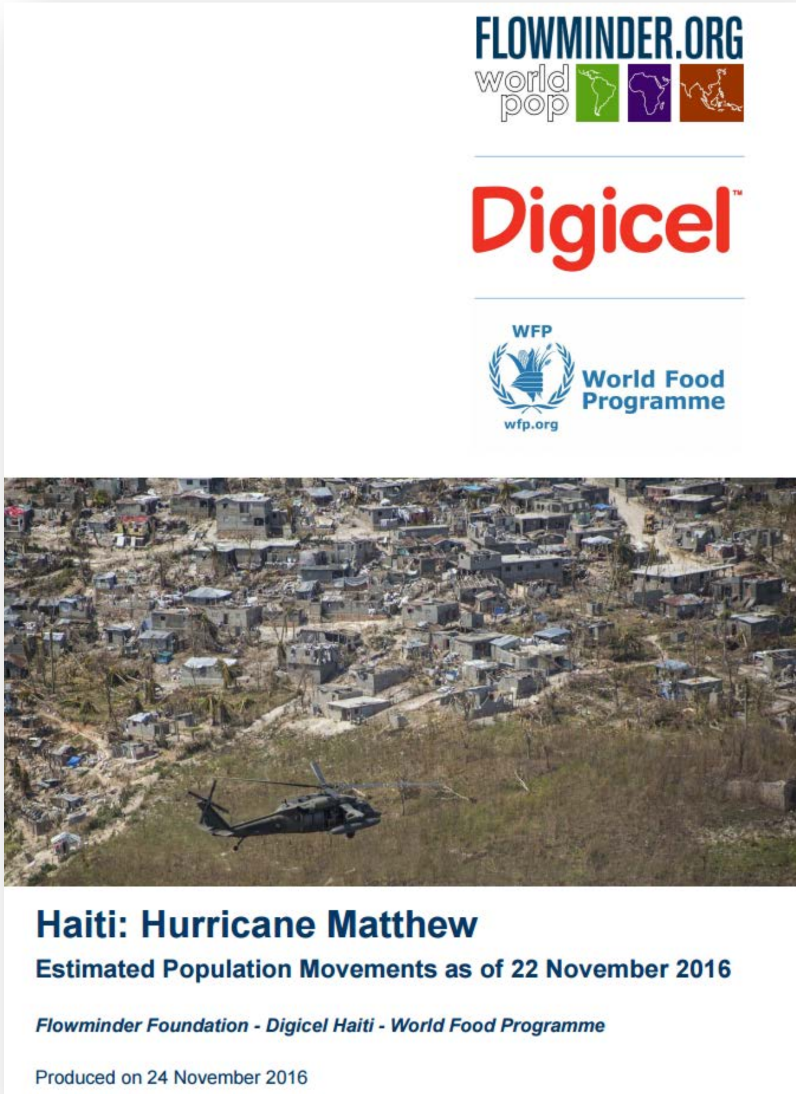
- An estimated 390,000 people more than normal had left the Kathmandu valley - comparing May 1 with the day before the earthquake April 24 (ratio to the population: 14%).
- An estimated 247,000 persons less than normal had come into the area during the same period(ratio to the population: 8.8%)
- People leaving Kathmandu Valley went to a large number of areas, notably the populous areas in the south and the Central and West Development Regions.

Above normal flows from Kathmandu Valley to other districts



Wilson et al., 2016 (PLoSCurr)

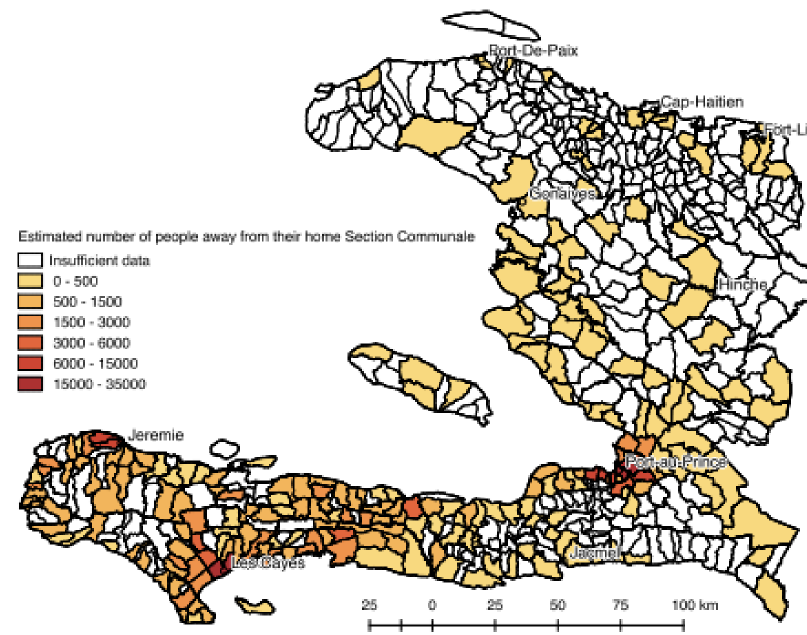
Population Displacements



Estimated population away from their home Section Communale[2]:

HOME DEPARTMENT:	GRANDE ANSE	SUD	NIPPES
POPULATION AWAY FROM HOME:	77500	132000	51000
% AWAY FROM HOME:	18%	17%	15%

24 October 2016, location of people away from their home Section Communale (out of those living pre-hurricane in Grande Anse, Sud and Nippes only)[3]



[2] Of the people normally resident within the given Département, we estimate the total number away from their home Section Communale on the given day.
[3] Section Communales are left blank where insufficient data is available.

CDRs Pros and Cons

Advantages

-Massi mple size, imposs to achieve ravel history surveys

- N cale

- Relatively reli le of stay

Alternative datasets are required in order to quantify and map mobility across continental scales

Bias in repres ents

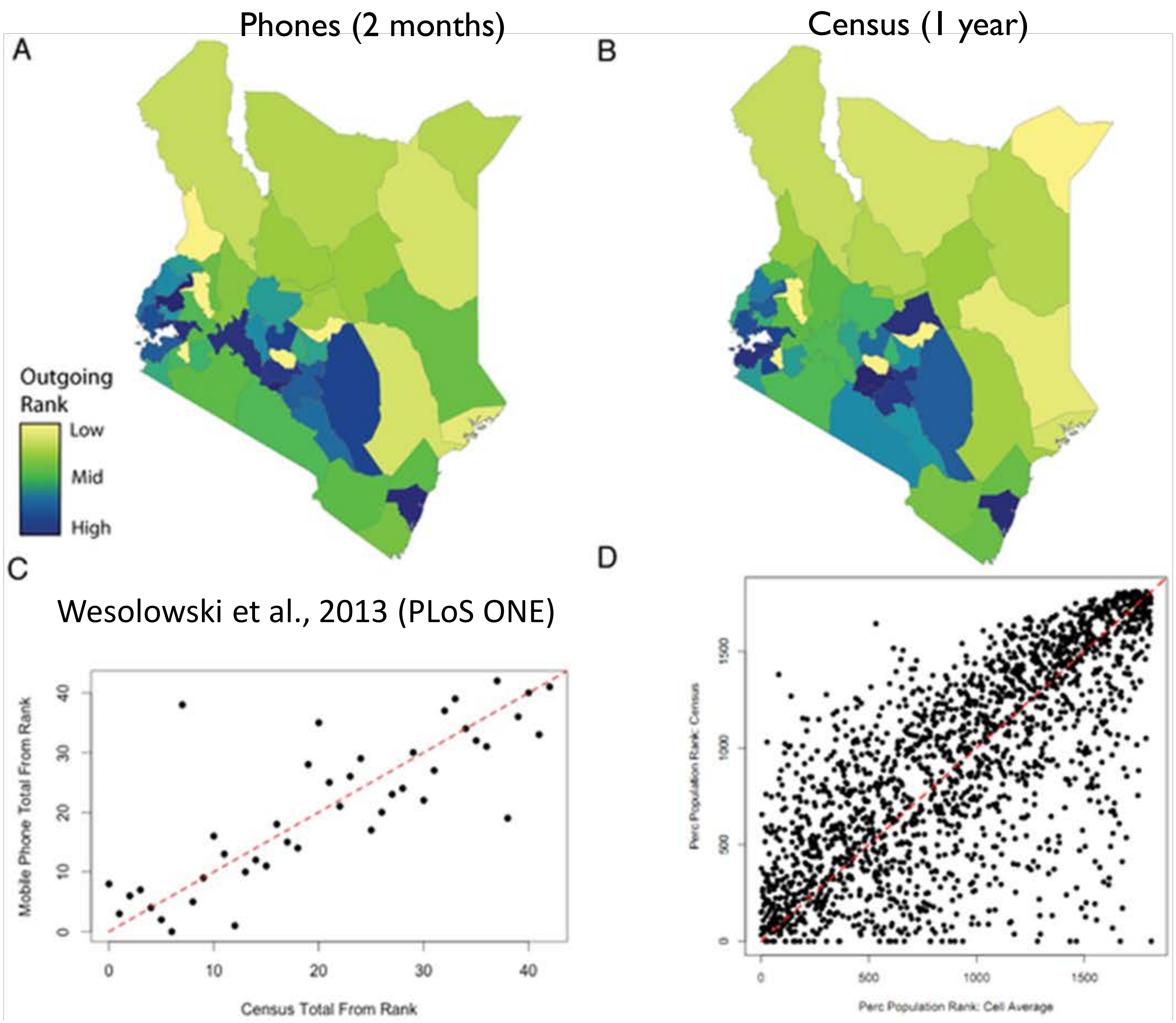
-Co e areas

-No dem ic inf on

-Cross-border mea nents feas but not easy

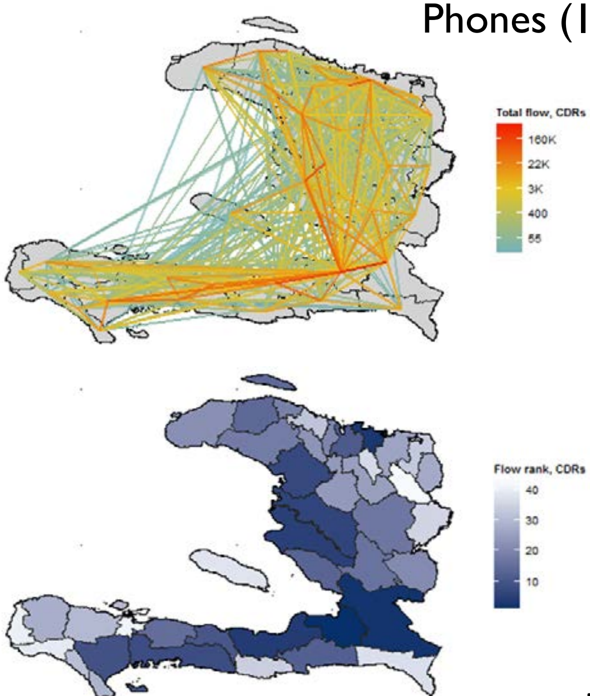
-Difficulties in sharing and a ssing (mostly ue to commercial and priv cy concerns)

A comparison of the ranked estimates of movement (I)

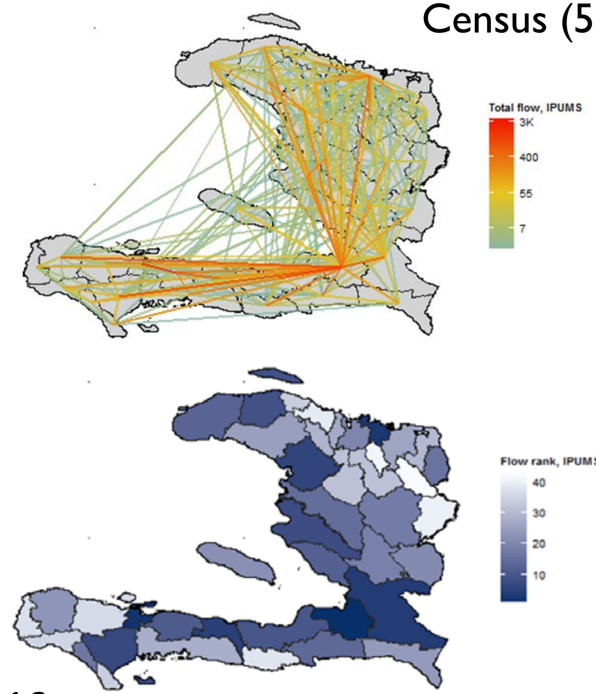


A comparison of the ranked estimates of movement (II)

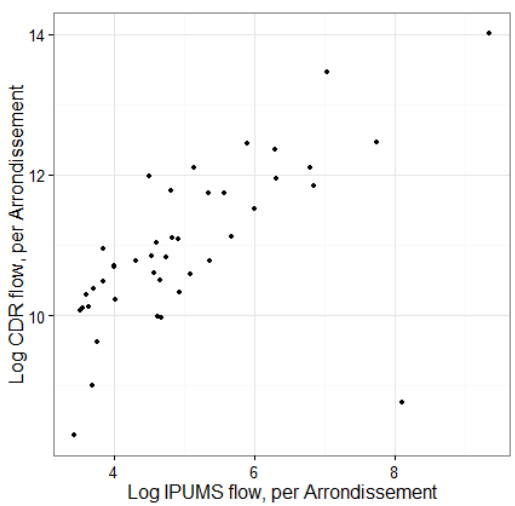
Phones (1month)

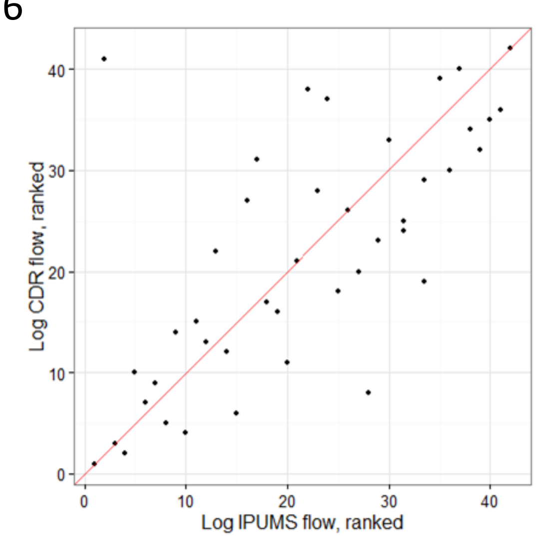


Census (5 years)

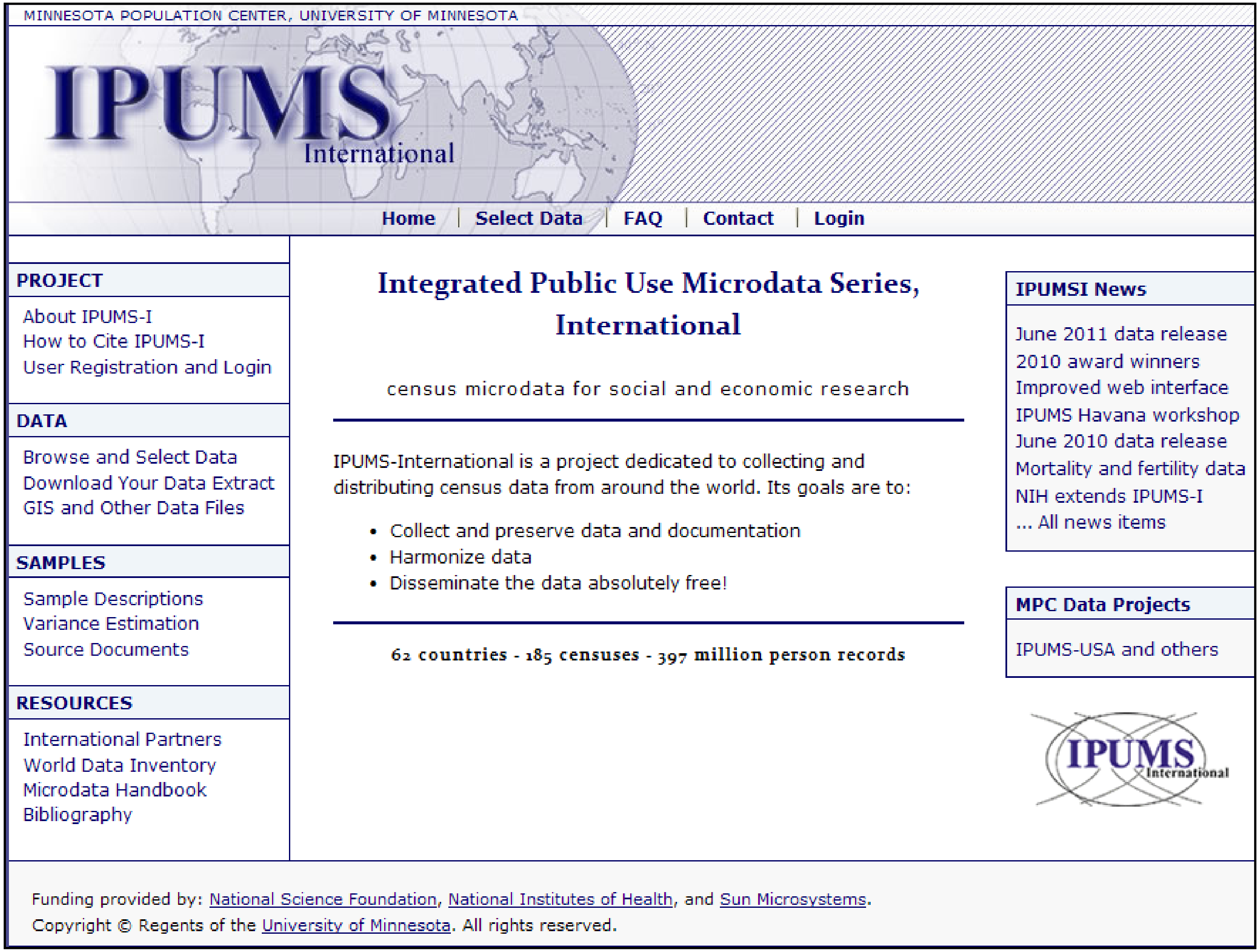


Ruktanonchai et al., 2016
(Malaria Journal)

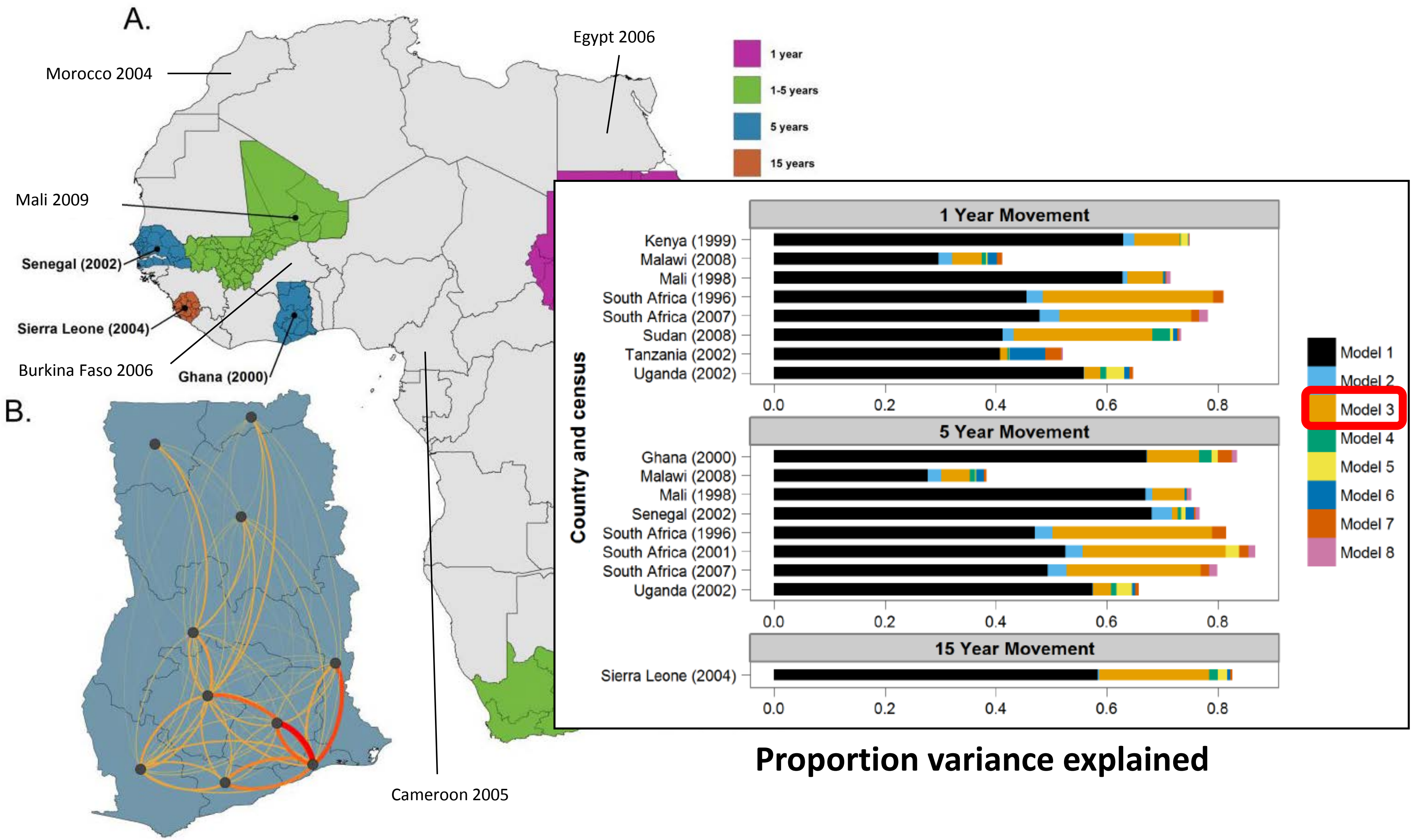




Modelling Internal Migration Using IPUMSI Census Microdata



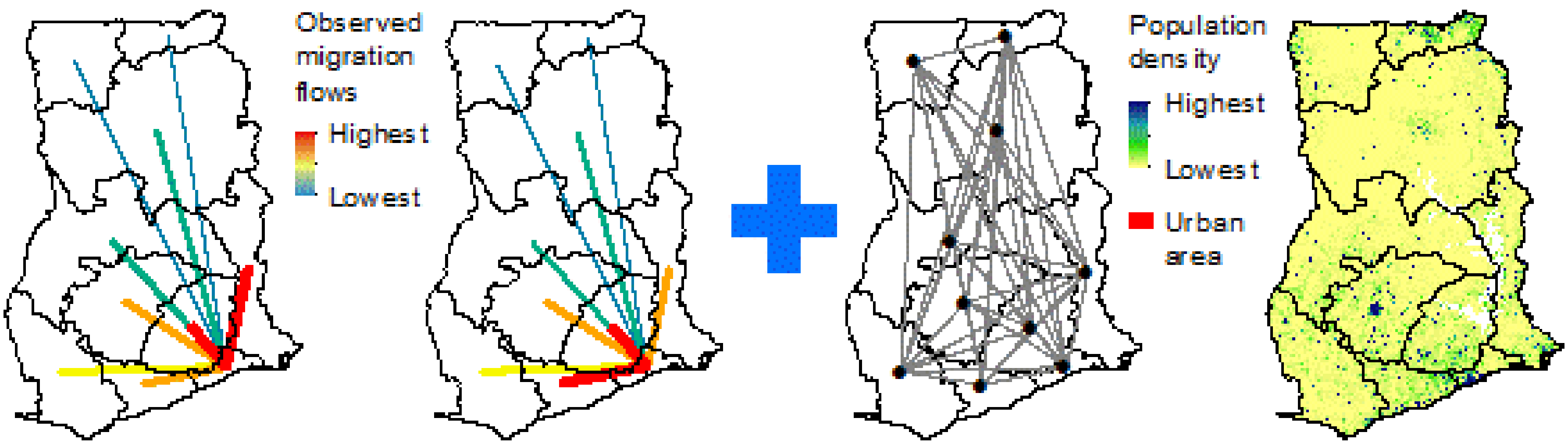
Modelling Internal Migration in Africa



Henry et al., 2013 (Appl. Geogr.); Garcia et al., 2014 (Migration Studies)

Response Variable and Covariates

In order to consistently model internal migration across all countries only globally available datasets proving to be able to explain most of the variance in the gravity models of Garcia et al. were explored.



Internal migration between each pair of admin units - Only inflows to (left) and outflows from (right) the Greater Acrra region are showed here

Distance and contiguity between admin units

Total population and proportion of urban population in each admin unit

http://www.worldpop.org.uk/data/methods/

Modelling Framework

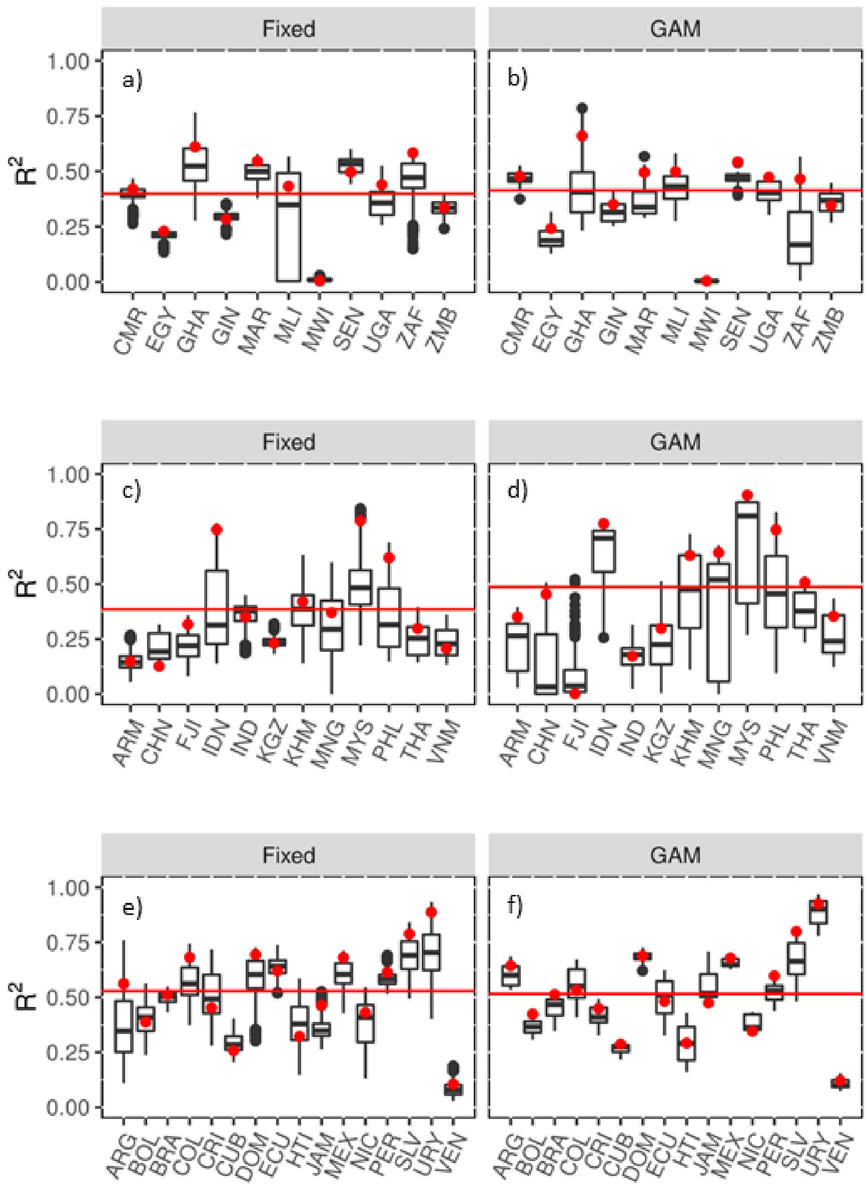
$$MIG_{ij} = \frac{P_i^{\alpha} P_j^{\beta}}{d_{ij}^{\gamma}}$$

With α, β, and γ being parameters, used to indicate the magnitude of the effect for each covariate, that are typically estimated in the statistical modelling framework

$$p_{ij} = \frac{e^{\beta_0 + \beta_1 P_i + \beta_2 P_j - \beta_3 d_{ij}}}{1 + e^{\beta_0 + \beta_1 P_i + \beta_2 P_j - \beta_3 d_{ij}}}$$

where p_{ij} = MIG_{ij}/TOT_j; with MIG_{ij} and TOT_j representing the number of people residing in *j* in the census year that were in *i* 5 years prior to the census and the total population residing in *j* in the census year, respectively

Models Common Across all Countries



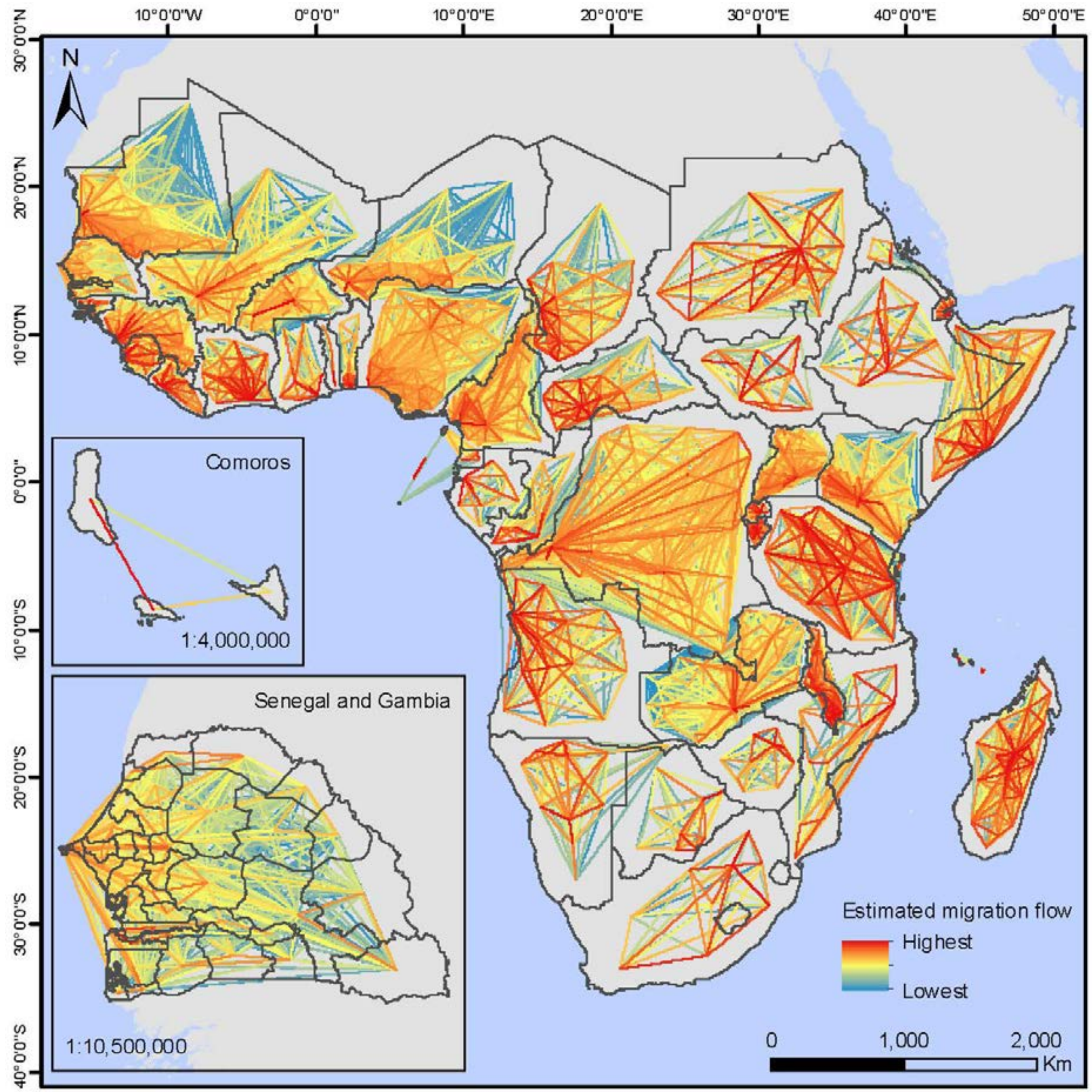
Multi-step approach to identify the model with the greatest predictive power

Best model was then selected using a **leave-one-out cross-validation approach**

R^2 values for all withheld countries were averaged and used to rank each models according to their predictive power averaged across all withheld countries.

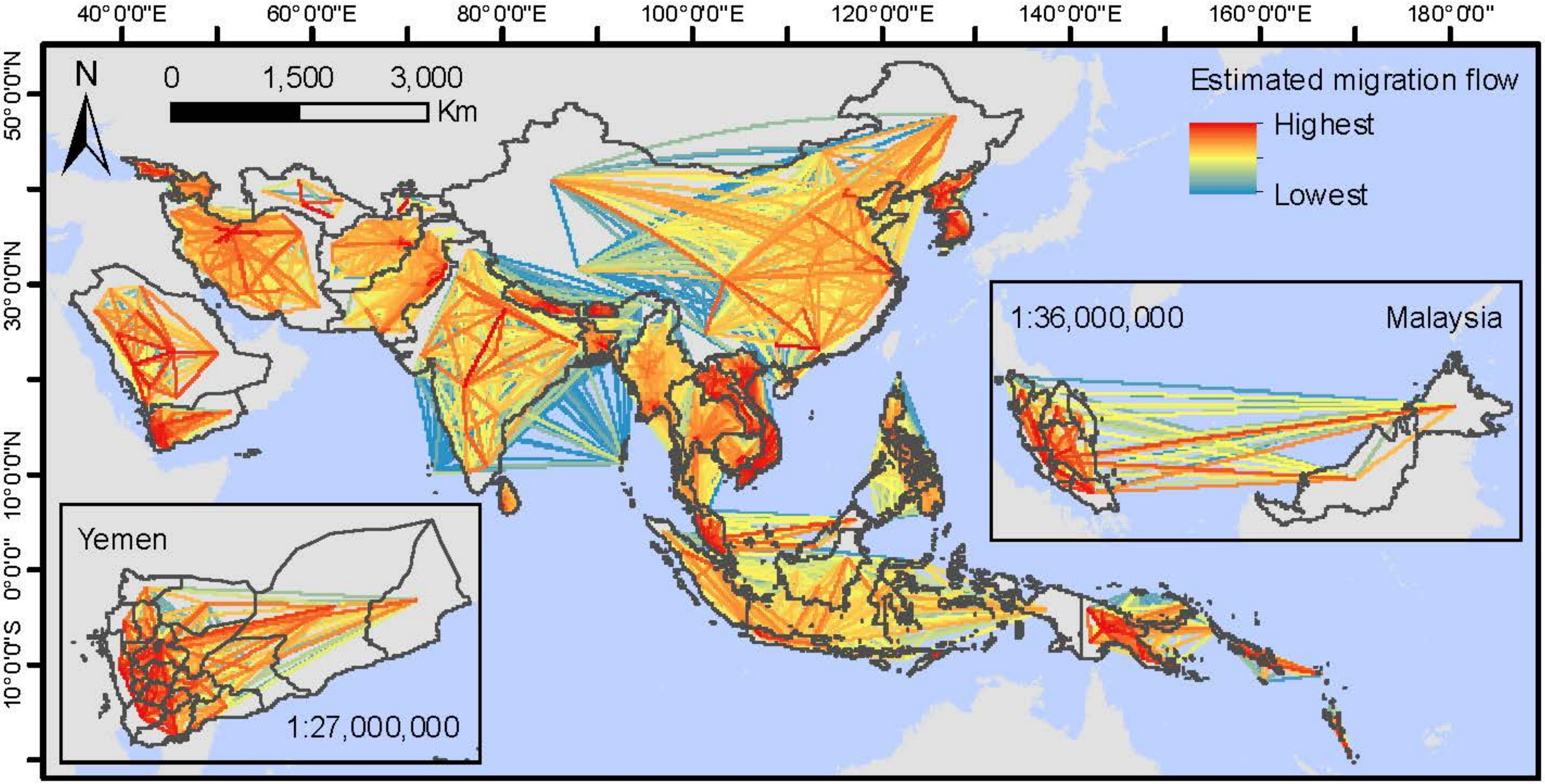
Sorichetta et al., 2016 (Nature Scientific Data)

Internal Migration Flows in Africa



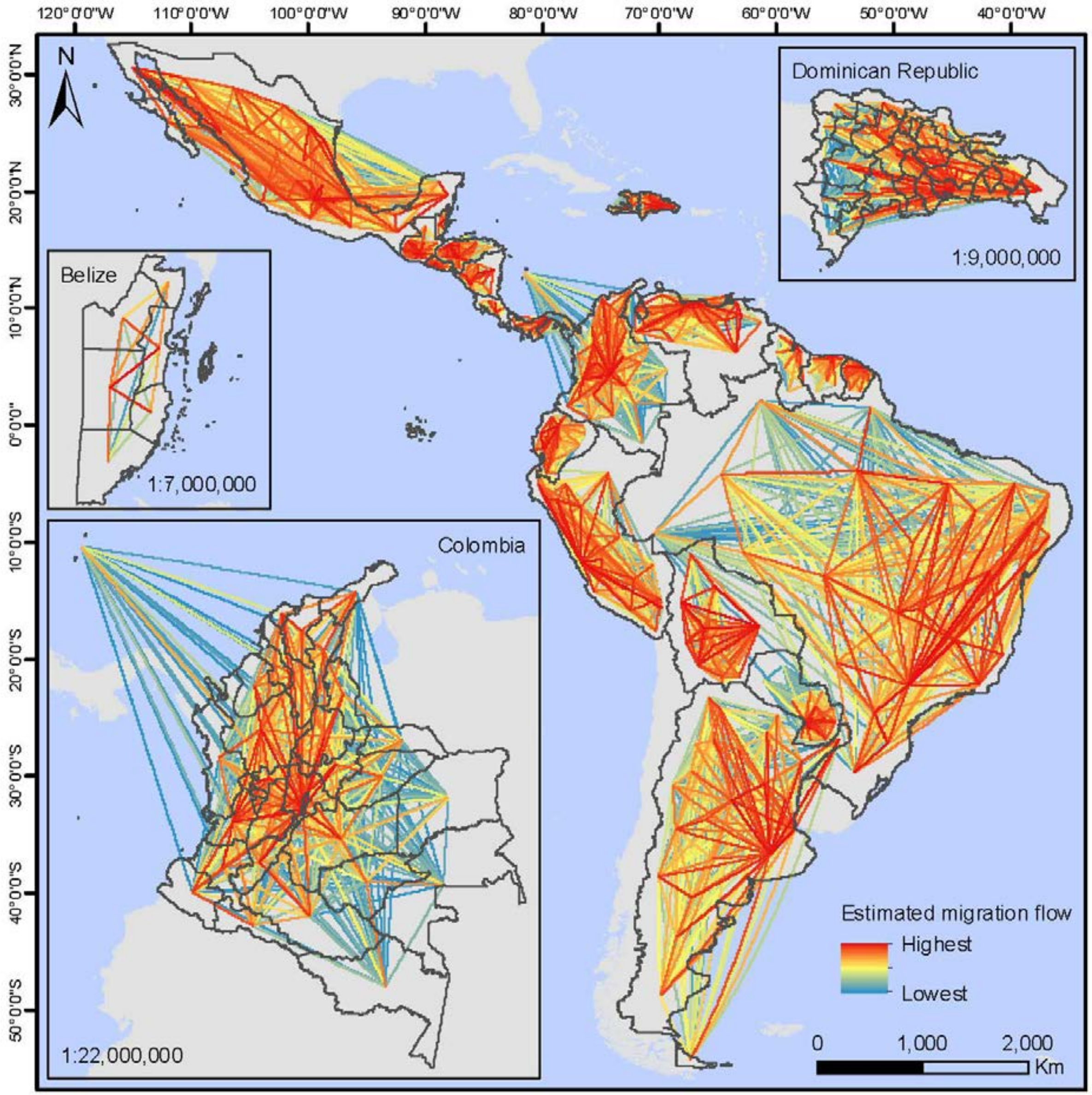
Sorichetta et al., 2016
(Nature Scientific Data)

Internal Migration Flows in Asia



Sorichetta et al., 2016 (Nature Scientific Data)

Internal Migration Flows in LAC



Sorichetta et al., 2016
(Nature Scientific Data)

Validation/Uncertainty

Continent	ISO code	R^2	Error p-value
AFRICA	CMR	0.60	0.07
AFRICA	EGY	0.21	0.20
AFRICA	GHA	0.68	0.21
AFRICA	GIN	0.39	0.09
AFRICA	MAR	0.52	0.14
AFRICA	MLI	0.51	0.14
AFRICA	MWI	0.02	0.06
AFRICA	SEN	0.54	0.12
AFRICA	UGA	0.50	0.11
AFRICA	ZAF	0.49	0.23
AFRICA	ZMB	0.37	0.22
ASIA	ARM	0.11	0.16
ASIA	CHN	0.08	0.19
ASIA	FJI	0.16	0.28
ASIA	KGZ	0.23	0.08
ASIA	IND	0.11	0.15

Error p-value is here defined as the average probability that predicted migration values do not belong to the observed migration dataset.

Sorichetta et al., 2016 (Nature Scientific Data)

Limitations and Caveats (I)

- For consistency, internal migration flows were estimated using a fixed set of pull and push factors common to all countries;
- Use of census data from many years ago for some countries may have generated inaccurate estimates for the period considered in this study (i.e., 2005-2010).
- The model fit varied between countries and could be improved by considering additional locally-specific migration drivers ;

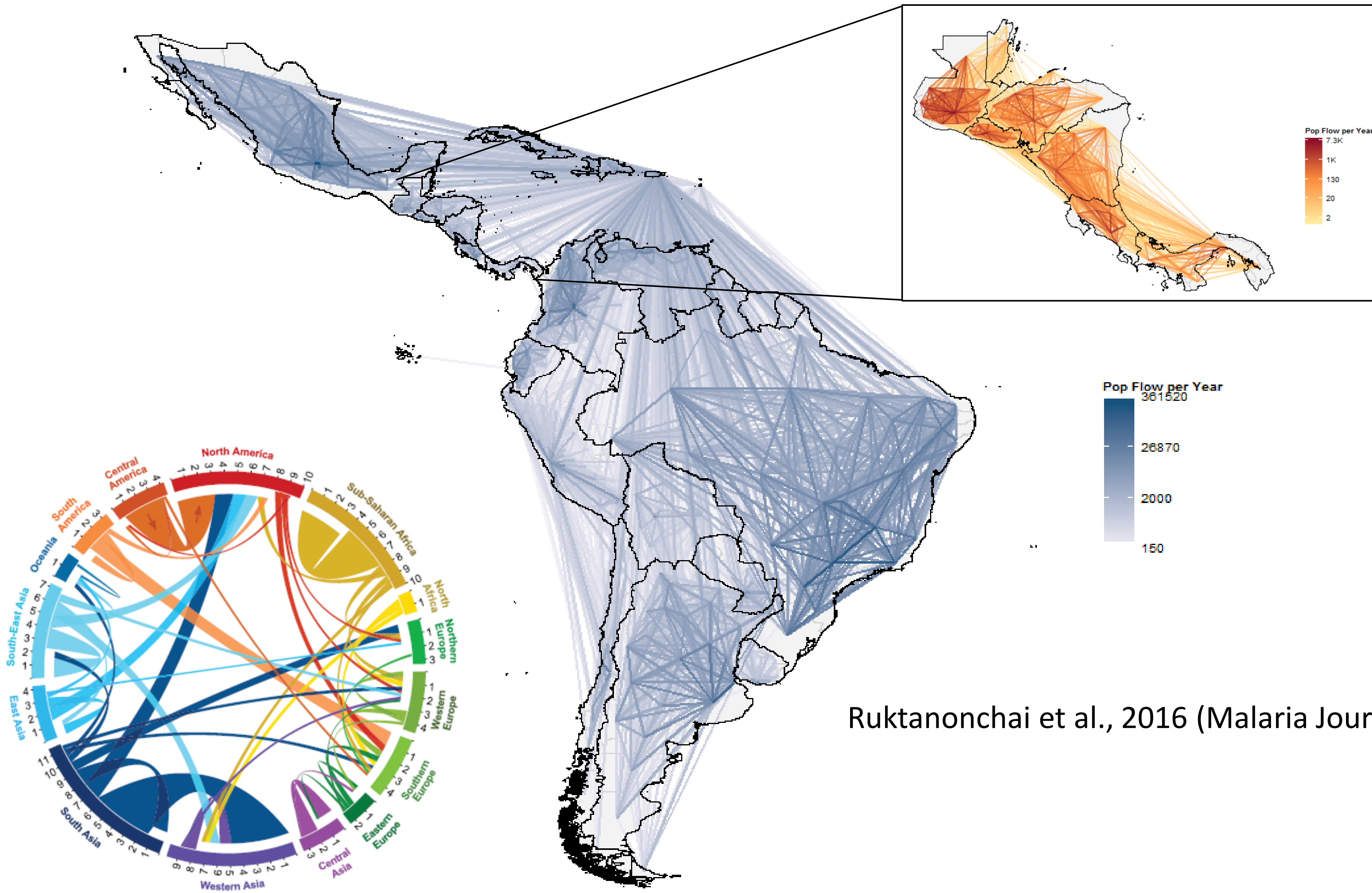
Limitations and Caveats (II)

- Migration models were fitted using only a small sample (ranging between 0.07% and 10%) of the full census for each country;
- The spatial detail at which migration is captured and summarized varies by country;
- The role of some of the pull and push factors, may not have been captured at the spatial level at which they influence internal migration as recorded in the census;

Limitations and Caveats (III)

- Ancillary datasets used to represent pull and push factors are modelling outputs in themselves having a degree of uncertainty that will carry over into the migration estimates;
- Other types of migrations, such as seasonal movements and forced displacements, may be not captured by the model.

Integrating International and Internal Migration Data



Ruktanonchai et al., 2016 (Malaria Journal)

Abel & Sander, 2014 (Science)

Next Step

- **Modelling international migration among subnational administrative units** in Africa, Asia, and LAC as a function of distance using an (**Iterative Proportional Fitting) double-constrain multilevel spatial interaction modelling framework** as described in Dennett & Wilson, 2013 (Environment and Planning A);
- Using IPUMSI-based estimates for internal migration;
- Using Abel & Sander, 2014 (Science) for international migration between countries.

Abel, Sorichetta et al., in preparation

Acknowledgements

UNIVERSITY OF Southampton

For Further Information

worldpop

www.worldpop.org

@WorldPopProject

FLOWMINDER.ORG

www.flowminder.org

@Flowminder